ERASMUS+ PROGRAMME

KEY ACTION 2: COOPERATION FOR INNOVATION AND THE EXCHANGE OF GOOD PRACTICES

STRATEGIC PARTNERSHIPS IN THE FIELD OF EDUCATION, TRAINING AND YOUTH

Blended Learning in Bioinformatics - The SMEs Instrument for Biotech Innovations

BIOTECH-GO
2016-1-BG01-KA202-023686

Biognosis, Greece
O3: BIOTECH-GO Joint Study Education Programme (JSTP)

Health Informatics
(additional)
# Contents

**Introduction to Health Bioinformatics** ................................................................. 4
Definition and overview of health informatics ......................................................... 4

**Medical and Health Informatics** ........................................................................ 6
Clinical Big Data ........................................................................................................ 8
   *Big Data in Health* .............................................................................................. 12
Electronic health records (EHRs) .............................................................................. 12
Social health .............................................................................................................. 17
Lifestyle environmental factors and public health ..................................................... 18

**Healthcare Informatics** ....................................................................................... 19
Clinical informatics .................................................................................................... 19
Integrated data repository .......................................................................................... 21
Clinical research informatics ..................................................................................... 23
   *Common data elements (CDEs) in clinical research* ............................................. 25
Sensor Informatics ..................................................................................................... 29
   *Wearable, implantable, and ambient sensors* ....................................................... 33
   *From sensor data to stratified patient management* ........................................... 36
   *Mobile health* ................................................................................................... 36
Imaging Informatics .................................................................................................... 38
   *Imaging across scales* ....................................................................................... 40
   *From morphology to function* .......................................................................... 40
   *Research initiatives to understand the human brain* ........................................ 41

**References** .......................................................................................................... 43
Introduction to Health Bioinformatics

The role and the importance of health care systems in the quality of life and social welfare in modern society have been broadly well recognized. Also, healthcare is a fundamental and most rapidly growing sector of economy; as such for a society to realize its dream and vision of sustainable development, it is important to be ensured mental and physical health of its individuals. “Health is wealth” goes the popular saying and therefore in every country, the health sector is critical to social and economic growth and development, with ample evidence linking productivity to quality of health care. Thus, any technology or system that is tailored towards improving the healthcare system is most likely to enjoy favorable public perception.

The evolution of medicine over the centuries is always combined with the corresponding technological developments. At the time the Human Genome Project, an international scientific research project which triggered a revolution in biology and medicine, was declared complete in 2003, new opportunities came in order to develop such data, analyze, and present the genomic data in a faster, cheaper, and more accurate and reliable way. This dramatic and rapid advance in technology has affected healthcare in terms of new prevention development, better diagnosis, and improved treatment for regular healthcare. The high volumes of various data from bioinformatics and healthcare informatics domains coupled with analytics are expected to deliver in the near future preventive, predictive and personalized healthcare aids. This can contribute to sustainable national growth and development (Oyelade et al. 2015).

Definition and overview of health informatics

What is Health informatics?

Health informatics is a multidisciplinary field that includes all information technology and telecommunications applications in the field of health, including all other relevant concepts such as medical informatics, health information systems, electronic medical record, telemedicine and mobile health.

Health Informatics (also called health care informatics, healthcare informatics, medical informatics, nursing informatics, or biomedical informatics) is a combination of information science and computer science within the realm of medicine and healthcare, which deals with various fields. It deals with the resources, devices and methods required to optimize the acquisition, storage, retrieval and use of information in health and biomedicine. Health informatics tools not only include computers but also clinical guidelines, formal medical terminologies, and information and communication systems.

Health informatics includes information applications in all areas of the health sector, resulting in a complex and diverse landscape. Thus, health informatics includes the following fields:

– Medical Informatics contains subsets such as Clinical Informatics, Pathology Informatics, Dental Informatics and Medical Education Informatics.

– Nursing Informatics, with sub-sector Nursing Education Informatics.
– Pharmaceutical Informatics.
– Bioinformatics and Clinical bioinformatics.
– Public Health Informatics.
– Informatics for Health Research.
– Healthcare Education Informatics.
– Consumer Health Informatics.

Figure 1. Health Informatics.

There are numerous current areas of research within the field of Health Informatics, including Bioinformatics, Image Informatics (e.g. Neuro-informatics), Clinical Informatics, Public Health Informatics, and also Translational Bioinformatics (TBI). Research done in all subfields of Health Informatics range from data acquisition, retrieval, storage, analytics employing data mining techniques, and so on. Health Informatics is applied to the areas of nursing, clinical care, dentistry, pharmacy, public health, occupational therapy, and biomedical research.

Various studies done on Health Informatics uses data from a particular level of human existence. Bioinformatics uses molecular level data, Neuroinformatics employs tissue level data, Clinical Informatics applies patient level data, and Public Health Informatics utilizes population data. On the other hand, TBI exploits data from each of these levels, from molecular level to entire population. TBI is used to mainly answer clinical level questions. These researches in various subfields are used to improve health care system.
Medical and Health Informatics

Medical informatics can be defined as the study of medical data translated to information, which in turn is translated to derive useful knowledge for the benefit of health care. It also involves their storage, retrieval, and optimal use for making health care-oriented decisions and problem solving. On more general terms,

Medical Informatics (MI) is the application of computers, communications and information technology systems to all fields of medicine and medical care.

It is derived from two words: medical and informatics where medical indicates the area of research and informatics indicates the methodology applied to support it (Figure 2).

Figure 2. Definition of medical informatics.

MI deals with resources, devices, and methods required for optimizing the acquisition, storage, retrieval and use of information technology in health care. MI is formed by intersection of different scientific fields, such as computers, engineering, biology, mathematics, and physics. Hence, a medical informatics specialist requires knowledge and skills in these different domains to survive and contribute towards growth of the field.

For any kind of analysis and processing, medical data is required. Medical data can be further divided into alphanumeric data, medical imaging, and pathological and physiological signals.

Biomedical informatics (Figure 3) or medical informatics is an emerging discipline underlying the acquisition, maintenance, retrieval and application of knowledge and information in research, education, and service in health-related basic sciences, clinical disciplines, and health care administration with computer science, statistics, engineering, mathematics, information technologies and management.
Biomedical informatics also provides the tools and skills needed for the development and application of new technology for improving patient care, medical education, health sciences and management for healthcare/hospital systems.

Biomedical informatics coalesces the related fields of Medical Informatics (now being named Health Informatics) and Bioinformatics. Health Informatics contains subsets such as Telemedicine, Clinical Informatics, and Dental Informatics, Pharmaceutical Informatics, Nursing Informatics and Public Health Informatics.

It is important here to define the difference between the terms data and information. Data is collection of record from a source, for example, a subject, books, and internet. Data can be of two types: primary and secondary. Information, on the other hand is processed data after data classification, which means converting raw data into meaningful form after it is classified. The central to both medical informatics and bioinformatics is the collection and analysis of information. While medical informatics is more concerned with structures and algorithms for the manipulation of the data and how it can be applied in healthcare, bioinformatics is more concerned with the data itself and its biological implications.

Medical informatics is further divided into four broad categories (see below Figure 4).

---

**Figure 3. Biomedical Informatics at a glance.**

**Figure 4. Categories of Medical Informatics.**
Health care is becoming an increasingly data-intensive field as doctors and researchers generate gigabytes of medical data on patients and their illnesses. While a patient visiting the doctor before 15 years may have only generated a few data points basic information such as weight, blood pressure, and symptoms a medical encounter today may leave a long trail of digital data from the use of high-definition medical imaging to implantable or wearable medical devices such as heart monitors. More importantly, as doctors and hospitals transition away from paper medical records, this data is increasingly being collected and made available in an electronic format. The availability of large data sets of digital medical information has made possible the use of informatics to improve health care and medical research. Often referred to as “in silico” research, informatics offers a new pathway for medical discovery and investigation. The field of bioinformatics has exploded within the past decade to keep pace with advancements in molecular biology and genomics research. Researchers use bioinformatics to gain a better understanding of complex biological processes by, for example, analyzing DNA sequences or modeling protein structures.

With the ability to deal with large volumes of both structured and unstructured data from different sources, big data analytical tools hold the promise to study outcomes of large-scale population-based longitudinal studies, as well as to capture trends and propose predictive models for data generated from electronic medical and health records. A unique opportunity lies in the integration of traditional medical informatics with mobile health and social health, addressing both acute and chronic diseases in a way that we have never seen before.

**Clinical Big Data**

The last decade has seen an exponential growth in the quantity of clinical data collected worldwide, triggering an increase in opportunities to reuse the data for biomedical research. The transition from paper medical records to electronic clinical systems has been accelerated by an emphasis on modernizing our health care infrastructure. This resulted in a significant growth in the amount of clinical data being collected.

Figure 5 illustrates the fast increase in the number of publications referring to “big data,” regardless of disciplines, as well as those in the healthcare domain. Although the popularity of big data is recent, the underlying challenges have existed long before and have been actively pursued in health research.
Figure 5. (a) Cumulative number of publications referring to “big data” indexed by Google Scholar. (b) Cumulative number of publications per health research area referring to “big data”, as indexed in IEEE Xplore, ACM Digital library, PubMed (National Library of Medicine, Bethesda, MD), Web of Science, and Scopus. (Andreu-Perez, Poon, et al. 2015)

Big Data is a term used to describe data sets with such large volume or complexity that conventional data processing methods are not good enough to deal with them. Big Data has been described disparately by different people. The most popular definition of Big Data is the 5Vs, which are Volume, Velocity, Variety, Verification/Veracity, and Value.

- **Volume** means the large amounts of data used. The amount of data being created is vast compared to traditional data sources.
- **Velocity** means the speed at which new data is generated. Data is being generated extremely fast - a process that never stops.
- **Variety** means the level of the complexity of data. Data comes from different sources and is being created by machines as well as people.
- **Veracity** is used to measure the genuineness (or trustworthiness) of the data obtained. Big data is sourced from many different places; as a result you need to test the veracity/quality of the data.
- The **value** gives how good the quality of data is.
- Another factor has also been considered, which is **Variability** (consistency of data over time).

These characteristics are summarized in Figure 6 along with the key features that each capture. The definition of Big Data might be subjected to technological advances in the future. Big Data infrastructure is a framework, which covers important components including Hadoop (hadoop.apache.org), NoSQL databases, massively parallel processing (MPP), and others, that is used for storing, processing, and analyzing Big Data.
What exactly is big data?

➢ A study published by The McKinsey Global Institute (MGI) in June 2011: Define Big Data as “datasets whose size goes beyond the ability to capture, store, manage and analyze database”.

➢ A report delivered to the U.S. Congress in August 2012 defines big data as “large volumes of high velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of the information”.

➢ Demchenko et al. defines big data by five Vs: Volume, Velocity, Variety, Veracity and Value.

➢ Wikipedia defines “big data” as “data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time”.

➢ In 2012 Gartner defined Big Data as “information assets characterized by their high volume, high velocity and high variety, which demand innovative and efficient processing solutions for the improvement of knowledge and decision making in organizations”.

However, the definition extends well beyond data volume and includes issues such as data heterogeneity and dynamism, each of which presents its own unique challenges. Beyond the technical considerations, the potential utility of such data has led to increasing awareness across many distinct disciplines. In oncology, for example, the term has become synonymous with the concept of evidence-based decision-making, where integration and analysis of large volumes of data aims to improve the quality, efficiency, cost and outcome of critical business decisions.

Big data includes information garnered from social media, data from internet-enabled devices (including smartphones and tablets), machine data, video and voice recordings, and the continued preservation and logging of structured and unstructured data. Big data refers to the tools; processes and procedures, which allow an organization to create, manipulate and manage very large data sets and storage facilities. Big data enables an opportunity for aggregation and integration leading to cost effective and patient care.
Another aspect of big data in biomedicine is the use of non-traditional data sources. These were well illustrated, both literally and figuratively, in a 2012 paper by Eric Schadt. A complex and detailed figure (Figure 7) showed various data types that could be mined for their effects on human health: weather, air traffic, security, cell phones, and social media among others. But strikingly to those reading the paper just a few years later, the list did not include personal activity trackers, e.g., FitBit, Jawbone, or even the Apple watch. This omission of such a popular technology today is indicative of what a fast-moving field this is.

![Figure 7. Heterogeneous and non-traditional sources of big data](image)

Big data enabled by technological advances in micro- and nano-electronics, nano materials, interconnectivity provided by sophisticated telecommunication infrastructure, massive network-attached storage capabilities, and commodity-based high-performance computing infrastructures. The ability to store all credit card transactions, all cell phone traffic, all e-mail traffic, video and images from extensive networks of surveillance devices, satellite and ground sensing data informing on all aspects of the weather and overall climate, and now to generate and store massive data informing on our personal health including whole genome sequencing data and extensive imaging data, is driving a revolution in high-end data analytics to make sense of the big data, drive more accurate descriptive and predictive models that inform decision making on every level, whether identifying the next big security threat or making the best diagnosis and treatment choice for a given patient.
**Big Data in Health**

Big Data in medicine can have many uses and applications, in areas such as epidemiology, clinical records, clinical operation, and administrative management, among others. The health area integrates large amounts of data, which must be stored, classified, analyzed and consulted, all of them with systematized and structured processes generate useful information for the achievement of advances in health discoveries and in the administration of medical resources.

In the health sector, the data on which Big Data’s analysis techniques can be applied are as diverse as they can be (Personal data, clinical data, administrative data). Existing analytical techniques can be applied to the vast amount of existing (but currently unanalyzed) patient-related health and medical data to reach a deeper understanding of outcomes, which then can be applied at the point of care. Ideally, individual and population data would inform each physician and her patient during the decision-making process and help determine the most appropriate treatment option for that particular patient. The information obtained once processed analyzed and classified the data allow the obtaining of a preventive, predictive, personalized and effective medicine.

*Figure 8. An applied conceptual architecture of big data analytics.* (Raghupathi & Raghupathi 2014)

**Electronic health records (EHRs)**

The electronic health record (EHR) is also known as Electronic Medical Record (EMR), Computerized Medical Record (CMR), Electronic Clinical Information Systems (ECIS) and Computerized Patient Record (CPR).

In May 2008 the National Alliance for Health Information Technology released the following definitions in an effort to standardize terms used in HIT:
**Electronic Medical Record**: “An electronic record of health-related information on an individual that can be created, gathered, managed and consulted by authorized clinicians and staff within one healthcare organization”.

**Electronic Health Record**: “An electronic record of health-related information on an individual that conforms to nationally recognized interoperability standards and that can be created, managed and consulted by authorized clinicians and staff across more than one healthcare organization”.

**Personal Health Record**: “An electronic record of health-related information on an individual that conforms to nationally recognized interoperability standards and that can be drawn from multiple sources while being managed shared and controlled by the individual”.

EHR is an evolving concept defined as a systematic collection of individual patients or populations electronically-stored health information in a digital format. It is capable of organizing clinical data by phenotypic categories. Also, EHR enables the communication of patient data between different healthcare professionals (specialists etc.), as it is capable of being shared across different health care settings, by being embedded in network-connected enterprise-wide information systems. Especially, this term refers to computer software that physicians use to track all aspects of patient care. Typically, this broader term also encompasses the practice management functions of billing, scheduling, etc.

EHRs are real-time, patient-centered records that make information available instantly and securely to authorized users. While an EHR does contain the medical and treatment histories of patients, an EHR system is built to go beyond standard clinical data collected in a provider’s office and can be inclusive of a broader view of a patient’s care. EHRs can:

- Include a whole range of data in comprehensive or summary form, including demographics, patient’s medical history, diagnoses, medications and allergies, treatment plans, immunization status, laboratory test results, radiology images, vital signs, personal stats like age and weight, and billing information.
- Allow access to evidence-based tools that providers can use to make decisions about a patient’s care.
- Automate and streamline provider workflow in health care settings.

One of the key features of an EHR is that health information can be created and managed by authorized providers in a digital format capable of being shared with other providers across more than one health care organization. EHRs are built to share information with other health care providers and organizations - such as laboratories, specialists, medical imaging facilities, pharmacies, emergency facilities, and school and workplace clinics - so they contain information from all clinicians involved in a patient’s care.
Figure 9. Samples of HERs.
Figure 10. Sample view of an HER by Practice Fusion, one of the largest free HER platforms, based on the Cloud. In the above sample we observe options such as: Charts, Past Medical History (PMH), List of prescriptions (Rx List), Drug allergies, Vaccinations, Dating and Referrals. [Online] Available: https://www.practicefusion.com/

Potential benefits of an EHR?

Benefits of an EHR can be categorized as follows:
A major challenge for clinical bioinformatics pertains to the accommodation of the range of heterogeneous data into a single, query able database for clinical or research purposes. HER has recently emerged and stimulated increased research interest. An ideal EHR provides complete personal health and medical summary by integrating personal medical information from different sources. The inclusion of genetic imaging and population-based information in EHR has the potential to provide patients with valuable risk assessment based on their genetic profile and family history and to carve a niche for personalized cancer management.

EHRs describing patient treatments and outcomes are rich but underused information. Traditional health data centers capture and store an enormous amount of structured data concerning a wide range of information. Mining EHRs is a valuable tool for improving clinical knowledge and supporting clinical research, for example, in discovering phenotype information. Mining local information included in EHR data has already been proven to be effective for a wide range of healthcare challenges, such as disease management support, pharmacovigilance, building models for predicting health risk assessment, enhancing knowledge about survival rates, therapeutic recommendation, discovering comorbidities, and building support systems for the recruitment of patients for new clinical trials. Most of this work focused on the analysis of very large multidimensional longitudinal patient data collected over many years. However, most clinical databases provide low temporal resolution information due to the difficulty in collecting rich long-term time-series data. To bridge this gap, current clinical databases can be enhanced by connecting with mobile health platforms, community centres, or elderly homes such that other information can be incorporated into the system to facilitate clinical decision-making and address unanswered clinical questions. One interesting direction will be to build patient-specific models using data already available in existing clinical databases, and, then, update the model with data that can be collected...
outside the hospitals. In particular, some chronic diseases are manifested with acute events that are unlikely to be predictable solely by sporadic measurements made within hospitals.

Taking thoracic aortic dissection, a relatively rare disease (3–4 per 100 000 people per year), as an example, the disease is typically manifested as a tear in the intimal layer of the aorta, which can later on develop into either type A (involving both ascending and descending aorta) or type B dissection (involving descending aorta only). Type-A patients would require immediate surgical intervention, whereas for type B dissection, it is generally considered as a chronic condition requiring careful long-term control of blood pressure (BP).

Individuals with connective tissue disorders such as Marfan syndrome (MFS) are often more susceptible to aortic aneurysms or tears. Large-scale population screening for this rare disease will, therefore, be useful in identifying people who are at higher risk of developing aortic dissection. For a tear to develop into type A dissection, while others into type B dissection, one hypothesis would be that it is due to different flow patterns generated close to the tear location and across the aorta. Although an initial model built from imaging can give good insights into the problem, this does not take into account progressive hemodynamic variation over time and the impact of lifestyle and daily activities. By incorporating ambulatory BP profiles, it is possible to create simulation results as a longitudinal model spanning over a longer period of time for a better understanding of disease progression as summarized in Figure 11.

**Figure 11. Integration of imaging, modeling, and real-time sensing for the management of disease progression and planning of intervention procedures.** This example of thoracic aortic dissection illustrates how risk stratification and subject-specific haemodynamic modeling substantiated with long-term continuous monitoring are used to guide the clinical decision process.

**Social health**

One of the primary tasks of telemedicine involves connecting patients and doctors beyond the clinic. However, this communication has been expanded, with the involvement of social networks, to new levels of social interaction. This new feature has opened up new possibilities of patient-to-patient communication regarding health beyond the traditional doctor-to-patient paradigm. One-fourth of patients with chronic diseases, such as diabetes, cancer, and heart conditions, are now using social
network to share experiences with other patients with similar conditions, thereby providing another potential source of big data. In addition to biological information, geolocation and social apps provide an additional feature to understand the behaviors and social demographics of patients, while avoiding resource intensive and expensive studies of large statistical sampling. This advantage has already been exploited by several epidemiological studies in areas, such as influenza outbreaks, collective dynamics of smoking, and the misuse of antibiotics. Text messages and posts on online social networks are also a valuable source of health information, e.g., for the better management of mental health. Compared to traditional methods, such as surveys, fluctuations and regulation of emotions, thoughts and behaviors analyzed over social network platforms, such as Twitter, offer new opportunities for the real-time analysis of expressed mood and its context. For example, when validating against known patterns of variation in mood, the $2.73 \times 10^9$ emotional tweets collected over a 12-week period in a study reported by Larsen et al. claimed to find some correlation between emotion tweets and global health estimates from the World Health Organization on anxiety and suicide rates.

Social media and internet searches can also be combined with environmental data, such as air quality data, to predict the sudden increase of asthma-related emergency visits. Similar models are anticipated to help other areas of public health surveillance.

**Lifestyle environmental factors and public health**

Climatological data, such as heat-stress and cold-related mortality, present another dimension to predict personal health. Recent remote sensing technologies and geographic information systems allow climate data for global land areas to be interpolated at a spatial resolution of 500 m to 1 km. Achieving high-resolution measurements are necessary so as to be able to monitor the real impact of pollution on human wellbeing in urban environments. With this aim the dense grid of wireless sensor networks facilitates the capture of spatiotemporal variability in toxic air pollutants. Such technologies will become increasingly important for connecting epidemic intelligence with infectious disease surveillance and launching effective heat response plans. Similarly, patterns of social factors influencing unhealthy habits such as smoking can be studied using the collective dynamics of social networks. As an example of this, Christakis and Fowler found that smokers mostly belonged to the periphery of social networks, and by the time of quitting, they behaved collectively. In addition, smokers with higher education tended to have a greater influence on their peers toward smoking behavior, compared to less educated smokers. As regards psychological states, emotional levels denoting hostility and stress, expressed in social media such as Twitter tweets, can serve as predictors of heart disease mortality per geographical area.

A mobile phone is an excellent platform to deliver personal messages to individuals to engage them in behavioral changes to improve health. Mobile phone messaging can be used as an alternative to deliver motivational and educational advices for changing population lifestyles. Although at present, there is limited evidence that mobile messaging-based interventions support preventive health care for improving health status and health behavior outcomes, a better understanding of how this platform can be used is an interesting area to explore. For example, type-2 diabetes is generally thought to be preventable by lifestyle modification; however, successful lifestyle intervention programs are often labor intensive. It has been shown that mobile phone messaging can be used as an alternative to deliver motivational and educational advices for changing population lifestyles.
**Healthcare Informatics**

Healthcare informatics deals with the best use of information, through the help of technology, in improving healthcare, public health, and biomedical research. It is more about information than the technology. Informatics is a blend of life-health sciences and computer and informatics sciences to make people better.

Healthcare informatics, in the simplest definition, is the division of healthcare that is involved in providing the right and useful information about a patient’s state of health to the right person, the healthcare professionals, at the right time it is required, thereby guiding them in making and taking informed decisions about their patient’s treatment. There is a synergy and exchange of information in Healthcare informatics between patients, healthcare professionals, healthcare planners, Internet Technology providers and management of healthcare facilities.

The name healthcare informatics originated from the coalition of similar terms (though they are sometimes used interchangeably, but there are lots of variances between them), these are: electronic health, popularly known as eHealth, Information management and technology also known as IM&T, Medical informatics and Telehealth.

Healthcare Informatics combines information and understanding from medical areas (pre-clinical, clinical and post-clinical), healthcare administration and management and information technology. Health Informatics experts understand information technology, application of technology to solve real life challenges and managing multifaceted processes of executing novel answers into organizations. Besides these common capabilities, an expert in Health Informatics is also conversant with the industry, both from the standpoint of offering medical services, and from the standpoint of managing, challenging and demanding medical organizations and procedures.

Medical Informatics makes available the essential tools to relate data and knowledge in the process of decision-making. Information management and technology ensures collection and information management from various sources, including the processing and delivery of the same. The objective is to assist in health care, increase the performance within organization, resolve business issues as well as assist service and business plans.

eHealth is a new term often used in European countries that implements electronic and digital technologies for healthcare practices, processes and communications.

Telehealth combines technology with health services that gives opportunity to people with medical conditions to access healthcare within the comfort of their home. A home Pod device installed in a patient’s home is an example of telehealth that can be used to monitor the patient’s health condition, the information that can then be communicated securely over the server to a remote clinician.

**Clinical informatics**

Clinical informatics, also known as healthcare informatics, is the study and use of data and information technology to deliver health care services and to improve patients’ ability to monitor and
maintain their own health. The data and clinical decision support involved in this field are developed for and used by clinicians, patients, and caregivers.

The field includes:

- Methods to collect, store, and analyze health care data
- The study of information needs and cognitive processes, and optimal ways to meet those needs
- Methods to support clinical decisions, including summarization, visualization, provision of evidence, and active decision support
- Optimizing the flow of information and coordinating it with care providers’ and patients’ workflows to maximize patient safety and care quality
- Methods and policies for information infrastructure, including privacy and security

![Figure 12. Clinical informatics.](image)

The development of electronic medical record (EMR) systems began as a means to document clinical activities for in-patients and out-patients. They have evolved as the primary front-line patient care clinical tool for medical professionals. The completion of the Human Genome Project opened the era of research in genomics and proteomics. Genome research provides keys to understanding the mechanisms of disease. In clinical informatics, the widespread adoption of the EMR system has generated large amounts of heterogeneous clinical data—some structured and others unstructured. In addition, the explosive health-related contents from online communities, mobile applications, and electronic personal health records increased the availability of non-traditional data on individual activities and life style. Integrated use of EMR and bioinformatics is beginning to influence the changes in the research paradigm—that is, rapid introduction of new concepts into the point of care. Dr. Want used clinical bioinformatics (CBI) with the definition of “the clinical application of bioinformatics-associated sciences and technologies to understand molecular mechanisms and potential therapies for human disease”, a new and important concept for the development of disease-specific biomarkers, mechanism-oriented understanding and individualized medicine. CBI is a new emerging science combining clinical informatics, bioinformatics, medical informatics, information technology, mathematics, and omics science together. Also, plays an important role in a number of clinical applications, including omics technology, metabolic and signaling pathways, biomarker discovery and development, computational biology, genomics, proteomics, metabolomics,
pharmacomics, transcriptomics, high-throughput image analysis, human molecular genetics, human tissue bank, mathematical medicine and biology, protein expression and profiling and systems biology. CBI aims to deal with the challenge of integrating genomic and clinical data to accelerate the translation of knowledge into effective treatment plan development and personalized prescription. It is to assist clinicians in various ways, including new biomarker discovery, identification of genotype and phenotype correlations, and pharmacogenomics at the point of care.

Understanding the interaction between clinical informatics and bioinformatics is the first and critical step to discover and develop the new diagnostics and therapies for diseases.

**Figure 13. Clinical Bioinformatics for Personalized Health.**

**Integrated data repository**

We are well into the era of “secondary use” of health data, of which an important part is reusing clinical data for both publishable research and quality improvement. A 2010 survey defined “integrated data repository” (IDR) as a data warehouse integrating various sources of clinical data to support queries for a range of research-like functions. An integrated data repository (IDR) containing aggregations of clinical, biomedical, economic, administrative, and public health data is a key component of an overall translational research infrastructure. Such a repository can provide a rich platform for a wide variety of biomedical research initiatives. IDRs for research are usually designed to allow “attribute-centric” queries: that is, queries for the set of patients who meet some criteria based on values of clinical observations or characteristics (also called “attributes” in the common Entity–Attribute–Value (EAV) model.

Current clinical and translational research increasingly relies on the existence of robust integrated data repositories (IDRs) with administrative, clinical, and “-omics” data.

IDRs are becoming an essential resource enabling the biomedical data reuse on larger amounts and sources of data. Several initiatives have been carried out on IDRs to provide access to biomedical research data, either as federated query tools or as centralized repositories. In most solutions, the adoption of a common data format was key. However, to our knowledge, the use of specific health information standards was limited. Besides, it is agreed that the reliability of data reuse depends greatly on its Data Quality (DQ). Certainly, DQ assessment is considered a key component to any IDR, where some successful examples can be found in the recent literature.
Following clear warehouse design principles can lower long-term maintenance costs for organizations that are currently building or significantly restructuring their data warehouses. Maintenance of those warehouses is very costly, and architectural changes are complicated by existing dependencies. Getting the right architecture early during the warehouse creation is crucial. The objective to provide an integrated data repository to researchers, clinicians, and administrators can be met in number of ways.

Integrated data repositories (IDRs) are indispensable tools for numerous biomedical research studies. There are three large IDRs (Informatics for Integrating Biology and the Bedside (i2b2), HMO Research Network’s Virtual Data Warehouse (VDW) and Observational Medical Outcomes Partnership (OMOP) repository) in order to identify common architectural features that enable efficient storage and organization of large amounts of clinical data.

<table>
<thead>
<tr>
<th>Table 2. Basic requirements for an IDR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Re-use:</strong> routine care data are re-used for research purposes or clinical purposes (e.g., inform care of patients based on past experience with similar patients)</td>
</tr>
<tr>
<td><strong>Integration:</strong> data are integrated to facilitate long-term lifetime analysis, such that data from disparate sources are linked to the corresponding patient (billing data, clinical data) and linked to the corresponding event (order entry, order fulfillment). Moreover, semantically identical or semantically related data are also linked.</td>
</tr>
<tr>
<td><strong>Organization:</strong> the IDR can accommodate a wide range of source systems and is easy to use and extend. It strikes a balance between graceful evolution and stability of the data structures. For example, major restructuring does not occur often, and most new data sources can be integrated without major schema change.</td>
</tr>
<tr>
<td><strong>Maintenance:</strong> the IDR is optimized for easy maintenance, especially with respect to adapting to changes in source systems and is robust to turnover of maintenance staff and data analysts.</td>
</tr>
</tbody>
</table>

Whereas the initial design of many clinical data repositories was driven by provision of decision support or evaluation of quality of care, their research use is rapidly increasing with significant impact on their design. As with similar efforts in informatics, adherence to general principles will provide some immediate benefits, with the potential for future, unanticipated benefits as well.

The results of a pilot project of the Spanish Ministry of Health, Social Services and Equality (2015/07PN0010), envisaged to the development of a National IDR of maternal-child care information. An IDR was developed as solution, which, based on health information standards,
ensured a common interface for monitoring BPs of different hospitals and regions, and having its DQ assessed ensures a reliable data reuse.

Examples might include correlative studies seeking to link clinical observations with molecular data, data mining to discover unexpected relationships, and support for clinical trial development through hypothesis testing, cohort scanning and recruitment. Significant challenges exist to the successful construction of a repository, and they include the ability to gain regular access to source clinical systems and the preservation of semantics across systems during the aggregation process.

Clinical research informatics

The documentation, representation, and exchange of information in clinical research are inherent to the very notion of research as a controlled and reproducible set of methods for scientific inquiry. Clinical research is the branch of medical science that investigates the safety and effectiveness of medications, devices, diagnostic products, and treatment regimens intended for human use in the prevention, diagnosis, treatment, or management of a disease. Clinical research enables new understanding and practices for prevention, diagnosis, or treatment of a disease or its symptoms.

Clinical research is critical to the advancement of medical science and public health. Conducting such research is a complex resource and information-intensive endeavor involving multiple actors, workflows, processes, and information resources. Ongoing large-scale efforts have explicitly focused on increasing the clinical research capacity of the biomedical sector and have served to increase attention on clinical research and related biomedical informatics activities throughout the governmental, academic, and private sectors. The result is the emergence and relatively rapid expansion of a new sub-discipline of biomedical informatics focused on clinical research, referred to as Clinical Research Informatics (CRI).

Clinical research informatics (CRI) is the rapidly evolving and dynamic sub-discipline within biomedical informatics that focuses on developing new informatics theories, tools, and solutions to accelerate the full translational continuum: basic research to clinical trials, clinical trials to academic health center practice, diffusion and implementation to community practice, and ‘real world’ outcomes. A recent factor accelerating CRI research and development efforts is the growing interest in sustainable, large-scale, multi-institutional distributed research networks for comparative effectiveness research. Given the large landscape that comprises translational science, CRI scientists are asked to conceive innovative informatics solutions that span biological, clinical, and population-based research. It is therefore not surprising that the field has simultaneously borrowed from and contributed to many related informatics disciplines.

In 2013, Peter Embi identified 6 categories of CRI, based on several articles:

➢ Data and Knowledge Management, Discovery and Standards;
➢ Clinical Data Re-Use for Research;
➢ Researcher Support and Resources;
➢ Participant Recruitment;
➢ Patients/Consumers and CRI;
➢ Policy, Regulatory and Fiscal Matters.
He also concluded that the field of CRI is broad and rapidly advancing.

The description of recent advances in CRI will be structured according to three use cases of clinical research: Protocol feasibility, patient identification and recruitment, and clinical trial execution. Basically, these three use cases cover the full range of clinical research.

– **Protocol Feasibility**

A key process in clinical research is protocol feasibility. The task is to estimate how many patients are available according to a set of feasibility criteria (e.g. diabetes type II patients, aged 18-60, HbA1c >8%) in a defined setting (e.g. hospitals A, B and C) and time frame (e.g. within past 12 months). A clinical study can only be successful, if a patient cohort of adequate size is existing. Patient counts are usually sufficient to answer this question, i.e. aggregated, irreversibly de-identified data.

– **Patient Identification and Recruitment**

Once a clinical study is initiated, eligible patients need to be identified. It is well-known that a large proportion of clinical trials are delayed or not successful due to issues with patient recruitment. In contrast to protocol feasibility, aggregated patient counts are not sufficient to support patient identification and recruitment. Candidate patient lists need to be generated and communicated to treating physicians. In a second step, local study teams get involved.

– **Clinical Trial Execution**

Data management in clinical trials is costly due to the high documentation workload – on average 180 pages per patient in a trial – and the need for high data quality. To support clinical trial execution, data can be transferred from EHR systems into electronic data capture (EDC) systems.

Given the rapid advances in biomedical discoveries, the growth of the human population, and the escalating costs of health care, the need to accelerate and improve clinical research is essential. In parallel, there is a strong need for global collaboration to address the huge challenges of efficient and effective data capture in clinical research. As such, the field of CRI has emerged as critical to solving many of the current challenges faced by clinical researchers and the research enterprise. There is little doubt the field is poised for continued and rapid growth, and that growth will be reflected in the biomedical informatics literature in the years to come. Open metadata, content standards with semantic annotation and computable eligibility criteria are key success factors for the future of CRI.
Common data elements (CDEs) in clinical research

Consistency in data collection is a fundamental principle of scientific research in general and clinical trials in particular. In any given study, each opportunity for data collection is expected to meet specifications independent of time, location, or people involved. While consistency of data collection within an individual study is essential for maintaining data quality and enabling analysis, consistency of data collection across multiple studies brings additional value. As biomedical research becomes more data-intensive, and as policy and practices promote increased data sharing, greater scientific
opportunities emerge from the comparison and secondary use of biomedical research data. Data sharing to support the combination of data across datasets for strengthening inferences and performing new analyses is rapidly becoming a general expectation. One empiric approach for achieving consistency in data collection within and across research studies is the use of Common Data Elements (CDEs).

What are CDEs?

The term “Common Data Element” was initially developed by Silva and Wittes in 1999 for case report forms used in National Cancer Institute clinical trials and has continued to evolve. As used currently, a CDE is a combination of a precisely defined question (variable) paired with a specified set of responses to the question that is common to multiple datasets or used across different studies. The primary context for CDEs is in research where precision, reproducibility, and cross-study comparison are priorities. A CDE can stand alone as a single variable or may be included in a structured collection of elements such as a multi-item scale or index or a complex case report form.

One critical characteristic of CDEs is the use of a defined value set, where, for a question that is designated as a variable for data collection, the permissible responses are restricted to a fixed list. For example, if the variable is current pregnancy status, the fixed value set could be limited to yes or no. If the variable is type of brain tumors that are gliomas of the highest grade, the fixed value set could be, based on the current classifications, glioblastoma multiforme, gliosarcoma, or gliomatosis cerebri.

For some CDEs, precision in defining the method of assessment may be part of the specification. For example, if a CDE for a clinical study is defined as the result of an immunoassay, the CDE may specify the specific way in which the assay is to be conducted. For example, with the enzyme-linked immunospot (ELISPOT) assay to detect either antibody or cytokine secretion, the results of several studies show ELISPOT results vary from laboratory to laboratory but can be harmonized through rigorous training, quality assurance, and quality control measures.

In practice, CDEs are identified by research communities from variable sets currently in use or are newly developed to address a designated data need. CDE development and selection is an iterative process guided by feasibility, utility, and acceptability that benefits from multiple stakeholders including clinicians, informaticists, terminologists, statisticians, patients, and others. CDEs that are specified using standardized vocabularies, codesets, and terminologies can ease the burden of data collection and data exchange and promote discovery and interoperability between systems, including patient registries and electronic health records.

There are no formal international specifications governing the construction or use of CDEs. Consequently, CDEs tend to be made available by research communities on an empiric basis.

What is the value of CDEs?

CDE use has some advantages within a single study if they are perceived and implemented as a standard or specification. CDEs can provide consistency and efficiency in establishing data collection infrastructure and minimize variability in training and implementation. Consequently, the use of CDEs...
can increase the efficiency, quality, clarity, and reproducibility of the overall research process and results.

CDEs can be used to design the logic of data collection and can be embedded in case report forms, patient registries, and integrated into collected and analytic datasets. CDEs can be expressed in machine-readable formats to be used in data analytic plans and structured routines and scripts to incorporate the CDE variables.

Enhanced value of CDEs is across studies to pool and combine data for meta-analyses, modeling, and post hoc construction of synthetic cohorts for exploratory analyses. CDEs can also be a tool to link datasets and examine relationships even if there is not a one-to-one mapping across all data elements in multiple datasets. CDEs can be used to link and aggregate variables across multiple datasets by identifying the CDEs and pulling the associated values into a new hybrid analytic dataset. CDEs can also be used to map associations across datasets. Well-constructed and implemented CDEs increase the precision and can eliminate the errors that come with other methods such as ad hoc transformations, conversion, and manual linking.

CDEs that are used in multiple studies are a tool to leverage the substantial investment made to collect quality data from clinical trials by increasing the consistency of data collection across studies. The use of CDEs, especially when they conform to accepted standards, can facilitate cross-study comparisons, data aggregation, and meta-analyses; simplify training and operations; improve overall efficiency; promote interoperability between different systems; and improve the quality of data collection.

**Sources of CDEs**

CDE are stored in collections, repositories, or libraries. Several are readily available and free although registration may be required for access. Researchers can find pre-defined CDEs already developed and available in sets for use with case report forms (CRFs). The National Institutes of Health (NIH) has an extensive repository of CDE collections, and researchers are encouraged to use them in human subject research, such as clinical studies, registries, or biorepositories. CDEs are developed within many disciplines and it is recommended they be used to establish uniformity when collecting or using clinical data.

NIH has an extensive collection of CDE repositories, which can be found on their Resource portal:

**CDE Repositories**

*General Categories:* These include CDEs used across several research disciplines. Examples from NIH listings include:

- NIH Common Data Element (CDE) Resource Portal - Collections and Efforts
- PhenX Toolkit
- PROMIS - validated patient-reported outcome measures
• NIH Toolbox - Validated measures of cognitive, emotional, sensory and motor functions

**Focused Categories:** These are specific CDEs are those that are less common to all subjects but more specific to research subcategories. Some examples include:

- NINDS Common Data Elements for neurological disorders and stroke clinical research
- NCI - biomarker harmonization
- NEI-AREDS - harmonize data across biorepositories
- NIDA - electronic health record extracted
- NCATS - Global Rare Diseases Patient Registry
- NIH - National Institute on Drug Abuse

**Other CDE Repositories:**
- Health & Human Services HHS.gov Common Data Element Repository (CDER) Library
  - A federal-wide, online searchable repository for grants-specific data standards, definitions, and context.

**Simple examples of CDEs:**

- Patient age
- Gender
- Marital status
- Diagnosis
- Tumor grade
- Number of alcoholic beverages consumed over past 30 days

**Development of CDEs:**

The development process generally includes four basic steps:

1. A need for a CDE or group of CDEs is identified. This may generate from an organization, an individual investigator, sponsor, or funding entity; it may initiate from a regulatory agency such as the FDA, or from a professional society or an entire research community.

2. Stakeholders & expert groups are identified and meet to develop or select CDE for an identified purpose

3. Iterations & updates are made, in initial development and with ongoing input from broader community
4. CDEs are endorsed by the developing body or organization. Their use is then required, recommended, encouraged, or acknowledged as an option for use.

Sensor Informatics

Sensor Informatics (SI) is an emerging technology that is concerned with the extraction of information from data generated by sensors, usually as part of a system designed to solve a larger application problem.

Global healthcare systems are struggling with aging population, prevalence of chronic diseases, and the accompanying rising costs. Also, escalated incidence and costs associated with lifestyle induced poor health (e.g., obesity), and non-communicable diseases such as cancer and cardiovascular diseases are major healthcare challenges globally. In response to these challenges, researchers have been actively seeking for innovative solutions and new technologies that could improve the quality of patient care meanwhile reduce the cost of care through early detection/intervention and more effective disease/patient management. Rather than relying on delayed intervention and expensive treatments, the future of a sustainable global healthcare system is one that is specifically focused on prevention, early detection and minimally invasive management of diseases. Thus, the future healthcare system should be preventive, predictive, preemptive, personalized, pervasive, participatory, patient-centered, and precise, i.e., p-health system (Zheng et al. 2014). Health informatics, which is an emerging interdisciplinary area to advance p-health, mainly deals with the acquisition, transmission, processing, storage, retrieval, and use of different types of health and biomedical information. The two main acquisition technologies of health information are sensing and imaging.
With increasing availability of sensor enriched smart, wearable devices, the knowledge about our own health and wellbeing is also evolving. It is no longer just limited to disease progression or the effect of therapeutic measures provided in clinical settings. Such a trend in sensor informatics has given rise to the big personal data, which is set to influence the future of healthcare. Preventing disease through promotion of healthy lifestyle choice is a potentially cost-effective approach to modern healthcare challenges. Choices such as diet, physical activity, sleep, smoking and alcohol, have all been associated with many medical conditions.

With the future of healthcare is shifting from reactive to preventive medicine, made possible by systems approach to disease through integrated diagnosis, treatment, and prevention, we are more focused on the quality of life and our individual wellbeing. As sensors get smaller, smarter and increasingly pervasive, the possibility of personalised, predictive, preventive and participatory medicine becomes increasingly realistic. Novel sensor informatics can allow the detection of disease at an earlier stage, stratify patient management with optimised and individualised treatment, and directly involve patients in both immediate and long-term continuous monitoring of therapeutic responses. In addition to device level developments, the myriad of data generated continuously in real-time represents one of the major challenges in sensor data analytics.

Advances in sensing hardware have been accelerating in recent years and this trend shows no signs of slowing down (Poon & Zhang 2008). In Table below, we summarize some of the state-of-the-art developments in sensing hardware covering devices used in research (Andreu-Perez, Leff, et al. 2015).
Recent advances in sensing technologies have made it possible to monitor health in an unobtrusive and seamless manner, transforming episodic, largely manual sampling processes to continuous, context-aware monitoring and intelligent intervention. Figure 24 outlines the evolution of allied technologies in the last 10 years. Three factors in particular have contributed to these advances: 1) increased data processing power, 2) faster wireless communications with higher bandwidth, and 3) improved design of microelectronics and sensor devices. The first two represent general trends in computing, whereas the third is of particular interest to pervasive health. Advances in sensor electronics have supported the development of a wide range of embedded systems, as well as devices that are small, lightweight and can be comfortably worn by an individual or ubiquitously placed in the environment with minimal power consumption.

<table>
<thead>
<tr>
<th>Sensor placement</th>
<th>Sensors type</th>
<th>Key technical features</th>
<th>TRL</th>
<th>Clinical focus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Electrodes on conductive fabric/elastic “heart sock”</td>
<td>9</td>
<td>Cardiac arrhythmia [14, 15]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Glucose (Decom®)</td>
<td>9</td>
<td>Diabetes (Decom®)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adenosine triphosphate (ATP)</td>
<td>9</td>
<td>Hemorrhage [16]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Accelerometer (17)</td>
<td>6</td>
<td>Rehabilitation [17]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Galvanic skin response (GSR) (15)</td>
<td>6</td>
<td>Obesity [18]</td>
</tr>
<tr>
<td>Heart</td>
<td></td>
<td>Temperature (18)</td>
<td>6</td>
<td>Infection [18]</td>
</tr>
<tr>
<td></td>
<td>Interscalular pressure (IPS) (19)</td>
<td>IOP by change in corneal curvature</td>
<td>6</td>
<td>Glaucoma [19]</td>
</tr>
<tr>
<td></td>
<td>Glucose (Small G) contact lens</td>
<td>BTDI readout</td>
<td>5</td>
<td>Diabetes (Small G)</td>
</tr>
<tr>
<td>Brain</td>
<td>Impact force (CheckLight™)</td>
<td>Impact logging</td>
<td>9</td>
<td>Concussion (CheckLight™)</td>
</tr>
<tr>
<td></td>
<td>Glucose/Oxygen (Pinnacle™)</td>
<td>Red blood capillary with multichannel photometer</td>
<td>7</td>
<td>Traumatic/hemorrhage (Pinnacle™)</td>
</tr>
<tr>
<td></td>
<td>EEG (NeuroP™)</td>
<td>8 channel EEG</td>
<td>7</td>
<td>Epilepsy (NeuroP™)</td>
</tr>
<tr>
<td>Ear</td>
<td>Acceleration (8, 20)</td>
<td>Three-axis accelerometer behind the ear</td>
<td>6</td>
<td>Clinical gas analysis (8, 20)</td>
</tr>
<tr>
<td></td>
<td>Auditory (21)</td>
<td>Linearized hearing aid</td>
<td>6</td>
<td>Hearing loss [21]</td>
</tr>
<tr>
<td>Tooth</td>
<td>Bacteria (22)</td>
<td>Anti-microbial peptide coated graphite as bacterial sensor</td>
<td>6</td>
<td>Infection [22]</td>
</tr>
<tr>
<td>Wrist/Arm</td>
<td>Activity level/energy expenditure (Nike®)</td>
<td>Custom metric for energy expenditure</td>
<td>9</td>
<td>Obesity (Nike®)</td>
</tr>
<tr>
<td></td>
<td>Skinfourmance (12)</td>
<td>Soft sensor for electromyographic activity</td>
<td>8</td>
<td>Emotional stress [23]</td>
</tr>
<tr>
<td></td>
<td>Accelerometer (20, 22)</td>
<td>Accelerometer on brace</td>
<td>6</td>
<td>Parkinson’s disease [24, 25]</td>
</tr>
<tr>
<td></td>
<td>Gyroscope and Inertimeter (26)</td>
<td>Wrist/Ankle mounted inertial tracker</td>
<td>6</td>
<td>Stroke rehabilitation [26]</td>
</tr>
<tr>
<td></td>
<td>EMG and EKG (35)</td>
<td>Multi-modal (EMG+Inertial) patch</td>
<td>6</td>
<td>Neuroprosthetic (65)</td>
</tr>
<tr>
<td>Foot</td>
<td>Accelerometer (24, 27)</td>
<td>Posture/reactive from heel acceleration and plantar pressure</td>
<td>6</td>
<td>Obesity [26, 27]</td>
</tr>
<tr>
<td></td>
<td>Gyroscope, force, head and pressure, electric field height, air pressure (28)</td>
<td>Glut shoe connector in shoe air pressure for ground contact force</td>
<td>6</td>
<td>Clinical gas analysis (28)</td>
</tr>
<tr>
<td>Hands/Fingers</td>
<td>Blood pressure, SpO2, (Health®)</td>
<td>Wireless network and pressure cuff</td>
<td>9</td>
<td>Hypertension (Health®)</td>
</tr>
<tr>
<td></td>
<td>Accelerometer (19, 30)</td>
<td>Sensor network on glove for hand gesture analysis</td>
<td>6</td>
<td>Surgical training [29, 30]</td>
</tr>
<tr>
<td></td>
<td>BondSens (39)</td>
<td>Pressure sensor network glove measuring range of motion</td>
<td>6</td>
<td>Arthritis [30]</td>
</tr>
<tr>
<td>Hip</td>
<td>Vibration (11)</td>
<td>Hip prosthesis tested with artificial thigh</td>
<td>6</td>
<td>Hip prosthesis [31]</td>
</tr>
<tr>
<td>Implantable/Ingestible wireless sensors/stimulators</td>
<td>pH (132)</td>
<td>pH capsule attached to esophageal wall</td>
<td>9</td>
<td>GERD (132)</td>
</tr>
<tr>
<td></td>
<td>Temperature, H2O evaporation (VidarBee®)</td>
<td>Ingrowth solution for wireless core temperature</td>
<td>9</td>
<td>Infection (VidarBee®)</td>
</tr>
<tr>
<td></td>
<td>Heart rhythm (Evites®)</td>
<td>Implantable defibrillator</td>
<td>9</td>
<td>Cardiovascular disease (Evites®)</td>
</tr>
<tr>
<td></td>
<td>Auditory nerve (Coilbust®)</td>
<td>Auditory nerve stimulation with wireless powering</td>
<td>9</td>
<td>Duodenal (Coilbust®)</td>
</tr>
<tr>
<td></td>
<td>Visible light (Bluesight®)</td>
<td>Retinal ganglion cells (RGC) stimulation</td>
<td>9</td>
<td>Bladder (Bluesight®)</td>
</tr>
<tr>
<td></td>
<td>Brain stimulator (Dolostim®)</td>
<td>Single lead implantable neurostimulator</td>
<td>9</td>
<td>PD, Tremor (Sota®)</td>
</tr>
<tr>
<td></td>
<td>Motoric ingestion (Penrose®)</td>
<td>Implantable pill with wireless interrogation for ingestion signum</td>
<td>6</td>
<td>Tasking ingestion management (Penrose®)</td>
</tr>
<tr>
<td></td>
<td>Force sensor (13)</td>
<td>Battery-less piezoelectric energy harvester knee implant</td>
<td>6</td>
<td>Knee replacement (Penrose®)</td>
</tr>
<tr>
<td></td>
<td>Pressure sensor (Carmel®)</td>
<td>Compliant artificial heart</td>
<td>6</td>
<td>Heart replacement (Carmel®)</td>
</tr>
<tr>
<td>Wearable for ambient environment</td>
<td>Ozone Chlorine, Methane, Carbon monoxide, humidity, temperature (EcoSense®)</td>
<td>Environmental sensing link with smartphones</td>
<td>9</td>
<td>Poisoning (SensorNine®)</td>
</tr>
</tbody>
</table>
According to the analysis in the BI Intelligence report (Garner) published at the end of 2014, the price of one MEMS sensor has decreased by half during the last. This has partly driven a paradigm shift of future Internet applications toward what is termed “the Internet of Things” (IoT). The Internet of Things (IoT) is a new concept, providing the possibility of healthcare monitoring using wearable devices. The IoT is defined as the network of physical objects, which are supported by embedded technology for data communication and sensors to interact with both internal and external objects states and the environment (Haghi et al. 2017). Moreover, enabling technologies ranging from nano- and microelectronics, advanced materials, wearable/mobile computing, and telecommunication systems, as well as remote sensing and geographic information systems have made it possible for sensing health information to be collected pervasively and unobtrusively as illustrated in Figure 25 (Andreu-Perez, Poon, et al. 2015).
Figure 25. Big sensing data in health are all around us, enabled by technologies ranging from nano- and microelectronics, advanced materials, wearable/mobile computing, and telecommunication systems as well as remote sensing and geographic information systems. The inner loop presents technologies for sensor components, while the middle loop presents devices and systems potentially owned by each individual or household. The outer loop presents sensing technologies required at the community and public health level.

Wearable, implantable, and ambient sensors

In the last decade, wearable devices have attracted much attention from the academic community and industry and have recently become very popular. The most relevant definition of wearable electronics is the following: “devices that can be worn or mated with human skin to continuously and closely monitor an individual’s activities, without interrupting or limiting the user’s motions” (Gao et al. 2016).

Three factors have contributed to the rapid uptake of wearable devices (Andreu-Perez, Leff, et al. 2015):

1. Increased data processing power,
2. Faster wireless communications with higher bandwidth, and
3. Improved design of microelectronics and sensor devices.

Advances in sensor electronics have supported the development of a wide range of embedded systems, as well as devices that are small, lightweight and can be comfortably worn by an individual or ubiquitously placed in the environment with minimal power consumption. Thus far, wearable devices are widely used to measure key health indicators such as electrocardiogram (ECG), heart rate, blood pressure, blood oxygen saturation (SpO2), body temperature, postures and physical activities (Andreu-Perez, Leff, et al. 2015).
Today, the range of wearable systems, including micro-sensors seamlessly integrated into textiles, consumer electronics embedded in fashionable clothes, computerized watches, belt-worn personal computers (PCs) with a head mounted display, glasses, which are worn on various parts of the body are designed for broadband operation. The field of wearable health monitoring systems is moving toward minimizing the size of wearable devices, measuring more vital signs, and sending secure and reliable data through smartphone technology. Although there has been an interest in observing comprehensive bio/non-bio medical data for the full monitoring of environmental, fitness, and medical data recently, but one obvious application of wearable systems is the monitoring of physiological parameters in the mobile environment. The majority of commercially available wearable devices are one-lead applications to monitor vital signs. However, most of such recreational devices are not suitable for the medical monitoring of high-risk patients (Haghi et al. 2017).

Example platforms include earlier systems with limited connectivity and single sensing elements developed solely for use in research laboratories to more recent ambient sensors as well as easy-to-wear wearable/implantable devices equipped with continuous multi-modal sensing capabilities and support for data fusion deployed in a wide range of clinical applications (Zheng et al. 2014). Furthermore, parallel developments in miniaturized sensor embodiment, microelectronics and fabrication processes, and the availability of wireless power delivery have made miniaturized implantable sensors increasingly versatile.

Implantable sensors address the challenges of both acute and chronic disease monitoring by providing a means of capturing critical events and continuous streamlining of health information. Recent advances in microelectronics and nanotechnology have greatly improved the sensitivity of different sensors. For example, based on metal nanoparticle arrays and single nanoparticles, the sensitivity of localized surface plasmon resonance optical sensors can be pushed toward the detection limit of a single molecule. This has enabled the development of the next generation of high-throughput sequencing technologies, as well as the detection of biomolecules, such as glucose, lactate, nitric oxide, and sodium ions. For diabetic patients, a myriad of new sensors for both wearable and implantable applications have been developed, which provide continuous monitoring and corresponding response to the time-varying glucose level, which is well known to be diet dependent.

There is a clear trend of moving from the scenario where a centralized large computing infrastructure is shared between multiple users toward one where each individual possesses multiple smart devices, most of which are sufficiently small to be wearable or implantable such that the use of these sensing devices will not affect normal daily activities. These sensor systems have the potential to generate datasets, which are currently beyond our capabilities to easily organize and interpret. Meanwhile, healthcare services delivered via ambient intelligence consisting of ambient sensors and objects interconnected into an integrated IoT represent a promising and supportive solution for the ageing society. It is important that such systems should take into account the sensor, service, and system integration architecture. Such distributed systems require decentralized inference algorithms, which are frequency explored, either in the framework of parametric models, in which the statistics of phenomena under observation are assumed to be known by the system designer, or nonparametric models, when the underlying data is sparse and prior knowledge is limited (Andreu-Perez, Poon, et al. 2015).
Figure 26. Four popular motion tracker wearable devices.

Figure 27. Unobtrusive wearable devices for various physiological measurement developed by different groups: watch-type BP device, PPG sensors mounted on eyeglasses, motion assessment with sensors mounted on shoes, wireless ECG necklace for ambulatory cardiac monitoring (courtesy of IMEC, Netherlands), h-Shirt for BP and cardiac measurements, ear-worn activity recognition sensor, glove-type pulse oximeter, strain sensors mounted on stocking for motion monitoring, and ring-type device for pulse rate and SpO2 measurement. (Zheng et al. 2014)
From sensor data to stratified patient management

Physiological sensing by these smart devices can be long term and continuous, imposing new challenges for interpreting their clinical relevance. For example, the current clinical practice defines hypertension based on measurements taken during infrequent hospital visits. Although automated oscillometric BP measurement devices are now available, studies in these areas are often limited to taking BP once every hour over a 24-hr period. With the newly emerging ambulatory devices, a comprehensive BP-related profile of an individual can be made available. Nevertheless, the interpretation of these data is non-trivial, since in many situations, they may not be equivalent to the clinical BP readings that are currently being used by practitioners. The signals, however, carry underlying physiological meanings that, if properly processed and managed, can be used as additional information for understanding uncontrolled hypertension or to enhance the current hypertension management schemes. In addition to vital sign monitoring, smart implantable sensors provide a promising technology to monitor postoperative complications, such as slow tissue healing and infections. Moreover, smart implants can also have a reactive role by delivering drugs for chronic pain and acting as brain stimulators for neurological diseases including refractory epilepsy and Parkinson’s disease. This makes smart implants not just another resource for data collection but also an integral part of early intervention.

With increased volume and acquisition speed of data from both wearable and implantable sources, new automated algorithms are needed to reduce false alarms such that they are sufficiently robust to support large-scale deployment, particularly for free-living environments. Automatic classifications are necessary since the dataset sizes are beyond the capability of manual interpretation within a reasonable time period. New compression-based measures are, therefore, proposed as high-quality cloud computing services to reduce the computation time for the automated classification of different types of cardiac arrhythmia. In many situations, measurements must be interpreted together with the context under which the data is collected. For example, many physiological parameters, such as BP or episodes of gastroesophageal reflux disease are posture dependent, which can be captured by inertial sensors. Therefore, multimodal integration and context awareness are essential to the analysis of pervasive sensing data. (Andreu-Perez, Poon, et al. 2015)

Mobile health

Mobile health technology (mHealth), which is the applying mobile communication technology to healthcare and patient wellness, has become increasingly popular. Nowadays, smart phones have become an inseparable companion for about 3.2 billion unique mobile users worldwide with over 30,000 available healthcare apps (Ross et al. 2014). Mobile health (m-Health) proposes to deliver healthcare services, surpassing geographical, temporal, and even organizational barriers. M-Health solutions address emerging problems on health services, including, the increasing number of chronic diseases related to lifestyle, high costs of existing national health services, the need to empower patients and families to self-care and handle their own healthcare, and the need to provide direct access to health services, regardless of time and place (Silva et al. 2015).

M-Health uses sensors, global positioning satellite receivers, and accelerometers that continuously monitor data. Once risk factors are determined for various conditions, the hope is to
integrate these factors into the EHR to assist physicians in identifying at-risk patients and build predictive models (Ross et al. 2014). Typical m-Health services architectures (presented in Figure 28) use the Internet and Web services to provide an authentic pervasive interaction among doctors and patients. A physician or a patient can easily access the same medical record anytime and anywhere through his/her personal computer, tablet, or smartphone.

M-Health services are even becoming popular in developing countries where healthcare facilities are frequently remote and inaccessible. Mobile applications for healthcare systems are rapidly growing and evolving. Research interest in this topic is expanding every day, as well as the diversification of the impact areas (Silva et al. 2015).

Due to the ongoing developments of network, mobile computing and computer storage, the storage and retrieval of big health data have attracted great attentions. With the increasing use of various kinds of long-term monitoring devices, there is an urgent demand for the intelligent management of big health data. These data could be integrated into the Personalized Health Record or EHRs, and would contribute to building an automated system to identify at-risk populations and send automated health messages to patients (Ross et al. 2014). A large number of cloud-based storage and retrieval systems are emerging in recent years (Zheng et al. 2014).

Many mobile applications are available for translating personal health and fitness signals into big data (e.g. The Apple ResearchKit). Add-on technologies and so-called “wearables” are also empowering the consumer with mobile diagnostics that can measure blood pressure level, monitor heart rate, identify cervical cancer, and even perform an eye exam (Taglang & Jackson 2016). Many sensors or wearable devices have a companion mobile application that uses Bluetooth, Zigbee, infrared waves, ultra band wireless communication or a USB-based sync service to update health-monitoring data from the wearable monitor to a connected computer or data aggregation database. As part of the mobile Health initiative (http://www.hhs.gov/open/initiatives/mhealth/), mobile applications have

\[ Figure 28. \text{Illustration of a typical architecture of m-Health services.} \]
been designed that provide software modules to harness the internal sensors in mobile phones for the capture of health data (Shameer et al. 2017).

The new generation of smartphones has a wide range of health apps with standardized protocol to connect to sensors provided by different companies. They can potentially serve as a platform to centralize health data, from which additional new information that was previously untraceable by individual sensors can now be mined. In fact, newer models of mobile phones are packed with sophisticated sensors that facilitate the extraction of different types of vital signs, even without the need for external devices. These sensors can provide valuable health information for the management of many long-term illnesses. For example, the video cameras of mobile phones can be used to collect heart rate and heart rate variability, embedded accelerometers and gyroscopes to track energy expenditure. Furthermore, the pulse transmission time as measured by time delays between electrocardiographic and photoplethysmographic sensors can be used as a surrogate measure for BP. This information can be calculated from two devices that connect with a mobile phone independently, one with an electrocardiographic sensor and the other one with a photoplethysmographic sensor. When connected to health providers, a closer level of interaction in healthcare can be maintained toward greater personalization and responsiveness (Andreu-Perez, Poon, et al. 2015).

The effective management of this huge inflow of mobile-generated data calls for the implementation of a big data solution. Therefore, healthcare organizations leverage big data solutions to manage all of the health information, improve care and increase access to healthcare. National Institutes of Health (NIH), awarded $10.8 million grant to the 11 national big data centers to develop innovative approaches in order to collect, analyze and manage health data generated by mobile and wearable sensors. The Mobile Sensor Data-to-Knowledge (MD2K) center is a part of NIH Big Data to Knowledge (BD2k), which aims to support advances in research, policy, and training needed for the effective use of big data in healthcare informatics [MD2K 2015] (Fang et al. 2016).

**Imaging Informatics**

Imaging MS is a new technology for direct mapping and imaging of biomolecules present in tissue sections. Imaging MS shows potential for several applications, including biomarker discovery, biomarker tissue localization, understanding of the molecular complexities of tumor cells, and intraoperative assessment of surgical margins of tumors.

Imaging informatics, also referred to as radiology informatics or medical imaging informatics, is the discipline that stands at the intersection of biomedical informatics and imaging, bridging the two areas to further our comprehension of disease processes through the unique lens of imaging, and improve clinical care. Imaging informatics addresses not only the images themselves, but encompasses the associated data to understand the context of the imaging study, to document observations and to correlate and reach new conclusions about a disease and the course of a medical problem.

Imaging informatics aims to improve the efficiency, accuracy, usability and reliability of medical imaging services within the healthcare enterprise. Imaging informatics concerns how information about and contained within medical images is retrieved, analyzed, enhanced, and exchanged throughout complex healthcare systems. When clinicians have immediate electronic access
to medical images, precious time is saved, allowing for timely medical decisions, reducing unnecessary repetition of exams and driving costs down.

Molecular imaging enables visualization of cellular and molecular processes that may be used to infer information about the genomic and proteomic profiles. As a result, the bioinformatic analysis of genomic and proteomic profiles may be valuable to assist the interpretation of images using molecular probes. Molecular diagnostics and molecular imaging can provide the two aspects of the disease: molecular diagnostics can provide the information of the exact mutation of a particular gene and classify the exact type of cancer, while molecular imaging can target the very same type of cells with that particular mutation in order to provide diagnostic information and disease staging (Schaffer et al. 2006).

The ever-increasing amount of annotated and real-time medical imaging data has raised the question of organizing, mining, and knowledge harvesting from large-scale medical imaging datasets. While established imaging modalities are getting pervasive, new imaging modalities are also emerging. These modalities are rapidly filling up the entire EM spectrum as shown in Figure 29. Many of these imaging techniques are now geared toward real-time in situ or in vivo applications, making multimodality imaging an exciting yet challenging big data management problem.

Recent developments in imaging are progressing in multiple frontiers. First, there is relentless effort in making existing imaging modalities faster, higher resolution, and more versatile. Take cardiovascular magnetic resonance imaging (MRI) as an example, imaging sequences are no longer limited to morphological and simple tissue characterization (e.g., via T1, T2/T2* relaxation times). Details concerning vessel walls, myocardial perfusion and diffusion, and complex flow patterns in vivo can all be captured. When facilitated with new minimally invasive interventional techniques, novel drugs and other forms of treatment, MRI now serves as a therapeutic and interventional aid, rather than solely a diagnostic modality. Similar advances can also be appreciated for ultrasound, computed tomography (CT), and other imaging modalities. Moreover, extensive efforts in combining different imaging modalities, not by postprocessing, but at the hardware level, e.g., MRI/PET and PET/CT, open up a range of new opportunities, particularly for oncological imaging and targeted therapy (Andreu-Perez, Poon, et al. 2015).

![Figure 29. Different imaging modalities across the electromagnetic spectrum.](image-url)
Imaging across scales

There have been extensive research efforts for developing new technologies that probe deeper into the biological system, from tissue (up to micrometer) to the protein level (micronanometer). In particular, recent advances in stimulated emission depletion fluorescence microscopy allow the generation of 3-D superresolution images of living biological specimens. It overcomes the classical optical resolution limit of light microscopy and pushes the spatial resolution of optical microscope toward the nanoscale. This opens up the possibility of imaging not only the fine morphological structure of many organ systems (e.g., microfibrils that form blood vessels), but also subcellular behavior and molecular signaling. The use of quantum dots or qdots also pushes the boundary of imaging resolution, allowing the study of intracellular processes at molecular levels (20-40 nm). Another class of fluorescent labels is made by conjugating qdots with biorecognition molecules, which emission wavelength can be tuned by changing the particle size such that a single light source can be used for simultaneous excitation of all different-sized dots. These technologies have already been used for immunofluorescence labeling of tissues, fixed cells, and membrane proteins, such as cancer markers, the hybridization of chromosomes, the labeling of DNA, and contrast-enhanced image-guided resection of tumors (Andreu-Perez, Poon, et al. 2015).

From morphology to function

The understanding of many biological processes requires the identification and representation of structure–function relationships. This expands across different spatial scales, namely proteins, cells, tissues, and organs. For instance, haemodynamic analysis combined with contractile analysis, substantiated with myocardial perfusion data, can be used to elucidate the underlying factors associated with cardiac abnormalities. Starting with modeling, the tissue and scaling up toward a more specific description of organ behaviors has made it possible to create integrative models of heart function. These architectural models fuse information such as fibrous-sheet geometrical models of tissue and membrane currents from ion channels at the subcellular level.

Amongst all organs that have been studied to define their function from its morphology, the brain is the one that has received the most attention recently. This is motivated by the fact that brain structure and function are keys to understand cognitive processes, hence the need for unveiling neuronal behavior from the molecular level up to the functioning of neural circuits. Super-resolution fluorescence microscopy has been applied to study neural morphology and their subcellular structures. These techniques may enable us to achieve a resolution as high as 20 nm. Needless to say, the myriad of markers necessary for each single type of cell and synapse would result in an enormous database.

Methods, such as functional MRI (fMRI) and functional diffusion tension imaging provide flexible information in the form of macrostructural, microstructural, and dense connectivity matrices. Improved fMRI sampling methods produce time-series data of multiple blood oxygenation-level-dependent volumes of the brain. In addition, there is an increasing trend in making neuroimaging multimodal. In some studies, several modalities are used to compensate the benefits and tradeoff of one another. Furthermore, information from lower cost and rapid noninvasive methods, such as wearable electroencephalography and functional near-infrared spectroscopy allows gathering brain functional data for examining cortical responses due to more complex tasks.
An indirect way of inferring function consists of a combination of imaging modalities as well as medical records, demographics, and lab test results. In order to maximize the information contained in these heterogenous sources, linking different metadata with features extracted from image modalities is key to characterize the structure, function, and progression of diseases. Solving this challenge presents a unique opportunity for bridging the semantic gap between images and more effective prediction, diagnosis, and treatment of diseases. However, this issue entails many independent yet interrelated tasks, such as generating, segmenting, and extracting enormous amounts of quantifiable spatial objects and features (nuclei, tissue regions, blood vessels, etc.). This requires the implementation of effective and optimized querying systems in order to reduce the computational complexity of handling these data. Fig. 9 represents a schema of what big data means for imaging, as defined by both structural and functional data.

![Figure 30. Processing schema of imaging toward big data. Nonfunctional medical imaging is acquired and processed to serve as a model to register organ activity in the resulting functional imaging. Results from image processing and functional imaging are stored in databases with specific metadata protocols. Large-scale big data analysis is performed in these databases linking then the features extracted through medical imaging processing and functional imaging.](image)

Existing efforts in improving the spatiotemporal constraints of brain imaging are significantly increasing the computational resources needed for neuroimaging studies. RAM memory is an important resource for neuroimaging analysis. For instance, to perform subject-, voxel- and trial-level analysis, a significant amount of fMRI images needs to be loaded into memory (Andreu-Perez, Poon, et al. 2015).

**Research initiatives to understand the human brain**

Another active topic in imaging is to study the functional connectivity of the human brain, which is fundamental to both basic and applied neurobiological research. Both U.S. and European Union (EU) have launched large-scale Human Brain projects in recent years with an aim to unravel
the organ’s complexity. The NIH-funded Human Connectome Project (HCP) aims at leveraging the latest advances in DTI to study brain areas in relation to their functional, structural, and electrophysiological connectivity. The idea behind the HCP is that neural connectivity is as unique as the fingerprint to each individual. Genetics, environmental influences, and life experience are factors contributing to the formation of each individual’s neural circuitry. This is supported by genome-wide association studies that link genetic variants with neurological and psychiatric disorders that have abnormal brain connectivity, e.g., variants at human clusterin (CLU) on chromosome 8 and complement receptor 1 on chromosome 1 are associated with Alzheimer’s disease as well as those specific markers associated with schizophrenia and dementia.

Recently, the EU commission supported financially the human brain project, which aims to develop a biological model of the brain that simulates different aspects of the nervous system, including point neuron models, neural circuitry, and cellular models at different scales. The main idea is to provide a simulation platform for theoretical neuroscientists to study how the brain processes information. For this purpose, it would require simulating all functions, architecture, and chemical properties for the 86 billion neurons and trillions of synapses of the human brain as estimated by Azevedo et al.. Therefore, the aim of the project is both ambitious and controversial (Andreu-Perez, Poon, et al. 2015).
References


Denny, J.C., 2014. Surveying Recent Themes in Translational Bioinformatics: Big Data in


