A.T. McCray

National Library of Medicine,
Bethesda, Maryland, USA

Commentary

The Theoretical Basis of Medical Information Science

Reflections on Marsden S. Blois' paper on *the proper use of man and machines*

Scott Blois, who as early as 1971 had already developed a sophisticated approach to the design of hospital information systems[1], was well aware of the benefits, but also of the limitations of applying computers to medical problems. He and others began to develop a theory of "medical information science", or "medical informatics", as it soon came to be known[1]. Blois' book, *Information and Medicine* [3], published in 1984, eloquently outlines his view of the relationship between medical knowledge and the information systems that attempt to portray and use that knowledge. He wishes to argue that until we have a proper theory of medical information science we will make little progress in the application of computers to medical problems. He says, "...there has been an irresistible urge to apply computers to medicine, but considerably less of an urge to attempt to understand where and how they can best be used" [3:xiii]. Some fifteen years later, the book is still surprisingly timely.

The first two chapters of the book discuss the nature of information itself, including the crucial role of context and knowledge of the world in the communication of information. Because information is imparted through descriptions and because descriptions are always abstractions of one form or another, it becomes necessary to choose a suitable level of abstraction for the purposes at hand.

In the third chapter, Blois makes a distinction between nominals (essentially, the names of objects in the world) and the attributes of those nominals. He claims, as have some philosophers of language, that the only way to describe an object is through its attributes. Further, as he states in chapter 4, it is precisely these attributes that are at the core of medical descriptions, and it is the attributes to which our attention is drawn. He gives the following example to clarify these ideas:

We may have a patient complaining of abdominal pain, which has been present off and on for several hours and which is now localized and persistent in the right lower quadrant of the abdomen. Examination reveals that the pain can be aggravated by gentle pressure in this region, and that the patient has a mild fever. The patient's white blood cell count is found to be 12,500 cells per cubic millimeter, and red blood cells are seen upon microscopic examination of the urine sediment. These attributes are those of the nominal "appendicitis", and to have this particular name come to mind is one of the meanings of diagnosis [2:66].

Importantly, descriptions can be classified according to their place in a hierarchy of descriptions, from simplest to most complex, thus forming a knowledge network. Blois illustrates this with a partial hierarchy of natural objects [3:47]. If the description "man" is, for example, at level zero in a hierarchy, then below that level might be organs and tissues, and below that will be cells, and so on. Similarly, above level zero might be "father" and above it would be "family" which is itself part of a larger social structure. By considering a description as it fits into its hierarchy, it is now possible to develop a theory of individual descriptions and their interconnection at different levels of the hierarchy. In particular, items toward the bottom of the hierarchy will be simpler and less ambiguous. As we move up the hierarchy, nominals become attributes of items at the next level up and, thus, become *embedded* in the higher level description (e.g., "man" becomes an attribute or, as others might say, a feature, of the nominal "father" at the next higher level). Attributes emerge at particular levels and thereby be-

[1] See Collen [2:37-43] for some discussion of the origin of the term itself.

same attributes of all higher levels, but not of lower levels (e.g., "cell" is an attribute of organ, and thereby also of "man", but "man" is obviously not an attribute of "cell").

Descriptions can now be compared by considering whether they share or differ in attributes. Given a context and a purpose, the absence of an attribute in one of the descriptions can be more significant than the presence of some other, less important attribute. Successful diagnosis of a medical condition, for example, depends heavily on recognizing the relative importance of those attributes that are present and those that are known to be absent. Descriptions are sometimes fuzzy in their expression, not because we intend to be unclear but, rather, because of the nature of the world. This has serious implications for reasoning processes that have as part of their premises fuzzy descriptions, a phenomenon often found in medical reasoning.

Having developed his theory of descriptions in the first four chapters of the book, Blois, in the next two chapters, applies that theory to the analysis of disease descriptions and to the process of diagnosis. Disease descriptions, like all other descriptions, consist of attributes, whether these are the known symptoms, signs, and course of the disease, or whether they are those that are observed, described, and treated in particular patients. Blois presents a hierarchical model of diseases that has at its highest level the patient as a whole and at its lowest level atoms and ions. Because clinicians attempt to treat the causes, not simply the effects of diseases, they are often treating lower-level phenomena. He notes [3:109]:

*When we speak of specific therapy, we are most likely found to be referring to lower-level things, to matters lying at the organ, cell, or molecular level. Here the disease process can be confronted in its early causality. Nonspecific or **symptomatic** therapy is concerned with the alleviation of abnormalities lying at higher levels...*

Diagnosis becomes a problem-solving activity involving isolating and classifying attributes. As our knowledge of the underlying causes of a disease increases, we can revise our classification methods based on the newly understood attributes of the disease in question. He notes that there has been only one serious attempt to describe diseases through their attributes, and that is the Current Medical Information and Terminology (CMIT) [5]. CMIT, though not originally intended to be a coding system, did serve as a standardized naming scheme, describing some 3,000 diseases. Blois points out [3:124] that "A long-standing difficulty in medicine has been the lack of a uniform nomenclature for diseases. Medical communications, record keeping, and the computerization of medical information have been plagued by this"[2].

The last four chapters of the book address the computability of medical decision making. While computers are excellent and efficient data processors, they are not generally successful at higher-level medical reasoning since this, as Blois points out, depends heavily on the interaction between observation of the particular patient and a vast store of medical, or world, knowledge. Systems that link clinical information systems to medical knowledge bases might have some hope of simulating medical decision making as it is practiced by physicians. If it is possible to recognize algorithmically the relationship of disease attributes to one another, then this might make the automatic generation of differential diagnoses feasible. Chapter 9 concludes with a discussion of nothing less than the field of medical informatics itself. Blois notes that new skills are needed by computer programmers who are attempting to design and implement medical information systems. These individuals will need appropriate training not only in computer science but, increasingly, in the domain of medicine itself. He says [3:234]:

This discontinuity in formalization between a manual (human) medical information process and the machine code necessary to accomplish comparable ends begins at a very high descriptive level and it is not itself a concern of computer science. If this concern is to be given a name at all, it must be regarded as concerning medical applications, and it is increasingly being referred to as "medical information science" in the United States, and as "medical informatics" in Europe. It will be the task of this new discipline to better understand and define ... medical information processes ..., in order that appropriate activities will be chosen for computerization, and to improve the man-machine system.

Chapter 10, *On the proper use of men and machines*, is the last chapter of the book and is reprinted here as part of this retrospective on medical informatics. Blois points out that in the past computers were most heavily used by researchers for particular tasks that would otherwise have been difficult, if not impossible, to carry out. Such users were strongly motivated and were willing to overlook some of the deficiencies of computerized systems if the systems provided them with

[2] Blois was an active participant in the early days of the UMLS project [6,7], which directly addresses the nomenclature problem in medicine, and his views on the nature of medical descriptions strongly influenced the direction of the project, particularly in the development of the most complex of its knowledge sources, the UMLS Metathesaurus.

critical assistance on the problems they were addressing. Blois predicts that computers will be much more widely used when communication networks improve, when there are standard protocols, when user interfaces are consistent, and when there are meaningful applications available for general use. He cautions that, given the inherent fallibility of computerized systems, they are only as good as the programming that created them - it is important to evaluate their results and, better yet, to design systems that explain their own behavior[3]. Blois goes on to discuss the inevitable reductionism and, therefore, loss of information that eventuates when computers are programmed to conduct complex tasks that are informal (and high-level) rather than formal (and low-level) in nature. He argues that computers are best used as tools for large-scale data storage and processing. He says [3:246], "Storage capacity and speed are the characteristics in which computers win hands down, if we can suitably formalize our applications and develop programs that can take advantage of them."

Blois concludes by noting that information systems for the foreseeable future will be "combinations of men and machines. The machine will be increasingly employed as a sophisticated and powerful tool..., but operating under the direction of a human, who alone is likely to be able to identify problems worth solving, and who must decide when to use a machine and when to turn it off"[4].

Many of the basic, and continuing concerns of medical informatics are either directly or indirectly addressed in this insightful and theoretically motivated work. Blois points, for example, to the need for an improved communications infrastructure, a matter that is currently being addressed through several efforts, including the Next Generation Internet project funded by the U.S. government. He argues for the establishment and use of standards, which is clearly an area of continuing research in the medical informatics field, as well as in the broader computer science community. HL-7, DICOM, Arden Syntax, and the UMLS represent a range of some of the current standardization activities in the medical informatics community. He indicates that when there are meaningful applications for general use in the medical community, user interfaces will need to be improved if busy clinicians are to use these computerized systems as part of their daily work. This is, of course, the subject of a good deal of research, including many of the developments in the design of physician workstations construed as "one-stop-shopping" systems. If computers are to be used for medical decision making (an application that Blois feels is well beyond the current capabilities of computer systems), then they must, at a minimum, be linked to some source of medical knowledge and, in particular, to clinical information systems. Some of the work on case-based reasoning as well as recent work on electronic patient record systems has explored such links, though not in exactly the form envisioned by Blois.

Perhaps, however, Blois has been most influential in his views concerning the field of medical informatics and of the research in which we are engaged. He has attempted to put the entire field on sound philosophical grounds by developing a theory of medical information science. The theory is based on the hierarchical nature of medical descriptions which, as Blois convincingly argues, has consequences for the way in which we are able to communicate about diseases, their diagnoses and their treatments. High-level descriptions are often abstractions and, thus, depend in their interpretation on rather sophisticated reasoning, as well as on significant amounts of world knowledge. Blois suggests that if we recognize that computers will have limited ability to handle higher-level phenomena, yet do rather well at handling lower-level phenomena, then we will have found the proper balance between man and machine.

[3] In a later work [8:33], Blois argues that if computer scientists have even some understanding of medicine, they will be more likely to design systems that are actually needed by medical professionals, and if medical professionals understand some of the basic principles of computer science, they will be in a better position to develop systems without repeating some of the mistakes of their predecessors.

[4] This sentiment has been echoed quite recently by William Stead in a short editorial commenting on the work of Nordyke and Kulikowski [9], which describes an informatics-based clinical practice that over many years has profitably made use of simple paper worksheets as input to a database management system. Stead comments that Nordyke and Kulikowski's work differs from the majority of work done on computer-based patient records in paying primary attention to the organization and utilization of information. "The computer is used only for those parts of the process where it does something that cannot be done reasonably any other way (retrieval and analysis of a population of records) or something that can be done without added work (generation of patient reports)"[10:131].

References

1. Blois MS, Henley RR. Strategies in the planning of hospital information systems. Journees d'Informatique Médicale, St. Lary: Institute de Récherche d'Informatique et d'Automatique, 1971.
2. Collen MF. A History of Medical Informatics in the United States: 1950-1990. Bethesda, MD: American Medical Informatics Association, 1995.
3. Blois MS. *Information and Medicine: The Nature of Medical Descriptions*. Berkeley: University of California Press, 1984.
4. Blois MS. On the proper use of men and machines. In: Blois.MS: Information and Medicine: The Nature of Medical Descriptions. Berkeley: University of California Press, 1984:235-55.

5. Gordon B ed. *American Medical Association. Current Medical Information and Terminology.* Chicago, 1971.
6. Tuttle MS, Blois MS, Erlbaum MS, Nelson SJ, Sherertz DD. Toward a biomedical thesaurus: building the foundation of the UMLS. In: Greenes RA ed. *Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Comp Soc Press, 1988: 191-5.
7. McCray AT, Miller RA. Making the conceptual connection: The UMLS after a decade of research and development. J Am Med Inform Assoc 1998;5:129-30.
8. Blois MS, Shortliffe EH. The computer meets medicine. In: Shortliffe EH and Perreault LE eds. *Medical Informatics: Computer Applications in Health Care.* Reading MA: Addison-Wesley Publ Comp, 1990: 3-36.
9. Nordyke RA, Kulikowski CA. An informatics-based chronic disease practice: case study of a 35-year computer-based longitudinal record system. J Am Med Inform Assoc 1998;5:88-103.
10. Stead WW. It's the information that's important, not the technology. J Am Med Inform Assoc 1998;5:131.

Address of the author:
Alexa T. McCray,
National Library of Medicine,
Bethesda, MD 20894,
USA
E-mail: mccray@nlm.nih.gov