



Archived at the Flinders Academic Commons:

<http://dspace.flinders.edu.au/dspace/>

'This is the peer reviewed version of the following article:

Maeder, A.J. (2008). e-Research Meets e-Health. In Proceedings of the second Australasian workshop on Health data and knowledge management (HDKM '08) Darlinghurst, NSW: Australian Computer Society. 2nd Australasian Workshop on Health Data and Knowledge Management. Wollongong, NSW. Jan 2008, pp. 3-6.

which has been published in final form at

<http://crpit.com/confpapers/CRPITV80Maeder.pdf>

Copyright © 2008, Australian Computer Society, Inc. This paper appeared at the Australasian Workshop on Health Data and Knowledge Management (HDKM 2008), Wollongong, NSW, Australia. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 80. James R. Warren, Ping Yu, John Yearwood and Jon D. Patrick, Eds. Reproduction for academic, not-for profit purposes permitted provided this text is included.

e-Research Meets e-Health

Anthony Maeder

e-Health Research Centre / CSIRO ICT Centre
20/300 Adelaide Street, Brisbane, Australia

anthony.maeder@csiro.au

Abstract

This paper considers some key aspects of e-Research methodology and infrastructure which are relevant to e-Health, and identifies some promising areas in e-Health where these aspects could be used beneficially.

Keywords: e-Health, e-Research, health databases, personal health, health monitoring

1 Introduction

e-Research (sometimes termed e-Science) (PMSEIC Working Group on Data for Science, 2006) has recently emerged as a powerful new paradigm in research methodology and has found application in many different disciplines. Health research has traditionally been conducted using established sciences or humanities research methodologies, and has yet to feel the full impact of e-Research. For this reason there has been little exposure in the health care sector to these methodologies and their benefits for non-research activities have not been recognised. In the operational environment of the health care sector, there is currently a strong migration towards use of electronic health records (EHRs) and the large scale deployment and integration of various types of health information systems, generally designated as Health Informatics or e-Health. Due to this move, a far greater amount of data is becoming available which could be used during operational activities leading to significant health care improvements. This paper considers some aspects of e-Research methodology and infrastructure which are relevant to e-Health and

identifies some promising areas in e-Health where they could be used beneficially.

2 Fundamentals

First we must clarify the scope of some terms and concepts associated with the two major knowledge domains with which this paper is concerned.

By *e-Research*, we mean the methods and approaches used in undertaking research activities which involve large collections of complex data sets, applying advanced and intensive computational techniques, and tools for enhancing human understanding and interaction, to support collaborative activities involving multidisciplinary teams.

By *e-Health*, we mean the broad range of information and communication technologies systems and tools deployed in healthcare applications, including software systems for health data management and processing/analysis, computer assisted clinical process and decision support systems, and systems interoperability components (including software supporting health care messaging, telehealth and medical device operations).

Next we consider the major contemporary issues affecting policy and operations in the health care sector, which have necessarily become drivers of change in Australia, and many other developed countries.

Foremost amongst these are the demands and costs of an *increasing burden of disease* (e.g. youth obesity and related impacts such as diabetes; chronic diseases such as cancer and cardiovascular; ageing population with multiple comorbidities and complex care requirements). Much of this burden arises from changes in demographic and socio-economic factors in the population profile, and is severely exacerbated by declining size and capability of the professional workforce in the health care sector.

A second major issue is the need for systemic engagement with *quality and safety improvements* (e.g. correct application of clinical guidelines and associated care planning; monitoring of patient condition and treatment after discharge; reduction of adverse events and adverse medication responses; ongoing clinical education, training and certification). These improvements are often slow and difficult to achieve, due to the uniform expectation of a rigorous evidence base to justify any changes in clinical practice, and the nature of processes needed for making changes in a complex systems environment with a high proportion of human involvement. On the other hand, strong pressure for such changes is applied by the increasing development of quality assurance procedures throughout the health care sector, the diversification of commercial business models for different activities in health services delivery, and the raising of consumer expectations fuelled to some extent by the prominence of health care issues in the public arena.

3 e-Research Concepts for e-Health

This section will revisit the fundamental characteristics of e-Research described above, and discuss how each of these could be applicable to (and take advantage of) the emerging e-Health operational environment. Indeed, it will be argued that many elements of the e-Research environment already exist to support health care and more are conceived for the future, so that significant progress could be made by exploiting the common aspects of methodology and infrastructure.

An important distinction is made in this work between data, methodology and infrastructure which is part of the normal health care operational environment, rather than that associated directly with health research. The latter has been set up explicitly and purposefully to support research activity independently of the functioning of the health care system. This separation of operational and research environments, while often considered prudent (e.g. for preserving patient data confidentiality) or convenient (e.g. by avoiding translation of data from conventional information systems) can nevertheless result in inefficiencies such as data duplication and re-interpretation. We therefore concentrate on situations pertaining to the “resource” (i.e. serving a particular scientific

area user community) and “reference” (i.e. serving a broad set of various user communities) types of data sets as described recently by the National Science Board (2005).

Large collections of complex data sets are widespread in health. A typical hospital patient information system might contain as many as millions of records, and each such record might hold tens or hundreds of fields of data, of many different formats. Often these collections have been constructed by the direct translation of previous paper based records to digital form. As a consequence, fields may be comprised of formats such as categorical and nominal data (e.g. history checklists, triage decisions), structured and semi-structured alphanumeric data (e.g. demographics such as name and address; pharmaceutical prescriptions), time sequence data (e.g. charts; medication administration) and even free text and freehand (sketch) data (e.g. reports from pathology, surgery, psychology). Relationships between data fields are often not explicit and they can be incomplete or erroneous, with ambiguities due to lack of enforced standards regulating the contents. Tagging or metadata markup of fields is not common, as the time investment required for this would be impractical, however meaning is usually implied by structure or naming of fields. These data sets can change over time and may be linked with other data sets to obtain further information about individual patients, or cohorts of patients with some common characteristics. Specialist forms of large complex data sets exist, such as medical image databases, which impose their own demands on users (e.g. the need to deal with multiple image sequences, or to perform 3D reconstruction or spatial registration). Other complex data sets could readily be constructed readily from data generated during clinical procedures but not routinely collected (e.g. physiological data from monitoring devices in ICU or during anaesthesia).

Advanced and intensive computational techniques are currently invoked for handling large amounts of health data or messaging efficiently. These functions are often provided by complex, multicomponent software systems which can be custom configured to suit site- and task- specific user requirements. A major functional area of their operation is based on database access and related activities, including security, warehouse/repository access, and audit trail logging. This may require some use of

sophisticated techniques (e.g. for approximate or incomplete query resolution, or for linkage between heterogeneous or distributed databases). The functions must also satisfy a wide range of typical management information systems needs such as aggregating and reporting, and must interoperate with other elements of software (e.g. for billing, ordering, document preparation etc.). Other more recently emerging areas of high performance computational demand include decision support systems such as guideline checkers, and computer assisted diagnosis applications (e.g. for breast screening from digital mammograms). These areas rely on optimisation strategies for classical “expensiv” algorithms such as traversal of rule trees, and classification of data by clustering.

Tools for advancing human understanding and interaction, to support collaborative activities involving multidisciplinary teams are usually provided as individual fixed solutions, such as statistical summaries and graphical display for reports, multiparty conferencing packages, and telehealth patient examination applications. Passing elements of incremental information between members of a care team asynchronously can be achieved using conventional messaging infrastructure (e.g. e-mail; wiki) along with annotations placed in charts via PCs or PDAs. The ability to collect data through intervention or continuously at the point-of-care (POC) (whether from human or device source) can also provide better information to support multiple carers. Other interactions assisted by e-Health tools are remote surveillance of patients in ICU/HDU, telecollaboration of clinicians in emergency situations, and multidisciplinary clinical team meetings to review patient situations.

4 e-Health Opportunities for e-Research

Many of the e-Research methods which are being developed to implement the above concepts could be adapted easily to improve on limitations in the current e-Health environment. We will now briefly describe two example situations where major potential impacts could be obtained by this means.

4.1 Large Scale Health Data Linkage

Linkage of health data sets can be beneficial in a number of clinical situations:

- information from different data sets on a patient may be helpful in assessing their circumstances more fully and accurately, and in providing individualised and localised care;
- mapping information across a cohort of similar patients may allow identification of genome-phenome subsets with different progression of disease or response to therapy;
- population based analysis of risk factors and state of health can help support community wide management and improvement programs, or coordinated care management for chronic disease or aged patients.

A major impediment to large scale linking of health databases is the wide variability and inconsistency in data formats and meaning. It is not sufficient to simply have access to many different databases with clear schema representations. In addition, it is necessary to have abstraction and regularisation of the data so that the combination is correctly achieved. It is also necessary to be able to extract carefully delimited subsets of the data conforming to given properties, when records may be incomplete. These restrictions have meant that in practice large scale data linking often requires a high degree of manual involvement and correction of results.

These requirements are also encountered in e-Research situations where collections of pooled “commons” data from different sources are constructed. A favoured approach is to provide terminological and ontological alignment between the different data sets, and to employ tools which allow semantic relationships to be discerned. Coupled with appropriate tools to synchronise the temporal or “longitudinal” aspect of patient data, this approach allows consistent data collections to be constructed, despite variations such as data types and annotations. An example of this higher order kind of connected data space (or information environment) formed by amalgamating all available patient records is being adopted as the underlying model for a “Patient Journey Browser” currently being developed.

4.2 High Volume Patient Data Logging

Collecting health condition data such as vital signs and mobility for individual patients by means of personal monitoring devices is widely proposed to improve the quality of health care (including self care). The availability of this detailed information on a fine grained time scale adds considerable power to possible clinical interventions, such as:

- compliance of patients with prescribed treatment/recovery regime
- adverse reactions to drugs or therapies
- automated extraction of functional assessments and comparisons
- understanding activities of daily living and modifying bad habits
- managing risks and providing alerts to increased risk circumstances.

The huge volume of data generated by a multichannel monitoring device with a high sampling frequency provides many technical challenges for realising data transfer, storage and access. Furthermore, it requires appropriate underpinning software to provide analysis, comparison and interpretation of the data, possibly by care team members with widely differing amounts of clinical training and ability. It also needs sophisticated computational techniques which can provide pattern recognition, data classification and machine learning functions that can help to automate human utilization of the data. These functions together form the basis for a “Care Assessment Platform” software concept that will support multisite, multiparty care team operations. If appropriately deployed, the aggregation of individual patient details can also be used to provide evidence for clinical guideline variations.

5 Conclusion

The two examples provided above are initial steps towards using the e-Research paradigm for managing information, in order to create new operational tools and methods for e-Health. As e-Research becomes more widespread and popular, it is anticipated that such tools will proliferate, especially as the

underlying data environment becomes more accessible.

6 References

PMSEIC Working Group on Data for Science (2006), From data to wisdom: pathways to successful data management for Australian science.

http://www.dest.gov.au/NR/rdonlyres/D15793B2-FEB9-41EE-B7E8-C6DB2E84E8C9/15103/From_Data_to_Wisdom_Pathways_data_man_forAust_scie.pdf

National Science Board (2005), Long-lived digital data collections.

<http://www.nsf.gov/pubs/2005/nsb0540/start.jsp>