From an *Information Poor* to an *Information Rich* Health Environment

A Research Strategy in Health Informatics for the
UK Institute for Health Informatics

Alan Rector, University of Manchester

1. Goals

The NHS, like most healthcare organisations, is currently information and knowledge poor. Little information is of a quality or in a form to be used effectively. This is paradoxical, because healthcare is an information and knowledge intensive activity.

However, current health informatics does not meet either public or professional needs. Ge tting the right information to the right people at the right time in the right form to make a difference is still the exception rather than the rule. Information on individual patients is rarely linked to general knowledge effectively. The result is suboptimal care, inefficient use of resources, preventable clinical errors, and restrictions on patients and public to take responsibility for their own health and care.   Poor information also restricts research and threatens to cripple efforts to link recent advances in basic biological science to practical clinical care.
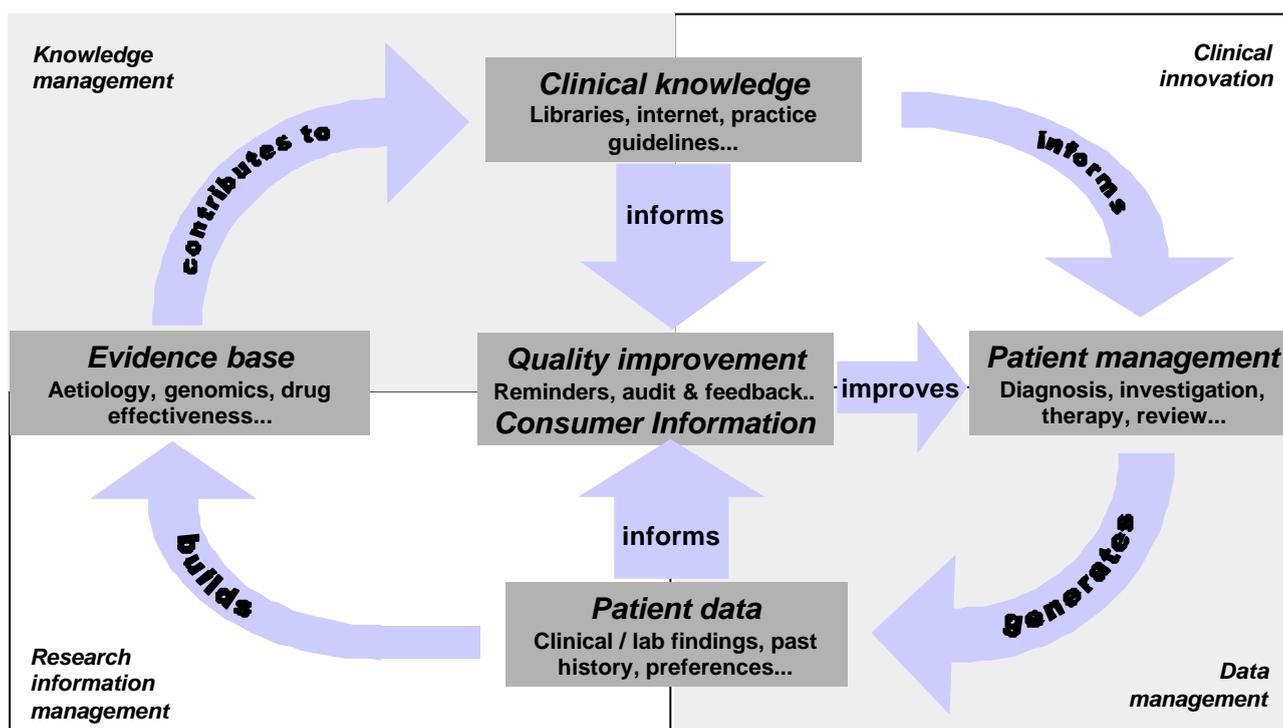
Figure 1: The Clinical Knowledge Cycle

The goal is to support the full knowledge cycle as shown in figure 1, and to support all three key stakeholders – consumers, clinical professionals, and researchers. In a world of 'e -Everything', new technology provides opportunities. However, major healthcare specific gaps remain which limit our ability to create and deliver rich health information in:

- *Organisational and human factors* –integrating IT successfully into the public's and professionals' healthcare practices requires a thorough understanding of a complex environment and many different interacting goals, tasks, and constraints.

- *Epidemiology and biomedical science* – getting the best value requires focusing priorities on areas of maximum impact and value and ensuring that the quality of the information leads to sound scientific results and management plans.
- *Information Technology and Computer Science* – meeting the stringent demands for healthcare information in dealing with flexibility, complexity, scale, safety, and security; formulating the vast range of health knowledge in coherent computationally sound form; making accessible information only available as text and speech.

The goal of Health Informatics Research is to close these gaps.

2.   Organisational and human factors: Doing the right things

- *Understanding the context and effects of IT* - Healthcare is amongst the most complex of human activities. System developers often underestimate its complexity and misunderstand what healthcare professions do. Healthcare professionals and systems developers both misunderstand the factors that affect the behaviour and experiences of patients and public. These mutual misunderstandings lead to systems which fail to deliver their intended benefits.

  Improved understanding is a prerequisite for more effective systems. We need to identify points of leverage, what IT interventions change and how, and why IT systems work or fail. To achieve this understanding requires sound qualitative and quantitative methods, often from behavioural sciences. To use this understanding requires building these techniques into our methodologies for designing, building, and disseminating software – and do so in ways which are practical within the constraints of healthcare.

- *Reducing clinical errors* - Healthcare is also an error-prone activity. Recent studies suggest errors are frequent, serious, and more most often the result of multiple failures in the healthcare system rather than simple human error. Studies also indicate that IT can be used to improve the overall system and reduce errors. However, the understanding of how best to use IT to minimise errors across the range of healthcare activities is only just beginning. The effect of various modes of 'decision support' is still controversial.

- *Interacting with Public and Patients* – Public and patients increasingly want access both to general information about health, disease, and therapy and to information about their own healthcare. There is strong evidence that patients benefit from access to understandable personalised information about their condition and treatment.

3.   Technical factors: Doing things right

- *Information Capture* – Information capture remains the Achilles' heel of clinical information. Most clinical information remains narrative text, with little or no structure. In study after study it has been shown to be collected inconsistently and inaccurately. Capturing high quality structured information is not integrated into clinical practice. The problem is acute for improving clinical care. Most management and diagnostic decisions remain based on 'history and physical'; the findings from images and signals must be described to be communicated and used for care or research. Yet too often, only numerical laboratory results are captured reliably. Accuracy, repeatability, and completeness of clinical data is often poor, and its capture often too time consuming to fit healthcare professionals' constraints. The problem is made acute by the research need to correlate genomic data with clinical information, an effort which will only be effective if we can capture high quality clinical information.

- *Language Technology* – Since so much clinical information exists only as narrative text, language technology for information extraction and understanding. Increasingly clinical texts are semi-structured documents, and specialised techniques of document management are needed to match the special characteristics of clinical documents. Narrative text is the natural mode for communicating much information to both professionals and patients. Hence, there is a growing need for language generation and language adaptation – making information originally held in databases or written for one group available to another. In the long run, the problem of understanding the mechanics of dialogue and explanation are critical both to information capture and to responding to queries from the public, patients, and professionals.

- *Complexity, Adaptable System, and Agents* –Adapting to the requirements of patients, professionals, clinical conditions, settings, and tasks leads to a to a combinatorially explosive number of variations. The heterogeneity of clinical information and queries means that information systems must cope with a combinatorially explosive number of types of information. The need to bridge levels of granularity from molecular to cellular to organ to organism results in a combinatorially explosive proliferation of detail. Managing this potentially explosive complexity is becoming a field in itself often focused on 'ontological engineering'.

- *Multimodal information management* – Clinical documents consist of combinations of images, signals, numeric, codes and text. Increasingly images are critical, and means to process, recognise, 'understand' and retrieve them are vital to clinical practice and research. Many medical images are not captured in ways which are intuitively understandable, but rather as two dimensional shadows, cross sections of three dimensional objects, still or moving.

- *Simulation and modelling* – Modelling and simulation have become fundamental tools to aid understanding of biomedical phenomena, from the expression of genes to the interplay of factors in the immune system to the dynamics of epidemics. Hybrid techniques are increasingly required to deal with the complexity of biomedicine.

- *Integration of heterogeneous systems* – All clinical systems exist in complex environments of other systems with which they must exchange information. Most interfaces between systems are built laboriously by hand. Real progress requires that many systems to meet many different needs integrate smoothly with minimal effort.

- *Knowledge acquisition, management, use and re-use* – Representing, managing and using knowledge are central objectives of healthcare IT. Most systems remain special built, from scratch, with their own information structures and terminologies. There are few re-usable resources, even in basic topics such as anatomy and physiology. Acquiring knowledge is labour intensive; making expensively acquired knowledge re-usable requires new techniques in its representation and curation. Decision making and inference procedures need to be represented in forms which are understandable by human experts, rigorously executable by software agents, safely re-usable within a wide, but clearly specified, scope.

4. Special problems of health informatics research

Health informatics is necessarily large scale and multidisciplinary. Research in healthcare informatics must often be conducted in healthcare institutions which are by there nature large and complex and have their own priorities. These factors bring with them special problems for health informatics research

- *Health informatics research is often dependent on local healthcare IT and other infrastructure* – which tends to be poor. Without infrastructure, it is difficult to disseminate projects widely enough to demonstrate their effectiveness. If research projects must finance new infrastructure, they become very expensive.

- *Health informatics research transcends the boundaries among funding agencies* – and is rarely the prime priority of any.

- *Health informatics research projects are inherently large and expensive* – as they require several staff from different disciplines, which puts them above the average for most grants from many funding agencies. Collaboration *per se* requires time and resources – a fact too seldom recognised. Addressing human factors adds to cost, and so is often omitted. Clinical salaries also raise costs.

- *Good research can be hard to recognise* – and can be lost in the cacophony of hype over 'The Web' and e-commerce. There is often confusion between short term development, evaluation, and long term research. Because there has been little funding for basic long term health informatics research in the UK, health informatics researchers are under represented on UK funding panels.

- *Intellectual property* – there are conflicts between the 'gift economy' and commercial ethos. Much potential intellectual property is too fine grained to be easily owned and too dependent on other information to be valuable in isolation.

5. Sample programmes

Two different orientations for programmes are needed. The first is oriented around the stakeholders in the healthcare process – consumers, healthcare professionals and researchers. The second is focused on particular areas of technology needed specifically by healthcare informatics. It is aimed, in particular at various aspects knowledge management –the upper left hand quarter of figure 1 – and at achieving sound foundations for long term development.

Programmes oriented around stakeholder needs

- *Consumer information management and knowledge delivery* - Information for the public that makes a difference. The long term research issues how to give patients particularly how to tailor information to patients' own problems and link it with their own healthcare records. Major efforts are needed in language and dialogue generation, specialised health vocabularies for consumers, links to general knowledge sources, and the means and effects of giving consumers ownership of their own records.

- *Clinical information capture, management and presentation for healthcare professionals*– integrating human factors, data capture, patient records, images management, advice, knowledge, and communication in ways that improve patient care, reduce clinical errors and relieve, rather than aggravate, professional stress. Major efforts are needed in human factors and data capture for Electronic Patient Record (EPR/EHR) and into the most effective means of supporting clinical decision making.   A particular focus should be the interface between models for terminology (ontologies), medical records and messages, and decision support.

- *Research information management: Knowledge creation and delivery* – making it easy to set up, manage, and analyse experiments and trials, especially those bridging the gap between clinical and genomic research. Issues include document management, information integration and fusion, bridging of levels of granularity from genomic to clinical, etc. Scalable solutions for managing large number of concurrent trials, pooling information,

and managing rapidly evolving information structures are required. Aids to curating, integrating, and indexing the rapidly exploding volumes of information will be a major focus.

Programmes oriented towards basic technology needed for health informatics

- *Rich common representation for biology and healthcare* – bridging the gap from molecular to tissue to organ to organism to disease and populations. Providing rich compatible scalable representations for genetics, anatomy, physiology, and clinical medicine is now a central research task for all areas of biomedical informatics which involves multidisciplinary collaboration between informatics, biology and medicine. Inference and search mechanisms must be capable of coping with the scale and complexity of health information resources running into hundreds of thousands, perhaps millions of concepts. Developments leading towards a 'semantic web' in biomedicine require a scope and flexibility of knowledge representation not yet achieved.

- *Scalable, highly adaptable systems* – scalable, enterprise wide knowledge management, information capture, user interface, problem solving, and information resources from coherent sources. Issues include managing complexity, user interfaces, interworking systems, knowledge management, terminology, evolving information systems, and information fusion. Scalability and adaptability have been major limitations of previous generations of health information systems and should be a major focus for future development in all areas, particularly data capture and knowledge management. There is a critical need for more flexible means of modelling evolving information systems for novel disciplines, such as genomic medicine,. where needs cannot be fully specified in advance. To meet the expanding needs of biomedical research, we must greatly improve our ability to re-use information to speed up the design and implementation of systems to answer novel questions.

- *Language processing for healthcare* – extracting information and knowledge from heterogeneous sources, both text and speech, related to individual patients and to clinical literature; presenting formal information effectively to different users in different situations. Underlying issues include special features of medical language, multiple speech recognition, complex dialogue management, and gathering sufficient corpora to support the research without compromising patient confidentiality. More robust techniques for supporting dialogue with both professionals and the public are required.

- *Integrated image processing and clinical research environment* – issues include structure recognition, content retrieval, enhancement, annotation, curation and indexing as well as user environments to make best use of three and four dimensional reconstructions and 'virtual reality'.

Prepared by: Alan L. Rector. *Professor of Medical Informatics, University of Manchester,* rector@cs.man.ac.uk

Prepared for: The UK Institute of Health Informatics

Date: January 2001