Journal of Biomedical Informatics 43 (2010) 104–110

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

What is biomedical informatics?

Elmer V. Bernstam [a,b,*], Jack W. Smith [a], Todd R. Johnson [a]

[a] School of Health Information Sciences, The University of Texas Health Science Center at Houston, Houston, TX, USA
[b] Division of General Internal Medicine, Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA

ARTICLE INFO

Article history:
Received 25 February 2009
Available online 13 August 2009

Keywords:
Biomedical informatics
Scientific discipline
Data
Information
Knowledge
Definition
Philosophy of information

ABSTRACT

Biomedical informatics lacks a clear and theoretically-grounded definition. Many proposed definitions focus on data, information, and knowledge, but do not provide an adequate definition of these terms. Leveraging insights from the philosophy of information, we define informatics as the science of information, where information is data plus meaning. Biomedical informatics is the science of information as applied to or studied in the context of biomedicine. Defining the object of study of informatics as data plus meaning clearly distinguishes the field from related fields, such as computer science, statistics and biomedicine, which have different objects of study. The emphasis on data plus meaning also suggests that biomedical informatics problems tend to be difficult when they deal with concepts that are hard to capture using formal, computational definitions. In other words, problems where meaning must be considered are more difficult than problems where manipulating data without regard for meaning is sufficient. Furthermore, the definition implies that informatics research, teaching, and service should focus on biomedical information as data plus meaning rather than only computer applications in biomedicine.

1. Introduction

Biomedical informatics has been an "emerging field" for decades. Concern about medical information and the desire to computerize health care are hardly new. Though originally focused on traditional paper-based medical records and their management rather than electronic medical records, the American Health Information Management Association (AHIMA) was founded in 1928 as the American Association of Medical Record Librarians [1]. Papers about medical reasoning were published in the 1950's [2]. Kaiser Permanente established a department of medical methods research in September of 1961; one of its goals was to "begin to use computers in the practice of medicine" [3]. In 1962, they obtained their first federal grants to automate and improve screening methods [4]. Recent developments have thrust informatics into the national spotlight as part of a massive economic stimulus package known as the American Recovery and Reinvestment Act.

Yet there is still no universally accepted definition of medical, health, bio- or biomedical informatics. Often, any activity that relates to computing is labeled "informatics" [5,6]. There is even some debate regarding the desirability of a definition since any meaningful definition has the potential to exclude good work [5] or restrict the use of informatics as a marketing term. We emphasize that a definition is not a value judgment. By defining informatics we are not claiming that informatics is better or worse than other fields. In order for there to be a field of informatics, there must be definable activities that are not informatics.

Academic informaticians, on the other hand, recognize that a compelling theoretically-grounded definition of informatics as a science is desirable [7]. In addition to our desire to define our academic field, a definition can help the field address practical issues, such as:

- *Educational program design:* provide a clear vision of our field to students, guide curriculum development and evaluation within training programs
- *Administrative decisions:* make a clear and consistent case for resources to administrators, to guide informatics units (academic and service-oriented) with respect to hiring faculty or staff, relationship to other organizational units and performance metrics
- *Communication:* including internal communication among informaticians and external communication with those outside of our field; a definition can help match current and potential collaborators, guide informatics societies such as the American and International Medical Informatics Associations (AMIA and IMIA, respectively), and help funding agencies and members of the general public understand our role and contributions
- *Research agenda:* provide a basis for identifying fundamental research questions, and to distinguish basic research in informatics from applied work

* Corresponding author. Address: School of Health Information Sciences, The University of Texas Health Science Center at Houston, 7000 Fannin Street, Suite 600, Houston, TX 77030, USA. Fax: +1 713 500 3929.
E-mail address: Elmer.V.Bernstam@uth.tmc.edu (E.V. Bernstam).

1532-0464/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2009.08.006

Still, articulating such a definition of our field has proven difficult. In this paper, we review the literature regarding definitions of informatics and propose a definition of informatics as a science that is grounded in theory. We then consider a number of important implications of this definition that begin to address some longstanding issues within the field.

2. Background

The "quest" for a definition of biomedical informatics and related concepts such as medical informatics, bioinformatics, clinical informatics and others is not new. Although, compiling an exhaustive list of definitions is not practical, it may be useful to consider categories of definitions modified and expanded from [8] and [9]. Although originally applied to definitions of nursing informatics, these categories are applicable to other areas [10] and the more general field of biomedical informatics. For each category, we briefly define the category, cite examples and discuss its advantages and limitations.

Information technology-oriented definitions focus on technologies and tools as being the defining property of informatics. These definitions usually emphasize computer-based technologies. Terms such as "clinical computing", "computers in medicine" and "medical computer science" are often used as definitions of informatics [7]. Similarly, Berman [11] defines biomedical informatics as "the branch of medicine that combines biology with computer science". Clearly, computers are very important tools for biomedical informaticians. Many activities associated with biomedical informatics such as data mining or electronic medical records would not be meaningful without computers. However, by focusing on computers, technology-based definitions emphasize the tools rather than the work itself [7]. A commonly cited simile is that referring to biomedical informatics as "computers in medicine" is like defining cardiology as "stethoscopes in medicine".

There are at least two unfortunate consequences of focusing on computer technology. First, emphasizing computers encourages us to insert computers whenever possible to solve problems in biomedicine. However, the question should not be: "how do we computerize health care". Indeed, recent studies show that computerizing health care does not necessarily improve outcomes [12,13]. The focus should remain on improving health care, rather than computerizing it.

Second, such definitions generally do not capture important informatics work that does not rely on computers (or computer science). For example, the study of information flow in clinical environments does not necessarily involve computers. Rather, it can focus on interruptions [14], errors [15] or how information is presented to the user [16]. Similarly, computerizing health care requires understanding culture, processes and workflow; indeed a great deal of work in this area has been done and published in informatics journals and/or widely cited in the informatics literature. Lorenzi listed change management among the four cornerstones of medical informatics [17]. Diane Forsythe's work on the influence of culture on information systems resulted in a prize named for the late Dr. Forsythe presented by AMIA [18].

Role, task or domain-oriented definitions focus on the roles of informaticians within organizations. For example, nursing informatics emphasizes the role of informatics – trained nurse specialists in supporting nursing practice and their grounding in nursing science: a specialty that integrates nursing science, computer science, and information science in identifying, collecting and processing, and managing information to support nursing practice, administration, education, and research and to expand nursing knowledge [19].

Role, task or domain-based definitions such as nursing or medical informatics imply that informatics projects are applicable only to the group included in their name (e.g., only applying to nurses, the domain of nursing or the tasks of nurses). Further, they imply that the techniques developed by informaticians are embedded in the "role, task or domain" where they were developed. There are multiple examples to the contrary. For example Protégé, developed at Stanford Medical Informatics, has been used for a wide variety of applications including ventilator management and elevator configuration [20].

Concept-oriented definitions focus on concepts such as data, information and knowledge. For example, Coiera [21] defines health informatics as "the study of information and communication systems in healthcare". Musen focuses on ontologies and problem solving methods as tools for organizing human knowledge and are therefore fundamental to biomedical informatics [7]. Such definitions focus on more fundamental concepts rather than tools, but often fail to provide definitions of those concepts that are sufficiently detailed or operationalized to provide a theoretical foundation for informatics as a science.

The following is a selected list of definitions including several authoritative textbooks:

- Greenes and Shortliffe [22] defined medical informatics as "the field that concerns itself with the cognitive, information processing, and communication tasks of medical practice, education, and research, including the information science and the technology to support these tasks". (task and domain-based)
- Shortliffe and Blois [23] define "biomedical informatics as the scientific field that deals with biomedical information, data and knowledge – their storage, retrieval and optimal use for problem solving and decision making". (Concept-based)
- Van Bemmel [24] writes that medical informatics "...comprises the theoretical and practical aspects of information processing and communication, based on knowledge and experience derived from processes in medicine and health care". (task and domain-based)
- Musen and van Bemmel [25] write that "[i]n medical informatics we develop and assess methods and systems for the acquisition, processing, and interpretation of patient data with the help of knowledge that is obtained in scientific research". (role, task and domain-based)

3. Formulating a definition of informatics based on data, information and knowledge

Despite the lack of agreement, most definitions, regardless of their category, focus on data, information and knowledge as central objects of study in informatics. However, there are no consistent definitions for data, information, and knowledge. Thus, these terms are often used interchangeably. Since data, information and knowledge are central to informatics, precisely defining them is a good starting point for an operational definition of the science of informatics.

A review of the literature on data, information, and knowledge revealed two main schools of thought: Ackoff's Data, Information, Knowledge, Wisdom (DIKW) hierarchy [26], and a related, but more precise set of definitions from philosophy (Table 1). In Ackoff's hierarchy, data are symbols. Information is data that have been processed to be useful. For example, to answer "who", "what", "when", or "where" questions. Knowledge is the application of data and information to answer "how" questions. Understanding is the appreciation of why, and wisdom is evaluated understanding. Since Ackoff first proposed the DIKW hierarchy, many have tried to clarify the meanings of the terms and their relationships. However, a review of recent textbooks describing the DIKW hierarchy found a lack of consensus with the only constant

Table 1
Alternative views of data, information and knowledge.

	Ackoff's hierarchy	Philosophy of information
Data	Symbols	Lack of uniformity
Information	Data that have been processed to be useful (answers to: who, what, when or where questions)	Data + meaning
Knowledge	Application of data and information to answer "how" questions	Justified, true belief
Wisdom	Understanding and appreciation of "why" questions	

being that knowledge is something more than information, and information is something more than data [27].

In contrast to the DIKW hierarchy, philosophers who study information have developed more precise, operational definitions of data, information, and knowledge. Although they have not yet reached consensus and issues remain to be clarified, these definitions are relatively precise and provide a useful starting point. To philosophers of information, a datum is simply a lack of uniformity, information is meaningful data, and knowledge is information that is true, justified, and believed [28].

As an example of how the philosophical definitions of data, information and knowledge can be applied, consider a mother who checks her toddler's temperature with a tympanic thermometer. She sees 102.1 on the display. The symbols "102.1" are data: a lack of uniformity on what would otherwise be a uniform surface (the thermometer display). The mother interprets these data as meaning that the baby has a temperature of 102.1 degrees Fahrenheit. This is now information (i.e., the symbols "102.1" refer to the baby's temperature). The mother next notes that since 102.1 degrees is higher than 98.6, the toddler has a fever. The difference between the normal body temperature and the toddler's is also a data item (or datum), whereas the resulting interpretation of this difference as fever is information.

We can only say that the mother "knows" the baby has fever, if that information is true and the mother has a justification (or understanding) of why it is true. In philosophy what counts as adequate justification is an open question [29]. Normal body temperature varies and the accuracy of tympanic thermometers is ±.5 degrees at best. Thus, the mother can never be absolutely certain that her toddler has a fever. Given a looser interpretation of what counts as an adequate explanation, if the toddler feels hot to the touch (another datum) and the mother takes one more confirmatory reading then there is sufficient justification for "knowing" that the toddler has a fever.

In informatics, we often use knowledge in a related, but slightly different sense: as general information believed to be justifiably true. For example, we record temperatures because we believe, on the basis of prior experience with many individuals over time, that deviations from the normal range may be dangerous. For example, very high or low temperatures may be indicative of an infection that can kill if not properly treated.

These definitions produce a natural hierarchy: there will always be more data than information, and more information than knowledge. Indeed, a significant amount of the information that we use and produce every day is not knowledge, either because it has no truth value (such as instructions like "Close the door on your way out"), or we cannot adequately justify why it is true.

In the above definitions, we have defined information using the terms "data" and "meaning". However, it also possible, and sometimes more convenient, to refer to data as the syntactic part of information and meaning as the semantic part. Syntax refers to the systematic arrangement of data in a representational system or language. Often a datum by itself does not have any meaning unless it is combined with other data according to an accepted syntax. For instance, a black dot on a white page may not mean anything. However, if that dot appears between two numbers, such as "5.2", the dot tells us that this is a decimal numeral and which parts of the numeral are fractional and which are integral.

The data part of a representational system may also be called its "form", in which case meaning is called its' "content". The use of the word "form" is important because of its relationship to formal methods, which are essentially methods that manipulate form using systematic rules that are dependent only on form, not content (meaning). Some symbols or inferences are meaningful. However, this is not captured in the formal rules of symbol manipulation. Formal methods, including computer programs, depend only on systematic manipulation of form without regard for meaning. Thus, ensuring that input to and output from formal methods correctly capture and preserve meaning remains essentially human.

For example, modus ponens:

If P then Q
P
Therefore Q

does not depend on the meaning of P or Q. If P denotes the character string "birds fly" and Q denotes the character string "cows fly" then modus ponens tells us that we can write the character string (i.e., we can logically conclude) "cows fly". This statement is just as legitimate a logical statement as "If xxqqyy then ppzz; xxqqyy; Therefore ppzz". Thus, the statements above are formally correct, but meaningless. To summarize, information can be identically defined as data + meaning, syntax + semantics, or form + content.

4. Definition of informatics

We propose that informatics is the science of information, where information is defined as data with meaning. Biomedical informatics is the science of information applied to, or studied in the context of biomedicine. Some, but not all of this information is also knowledge.

Informaticians study information (data + meaning, in contrast to focusing exclusively on data), its' usage, and effects. Thus, practitioners must understand the context or domain, in addition to abstract properties of information and its' representation.

The definition of information as data + meaning, immediately identifies a fundamental challenge of informatics: how to help human beings store, retrieve, discover, and process information, when our tools (information technology) are largely limited to manipulating data and have only rudimentary information processing capabilities. In other words, the fundamental challenges in informatics result from the difficulties of automating the processing of meaning using tools that actually process data. Since all knowledge is also information, manipulating knowledge using currently available tools is also difficult.

The gap between human information needs and the capabilities of our information technology is at the heart of informatics. Human beings are best at constructing and processing meaning; whereas computers are best at processing data. Although formal methods such as algebra and logic are very useful, they do not manipulate meaning. Compared to computers, human beings are slow and error prone at formal manipulation of data. In contrast, computers are much faster and more accurate when processing data, but have only a rudimentary ability to process meaning. Difficult problems in informatics often involve trying to get computers to process meaning, or at least to appear "as if" they are processing meaning. Although this gap presents a problem, it also means that human beings and computers are naturally complementary.

To better illustrate the fundamental differences between data processors and meaning processors – between computers and human beings – we need only examine some basic results from cognitive psychology. The first general result is that human beings tend to remember the meaning of a sentence or picture instead of its exact form [30–33]. Experimental subjects tend to classify sentences with the same or similar meaning as being identical, ignoring wording differences (syntactic forms). For instance, given the sentence "The doctor diagnosed the patient with pneumonia", participants are more likely to make errors when later presented with sentences like "The doctor decided the patient had pneumonia", or "The patient was diagnosed with pneumonia", than when they are given "The doctor diagnosed the patient with a brain tumor", even though the latter is syntactically (but not semantically) more similar to the original sentence. This is exactly the opposite of computers, which excel at storing and matching exact syntactic forms, but require considerable programming to have even a rudimentary ability to equate different forms with the same meaning. Similarly, recent experiments in ecological psychology have shown that many of the psychological biases found in classic studies of human reasoning and decision making can be greatly reduced or eliminated when human beings are given meaningful problems that relate to their real-world experience [34–36].

5. Discussion

Earlier we indicated that a clear definition of informatics will help the field address practical issues, including educational program design, administrative decisions, communication, and to develop a research agenda. The definition we proposed above does not, by itself, resolve these issues. However, it does offer a perspective on informatics that has significant implications for the field that can help us to address these issues. In this section we discuss several of these implications.

5.1. Implication #1: defining informatics as the study of data + meaning clearly distinguishes informatics from important related fields

Defining the central object of study of informatics as data + meaning allows us to distinguish informatics as a science from computer science, mathematics, statistics, the biomedical sciences and other related fields. It also clarifies the role of each of these fields in informatics.

Computer science is primarily the study of computation. Computer scientists seek to provide solutions to general problems by classifying computational problems in terms of formal abstract properties and deriving effective, efficient algorithms (sequences of syntactic rules) for solving them. For instance, computer scientists talk about network traversal problems and algorithms for traversing networks. What is meant by networks in this context are not the myriad real-world objects we might think of as networks but the formal mathematical objects categorized as networks. The meaning of the data being manipulated by an algorithm is not important. An algorithm to find the shortest path connecting two nodes in a network depends only on the length of the edges, not whether the edges and nodes represent a geographical map, computer network, or social network.

On the other hand, computer science plays an important role in informatics. There can be no information without data, and computers are the best medium we have for reliably storing, transmitting, and manipulating data. Thus, some informaticians develop methods that allow computers to process data "as if" the computer understands the meaning; and to produce tools that allow human beings to make more sense of data displayed by the computer, thereby turning it into information. Information retrieval and formal ontologies are examples of research on the former; whereas work on data visualization and exploratory data analysis are examples of the latter.

Within computer science, the field of artificial intelligence (AI) deserves particular attention in regard to the issues of representation and meaning. There are a variety of definitions of AI and considerable controversy regarding its scope, achievements and appropriate goals for the discipline. John McCarthy, one of the founders of AI, defined the field as "the science and engineering of making intelligent machines, especially intelligent computer programs" [37]. He goes on to define intelligence as "the computational part of the ability to achieve goals in the world". Clearly, there can be a variety of goals, some of which depend on meaning and are difficult to reduce to formal methods (e.g., identify "sick" patients) and some that are relatively simple (e.g., 5 + 2 = ?). Some AI researchers spent decades attempting to develop machines that can process meaning. Indeed, a (somewhat pejorative) definition of AI is "[t]he study of how to make computers do things at which, at the moment, people are better" [38]. Thus, biomedical informatics does not have an exclusive claim on "processing meaning". AI researchers have been trying for decades. However, AI researchers generally (but not exclusively) focus on computational aspects of intelligence; as per McCarthy's definition. In contrast, informaticians are concerned, more broadly, with information and our use of it, either individually, as teams, or in concert with the artifacts that we use to store, transmit, and manipulate it (e.g., paper, whiteboards, phones, computers, etc.).

Like computer science, mathematics and statistics provide important tools and methods for informatics, but their central object of study relates to formal abstract patterns and features of data, not meaning. Their utility in informatics projects is due to their ability to manipulate and reveal patterns in data and to draw formally correct conclusions that we (human beings) may then see as meaningful. For example, we can apply statistical methods to text and provide semantic similarity measures that, in some cases, closely correspond to human judgment. There are also sophisticated statistical tools for detecting differences, and hence new data to which we may choose to attach a meaning.

In a similar way, biomedical science is fundamentally different from informatics because biomedical science seeks to answer questions concerning biomedical issues, such as genetic factors that may affect lung cancer. Within biomedical science, informatics has grown in importance because of the increasing amount of information, both research and clinical, required to solve important problems. As we discuss below, biomedical science is a challenging application domain for informatics, because the relevant concepts are difficult to relate to formal representations.

Human factors and cognitive science are increasingly recognized as important in the design and application of information systems. Information systems are designed to support human activity. Therefore, to design usable and useful information systems, it is important to understand human cognition. Further, since current information systems process data (form), rather than meaning, human beings must ultimately assign meaning to the data, thus turning it into information. Thus, there is significant overlap with informatics. However, "[c]ognitive science is the interdisciplinary study of mind and intelligence..." [39]. Thus, its' object of study is cognition, not information or knowledge.

Finally, biomedical engineering is sometimes confused with biomedical informatics. Again, there are some projects that blur the distinction. However, biomedical engineers seek to solve biomedical problems using engineering methods. These solutions may take the form of devices or computer programs (e.g., simulation of biomedical processes). However, the focus is on the biomedical problem to be solved, not data, information or knowledge.

Please note that the above discussion does not imply computer science, statistics/mathematics or biomedical engineering are somehow less important than informatics; only that they have a different primary focus. In some cases, these fields adopt a different perspective on the same problem. Clinicians care for patients. Informaticians develop methods for applying and/or retrieving the information needed to support effective care. Computer scientists provide efficient algorithms to manipulate the data underlying the information.

There are, of course, frequent areas of overlap and we do not argue that the world is clearly demarcated into informatics and non-informatics. For example, magnetic resonance imaging (MRI) of the human brain may be the subject of research for computer scientists. In those cases, the question becomes: to what extent is information the "central" focus of the activity? For example, if the goal is to transmit images that happen to be MRI images of human brains, then the goal is more within the scope of electrical engineering or computer science, not informatics. On the other hand, if the goal is to deal with the information from an MRI and diagnosis of human disease (e.g., retrieve all patients whose MRI shows glioblastome multiforme), then the project is more related to informatics than to computer science.

It is worth noting that "information science" is an active field of study. There are schools of information science. If information science focuses on information, where information is defined as data + meaning, then information science is fundamentally and scientifically the same as informatics. The distinction between information science programs and biomedical informatics programs is thus a matter of application domain, rather than fundamental science. Indeed, some schools are changing their names to "schools of informatics" (e.g., Indiana University School of Informatics).

Finally, we do not wish to imply that these are the only fields of importance to informatics. Because human beings ultimately construct and manipulate meaning, any field that has meaning as a central object of study must use techniques, theories and results from fields such as cognitive science, psychology, linguistics, and sociology, among others.

5.2. Implication #2: computation is an important tool for informatics, but is not the primary object of study and is neither a necessary nor sufficient condition for informatics

In our definition, information, not computation, is the primary object of study of informatics. Many activities in informatics have nothing to do with computation (i.e., computers). Within health care, time-based, source-based, and problem-oriented medical records are all important informatics products that predate computers. Thus a central concern in informatics is: what information is needed and how is it best represented to support a specific set of human activities [40]. Information architecture, ontologies, and book indices are all important informatics tools that do not depend on computers. However, computation is increasingly important as the amount of available data increases exponentially. Simon pointed out some time ago that scarcity of attention, rather than scarcity of data is a fundamental barrier to effective use of information [41].

5.3. Implication #3: defining informatics as the study of meaningful data informs informatics curriculum design

Our definition provides clear guidance regarding the core skills and knowledge sets required of a well-trained informatician. The primary goal of an informatics education should be to prepare students to work with information (data + meaning). Academic informaticians may develop new theories, models, and tools for solving problems that deal with information, such as information needs, information architecture, information retrieval, and the characteristics of information. Since all information must have some data representation, informaticians must also be well versed in tools that help us store, retrieve, and manipulate data. This includes skills in computer science such as databases, data warehouses, and so on. They must also understand techniques for deriving new data, and possibly new meaning, from existing data. For example, artificial intelligence (AI) techniques, such as machine learning, can reveal relations among data that may be meaningful.

Another class of skills relates to the study of representations and algorithms that permit computers to appear as if they understand meaning, even if in a rudimentary way. Thus, ontologies and semantic applications are essential to informatics. Finally, since human beings construct meaning by looking at representations, informaticians must understand how representations (such as visual, haptic, aural, etc.) and a person's interaction with them affect a person's ability to construct meaning. Thus data visualization, exploratory data analysis tools, and human factors engineering all play a major role in constructing tools that help human beings discover, understand, and use information.

5.4. Implication #4: The emphasis on meaning allows us to see why some informatics problems are easier than others

This definition allows us to understand why some informatics problems are easier than others. Consider the banking system.[1] Clearly it is quite complex and involves a great deal of data and meaning. Why do all banks use computers? In contrast to biomedicine, we hear no arguments regarding the suitability of computers to track accounts. Why is this? We argue that in the case of banking, there is a very narrow "semantic gap". In other words, the correspondence between the data (numbers) and information (account balances) can be very direct. As we manipulate representations of numbers, the meaning of these manipulations follows easily.

Namely, if the problem relates strictly to form (data), or is easily reduced to a form-based problem, then computers can easily solve it. Retrieving all abstracts in PubMed containing the string of characters for the term "obesity" is a question related to data and is easily reducible to a form-based data query; whereas retrieving all abstracts in PubMed that report a positive correlation between beta blockers and weight gain is an information retrieval question that depends on the meaning of the query and the meaning of the text in the abstracts. This is not easily reducible to form and is therefore much harder to automate.

In general, concepts definable with necessary and sufficient conditions are relatively easy to reduce to form, and thereby permit some limited automated processing of meaning. However, concepts without necessary and sufficient conditions (e.g., recognizing a cup or a sick patient, or defining pain) cannot be easily reduced to data and are much more difficult to capture computationally.

Biomedical informatics is interesting, in part, because many biomedical concepts defy definition via necessary and sufficient conditions. This is true because biomedicine studies naturally

[1] We are referring here to the banking function of tracking accounts. Clearly, the financial system as a whole, and banks in particular, do much more than track accounts.

evolved systems as opposed to human-engineered systems. Evolution implies a chain of propose, copy and modify with a selection pressure. In other words, a population of individuals with (usually minor and relatively random) variations is exposed to an environment in which some are better able to reproduce (and their progeny to survive) than others. The population is, in most descriptions, composed of individual biological organisms such as plants, animals or human beings. Representations and symbol systems can also be created using a copy, modify and test method [42]. Variation between individuals is tolerated over time as long as it has a neutral or positive effect on reproduction. Variation that imparts a reproductive disadvantage relative to competitors is gradually removed from the population.

Systems that evolve tend to have specific properties that make them difficult to represent mathematically and thus, to compute upon. Evolved systems tend to be non-decomposable or, at best, nearly decomposable [41]. For example, consider the functional systems of an airplane. In order to fly, it must generate lift (force that counteracts gravity) and thrust (force that propels the airplane forward). The airplane has two distinct subsystems to develop lift and thrust: wings that develop lift and engine(s) that develop thrust. Clearly, these systems interact (a stationary wing develops no lift), but they are clearly distinct. We note that engineered systems often go through multiple iterations based on experience (e.g., Boeing 707 → 737). However, this process is better described as "re-engineering" than evolution.

On the other hand, a bird's wing develops both lift and thrust and these are not decomposable. One cannot remove the "thrust" component of a bird's wing. In addition to lift and thrust, a bird's wing has multiple other functions such as protecting the vital organs from trauma, conserving body heat, etc. Thus, one cannot consider (and model) the functions of a bird's wing in isolation from each other except as an approximation.

Similarly, it is difficult to clearly separate body systems. For example, the kidneys are not generally considered to be part of the circulatory system, but they have a very important role in maintaining blood pressure and preventing fluid overload. Indeed, some of the most common treatments for congestive heart failure, diuretic medications, act primarily on the kidneys and not the heart. Consequently, drawing distinct boundaries between evolved systems and their components is difficult.

Blois [43] argued that, in order to compute upon a system, one must first determine the system's boundaries. In other words, one must define all of the relevant components and assume that everything else is irrelevant. However, this is very difficult to do for evolved systems. If we want to model the circulatory system, can we exclude the renal system? The endocrine system that includes the adrenal glands (releases epinephrine that constricts blood vessels and raises blood pressure)? The nervous system? And so on.

Evolution tends to satisfice [41] and not optimize. If an individual survives long enough to reproduce and pass on its genetic material, that is good enough. There is no requirement for optimal fitness. Thus, some variability is tolerated in a population and is even desirable since the future environment progeny will encounter is unpredictable. No two human beings are exactly the same. In contrast, engineered systems are made identical in many important characteristics. They have interchangeable parts – a wing from one airplane will fit another airplane as long as they are the same model. All other things being equal, an airplane will react the same as another example of that model to damage or set of environmental conditions (e.g., wind shear, turbulence). In contrast, two human beings may react very differently to the same drug or the same surgical procedure.

We note that engineered systems are not necessarily less complex than evolved systems. Indeed, quantifying and comparing the complexity of two systems is not straightforward. However, few would argue that a Boeing 747 or the space shuttle are not complex systems. Thus, the evolved systems are not simply complicated or more complicated than engineered systems. Instead, they are complex in a different way compared to engineered systems. This property makes them less likely to be reducible to form and thus amenable to automation through computation.

6. Conclusion

Biomedical informatics is the application of the science of information as data plus meaning to problems of biomedical interest. This definition is sufficiently broad to include the majority of activities currently considered to fall within the scope of biomedical informatics while excluding activities that are traditionally considered to be outside of our field. As such, our definition can serve as a guide to students, educators, practitioners and researchers. Significant work remains be done to understand and operationalize the implications of this perspective. However, we believe that this definition captures the intuition behind many of the definitions of informatics, while also opening the door for a paradigm shift in how we view and practice informatics.

Patel and Kaufman [44] argued that biomedical informatics is a "local science of design". Local in the sense that biomedical informatics is a "science where principles simplify and explain parts of the domain of interest rather than provide universal coverage or a unifying set of assumptions". However, "the collection of particulars (derived from specific systems and approaches) advanced by individual institutions leads to the development of notions that are nearly universal (i.e., principles, paradigms, and theories), and they in turn shape the discipline and guide development". We hope that this work is a step toward the development of such (nearly) universal principles, paradigms, and theories. Informaticians are often asked by collaborators and members of the general public – "What is informatics? It behooves us to have a clear answer.

Acknowledgments

The authors thank Drs. M. Sriram Iyengar and Dean F. Sittig for valuable discussions regarding the ideas expressed in this manuscript. Supported in part by the Center for Clinical and Translational Sciences at UT-Houston (1UL1RR024148).

References

[1] AHIMA facts. 2007 [cited 2007 December 17]. Available from: http://www.ahima.org/about/about.asp.
[2] Ledley RS, Lusted LB. Reasoning foundation of medical diagnosis. Science 1959;130(3366):9–21.
[3] Collen MF. Health care information systems: a personal historic review. In: Proceedings of ACM conference on History of medical informatics. Bethesda, MD: Association for Computing Machinery; 1987.
[4] Hammond WE. Patient management systems: the early years. In: Proceedings of ACM conference on History of medical informatics. Bethesda, MD: Association for Computing Machinery; 1987.
[5] Musen MA, van Bemmel JH. Challenges for medical informatics as an academic discipline. Methods Inf Med 2002;41(1):1–3.
[6] Friedman CP, Ozbolt JG, Masys DR. Toward a new culture for biomedical informatics: report of the 2001 ACMI symposium. J Am Med Inform Assoc 2001;8(6):519–26.
[7] Musen MA. Medical informatics: searching for underlying components. Methods Inf Med 2002;41(1):12–9.
[8] Staggers N, Thompson CB. The evaluation of definitions for nursing informatics: a critical analysis and revised definition. J Am Med Inform Assoc 2002;9(3):255–61.
[9] Turley JP. Toward a model for nursing informatics. Image J Nurs Sch 1996;28(4):309–13.
[10] Lusignan Sd. What is primary care informatics? J Am Med Inform Assoc 2003;10(4):304–9.
[11] Berman JJ. Biomedical informatics. Sudbury, MA: Jones and Barlett Publishers; 2007.

[12] Han YY et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. Pediatrics 2005;116(6):1506–12.
[13] Koppel R et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005;293(10):1197–203.
[14] Brixey JJ et al. Interruptions in a level one trauma center: a case study. Int J Med Inform 2008;77(4):235–41.
[15] Zhang J et al. A cognitive taxonomy of medical errors. J Biomed Inform 2004;37(3):193–204.
[16] Wang TD et al. Aligning temporal data by sentinel events: discovering patterns in electronic health records. In: Twenty-sixth annual SIGCHI conference on human factors in computing systems. Florence, Italy: ACM; 2008.
[17] Lorenzi NM. The cornerstones of medical informatics. J Am Med Inform Assoc 2000;7(2):204–5.
[18] Forsythe DE. New bottles, old wine: hidden cultural assumptions in a computerized explanation system for migraine sufferers. Med Anthropol Q 1996;10(4):551–74.
[19] ANA publication NP-907.5M. The scope of practice for nursing informatics. Washington, DC: American Nurses Association; 1994.
[20] Park JY, Musen MA. VM-in-Protege: a study of software reuse. Medinfo 1998;9(Pt. 1):644–8.
[21] Coiera E. Guide to health informatics. 2nd ed. New York: Oxford University Press, Inc.; 2003.
[22] Greenes RA, Shortliffe EH. Medical informatics. An emerging academic discipline and institutional priority. JAMA 1990;263(8):1114–20.
[23] Shortliffe EH, Blois MS. The computer meets medicine and biology: the emergence of a discipline. In: Shortliffe EH, editor. Biomedical informatics: computer applications in health care and biomedicine. New York, NY: Springer Sicence + Business Media, LLC; 2006. p. 3–45.
[24] van Bemmel JH. The structure of medical informatics. Med Inform 1984;9:175–80.
[25] Musen MA, van Bemmel JH. Handbook of medical informatics. March 25, 1999 [cited 2007 December 19]. Available from: http://www.mieur.nl/mihandbook/r_3_3/handbook/homepage_self.htm.
[26] Ackoff RL. From data to wisdom. J Appl Syst Anal 1989;16(1):3–9.
[27] Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. J Inf Sci 2007;33(2):163–80.
[28] Floridi L. Semantic conceptions of information. October 5, 2005 [cited 2008 November 13]. Available from: http://plato.stanford.edu/entries/information-semantic/.
[29] Adams F. Knowledge. In: Floridi L, editor. The Blackwell guide to the philosophy of computing and information. Malden, MA: Blackwell Publishing Ltd.; 2004. p. 228–36.
[30] Anderson JR. Verbatim and propositional representation of sentences in immediate and long-term memory. J Verbal Learn Verbal Behav 1974;13(2):149–62.
[31] Mandler JM, Ritchey GH. Long-term memory for pictures. J Exp Psychol Hum Learn Mem 1977;3(4):386–96.
[32] Sachs JS. Recognition memory for syntactic and semantic aspects of connected discourse. Percept Psychophys 1967;2(9):437–42.
[33] Sachs JS. Memory in reading and listening to discourse. Mem Cogn 1974;2(1a):95–100.
[34] Cosmides L, Tooby J. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. Cognition 1996;58(1):1–73.
[35] Gigerenzer G. How to make cognitive illusions disappear: beyond "Heuristics and Biases". Eur Rev Social Psychol 1991;2(1):83–115.
[36] Gigerenzer G. The taming of content: some thoughts about domains and modules. Thinking Reasoning 1995;1:324–32.
[37] McCarthy J. What is artificial intelligence? 2007 [cited 2009 May 17]. Available from: http://www-formal.stanford.edu/jmc/whatisai/node1.html.
[38] Rich E, Knight K. Artificial intelligence. 2nd ed. McGraw-Hill; 1991.
[39] Thagard P. Cognitive Science. Stanford Encyclopedia of Philosophy 2007 [cited 2009 February 23]. Available from: http://plato.stanford.edu/entries/cognitive-science/.
[40] Friedman CP. A 'fundamental theorem' of biomedical informatics. J Am Med Inform Assoc 2009;16(2):169–70.
[41] Simon HA. Sciences of the artificial. 3rd ed. Cambridge, MA: MIT Press; 1996.
[42] Koza JR. Genetic programming: on the programming of computers by means of natural selection. Cambridge, MA: MIT Press; 1992.
[43] Blois MS. Information and medicine: the nature of medical descriptions. Berkeley: University of California Press; 1984.
[44] Patel VL, Kaufman DR. Science and practice: a case for medical informatics as a local science of design. J Am Med Inform Assoc 1998;5(6):489–92.