Chapter XXIV
The Data–Information–Knowledge Model

Andrew Georgiou
University of New South Wales, Australia

ABSTRACT

The generation and transformation of data into information and knowledge is a basic formula in health informatics. This process is often represented in a model that portrays each component hierarchically with data at the bottom followed by an intermediary layer of information and topped by the knowledge layer. This model is a simple way to conceptualize important components of the informatics process, but it also has major limitations. The capture of data does not lead seamlessly to information or knowledge. The process is much more complex involving a multi-faceted web of interactions and issues.

INTRODUCTION

Health informatics is still an emerging and rapidly expanding academic discipline (Greenes & Shortliffe, 1990), located at the intersection of ICT and the many areas of healthcare. Its growth is a direct consequence of the dramatic expansion of ICT across health services over the last two decades. More than just the study of computers within medicine and its related fields, health informatics embraces a number of different fields and activities including patient care, healthcare research, education, and organizational planning and management (Peel, 1994; Shortliffe, 1991).

The heterogenous nature of the discipline means that it finds itself enmeshed in the many methodological and epistemological issues involved in the practice of healthcare (Georgiou, 2002). Indeed, even health informatics' most basic formula—the generation and transformation of data into information and knowledge—invites divergent opinions about the assumptions that underpin its claims (Hirschheim et al., 1995).

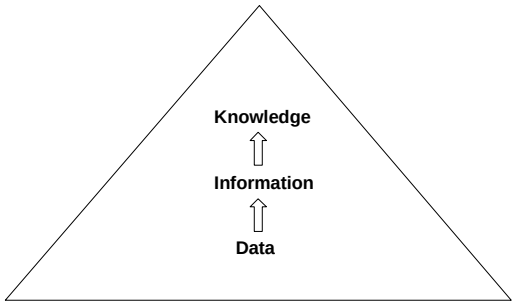
THE HEALTH-INFORMATICS MODEL

The process of distinguishing different forms and objects, and naming and categorizing them is an essential part of our ability to interact and communicate. This process is no less important to medical science and its need for communication channels to describe illnesses and share treatment options (Coiera, 2003). Models form an important element to the way we perceive and interpret the world (NHS Executive, 1996). Not only do they help us to describe and communicate aspects of the world, they also assist in understanding what is going on, perhaps even helping us to change and manage a part of reality.

The model often referenced within health informatics is the data-information-knowledge model. It is used to help identify and understand the different components and interrelationships within healthcare (Abdelhak et al., 1996; Coiera, 2003; Degoulet & Fieschi, 1997; Sheaff & Peel, 1995). The origin of this basic model can be traced back to the 19th century and the development of the functions of taxonomy and classification. Early statisticians used and developed classification systems as knowledge repositories developed from data and information (Desrosieres, 1998). Their model involved three essential parts arranged hierarchically, with data at the bottom and an intermediary layer of information topped by the knowledge layer.

Within this model (see Figure 1), data take on the character of facts or observations, which, in and of themselves, have little or no meaning. They take on significance only when they are provided with a contextual framework to manage and make sense of them. Information is assumed to be the product of processed data. The generation of knowledge then proceeds through a complex process involving deduction

Figure 1. The informatics model



(based on principles of logical implication, e.g., statements of certainty), abduction (which aims to establish links between observations such as cause and effect, e.g., statements of possibility), and induction (whereby generalizations are generated from specific examples to formulate general rules; Coiera, 2003; Degoulet & Fieschi, 1997).

This hierarchical model has underpinned many of the information infrastructures within healthcare including the use of disease classification data to help research and plan healthcare. The United Kingdom National Health Service's Information Management and Technology strategy in the 1990s described the process as a "language of health" (NHS Centre for Coding and Classification, 1996). It posited healthcare information as based on clinical terms that could then be transformed into classification systems such as the *International Statistical Classification of Diseases and Related Health Problems* (World Health Organization [WHO], 1993), and from there into casemix groups for cost and resource management.

Even the growth of evidence-based medicine (EBM) with its commitment to using science, research, and evidence to guide decision making (Appleby et al., 1995) envisages information flows that broadly replicate the