

The Development of the IMIA Knowledge Base

Author:
Graham Wright¹

Affiliation:
¹Faculty of Health Sciences,
Walter Sisulu University,
South Africa

Correspondence to:
Graham Wright

Email:
profwright@gmail.com

Postal address:
PO Box 294, Bathurst 6166,
South Africa

Dates:
Received: 29 Oct. 2010
Accepted: 18 May 2011
Published: 14 Oct. 2011

How to cite this article:
Wright, G., 2011, 'The
Development of the IMIA
Knowledge Base', *SA Journal
of Information Management*
13(1), Art. # 458, 5 pages.
doi:10.4102/sajim.v13i1.458

© 2011. The Authors.
Licensee: AOSIS
OpenJournals. This work
is licensed under the
Creative Commons
Attribution License.

Background: The discipline of health or medical informatics is relatively new in that the literature has existed for only 40 years. The British Computer Society (BCS) health group was of the opinion that work should be undertaken to explore the scope of medical or health informatics. Once the mapping work was completed the International Medical Informatics Association (IMIA) expressed the wish to develop it further to define the knowledge base of the discipline and produce a comprehensive internationally applicable framework. This article will also highlight the move from the expert opinion of a small group to the analysis of publications to generalise and refine the initial findings, and illustrate the importance of triangulation.

Objectives: The aim of the project was to explore the theoretical constructs underpinning the discipline of health informatics and produce a cognitive map of the existing understanding of the discipline and develop the knowledge base of health informatics for the IMIA and the BCS.

Method: The five-phase project, described in this article, undertaken to define the discipline of health informatics used four forms of triangulation.

Results: The output from the project is a framework giving the 14 major headings (Subjects) and 245 elements, which together describe the current perception of the discipline of health informatics.

Conclusion: This article describes how each phase of the project was strengthened, through using triangulation within and between the different phases. This was done to ensure that the investigators could be confident in the confirmation and completeness of data, and assured of the validity and reliability of the final output of the 'IMIA Knowledge Base' that was endorsed by the IMIA Board in November 2009.

Background

The author undertook this research over a four-year period with a number of collaborators in five discrete phases, which utilised quantitative and qualitative approaches. The discipline of health or medical informatics is relatively new in that the literature has existed for only 40 years. The British Computer Society (BCS) health group was of the opinion that work should be undertaken to explore the scope of medical or health informatics. A qualitative approach was used to gather expert opinion and construct a cognitive map of the discipline of health informatics. Once the mapping work was completed the International Medical Informatics Association (IMIA) expressed a wish to develop it further to define the knowledge base of the discipline and produce a comprehensive internationally applicable framework. Various data extraction methods were then used to identify the most commonly used keywords in the health informatics published literature followed by a consensus method to produce a final framework and knowledge base. This mixed method approach was adopted as a pragmatic means to address the development of what the discipline considered the current knowledge base and thus a reflection of the thoughts and publications of the discipline. The work was overseen by an International Research Advisory Board and refereed by Professor Lorenzi on behalf of the IMIA Board and General Assembly.

Research problem

The discipline of health informatics had not been formally defined and many definitions of the discipline have emerged in the literature. Not only was there a lack of agreed definition in that the discipline was variously called: health informatics, medical informatics, clinical informatics and latterly bioinformatics, but the scope of the discipline had not been adequately defined. Some of the consequences included misunderstandings regarding standards and use of terminology, lack of consistency within educational curriculum and a lack of a framework for defining skills and workforce requirements.

Objectives

The aim of the project was to explore the theoretical constructs underpinning the discipline of health informatics and produce a cognitive map (Eden & Ackermann 2004) of the existing understanding of the discipline.

Subsequent aims of the project were to develop the knowledge base of health informatics, which was seen as central to the IMIA strategy (Murray 2008; Lorenzi 2007), and to undertake the task of exploring the current perceptions of the Health Informatics community as to the scope of the discipline.

Method

The project's international advisory board of health informatics experts provided advice on the methods that were used and facilitated access to source materials. The mixed methods used in the project were:

- a consensus conference using a cognitive mapping exercise
- workshops to verify international interpretation
- extraction of keywords from the entire published index papers on health informatics using computer software packages and techniques
- workshop to examine keywords and exclude terms
- voting in of keywords by international volunteers using a voting system based in an Excel spreadsheet.

The aim was to obtain different perspectives (data) on the issue of mapping the discipline of health informatics with the belief that the analysis would provide confidence and confirmation that the data was complete and the final outcomes from all the phases of the project were not just artefacts of one particular method of data collection or analysis. This process of data gathering and systematic analysis reflects the principles of grounded theory where the researcher begins with an area of study and allows the theory to emerge from the data. In this project the area of study was the discipline of health informatics and the knowledge base was derived from data systematically gathered and analysed through the research processes undertaken (Strauss & Corbin 1998).

Wolf (2010) in a recent article says the final consideration in using a mixed method approach is 'to consider thoroughly whether to engage in triangulation, and if doing so, to use tailor-made triangulation strategies fitted to the research questions and interests'.

The project undertaken to define the discipline of health informatics used all four forms of triangulation (Denzin 1970) and this article describes how each phase of the project triangulated with the other phases for confirmation and completeness of data, and validation and verification of the project outputs.

Triangulation is a strategy to 'overcome the intrinsic bias that comes from single methods, single observer and single theory studies' (Patton 1990). Its objective is the confirmation and completeness of data through cross checking data

from several sources to seek out consistencies in the data (Begley 1996; O'Donoghue & Punch 2003). Many researchers also advocate triangulation as a means of resolving the quantitative and/or qualitative question through integrating the two approaches in one study and contributing to methodological rigor in order to validate the findings (Begley 1996; Cohen *et al.* 1994).

Denzin (1970) identified four forms of triangulation: data, investigator, theoretical and methodological. Data triangulation involves gathering data using different sampling strategies, so that segments of data are collected at various times, social situations and with different people. Investigator triangulation requires the use of more than one researcher in collecting and interpreting data. Using more than one theoretical position for data interpretation is called theoretical triangulation, whereas the most common form of triangulation, methodological triangulation, refers to the use of more than one method of data collection.

The ability to generalise findings to wider groups is one of the most common tests of validity for quantitative research. Triangulation is typically a strategy for improving the validity and reliability of research findings. Patton (2002:247) advocates the use of triangulation stating 'triangulation strengthens a study by combining methods'. However, the idea that triangulation is simply the combination of different methods of investigation is a restricted one, and researchers need to increase their use of the other less frequently employed forms of triangulation. When using triangulation of methods, researchers should also reflect on whether the use of within-method triangulation would be advantageous to their project. Within-method triangulation involves using dissimilar aspects of the same method in one study; for example, a questionnaire might contain two different scales to measure emotions. Between-method triangulation involves using different research methods, for example a questionnaire and observation to collect data (Bryman 2003; Begley 1996). Sequential use of quantitative and qualitative methods may also be more effective for some projects rather than simultaneous use, which do not permit the development and refinement of the methodologies. The deliberate use of multiple data sources and methods to cross-check and validate findings, should pervade all projects and lead to the objective of confirmation. Triangulation should be chosen intentionally, and a description of its rationale, planning and implementation is essential in project reports to give authority to triangulation and the project outcomes (Begley 1996).

The project explored the theoretical constructs underpinning the discipline of health informatics. The early project work was situated within a theoretical educational framework. Bloom's taxonomy affords a hierarchical scheme for categorising levels of complexity for objectives within educational settings (Bloom *et al.* 1984). It also overlays well against other academic levels, such as the progression from undergraduate to postgraduate levels (Furst 1981; Seddon 1978).

Bloom classified three domains of educational activity (Forehand 2005):

1. cognitive, describing knowledge and mental skills
2. affective, describing attitude, feelings and emotions
3. psychomotor, describing manual or physical skills.

Bloom identified six levels of educational objectives within the cognitive domain; from the lowest level, knowledge, through comprehension, application, analysis, synthesis and evaluation (Forehand 2005; Anderson *et al.* 2001). The first phase of the project was a mapping exercise that was based on these concepts.

The five phases of the project

Consensus conference

The 2005 Consensus conference was an intensive 24-hour workshop involving small group and plenary discussions, with participants and researchers in residence overnight. There were 24 invited participants drawn from a sample frame that had professions down one axis and organisations across the other. Organisations included health providers, family medicine, ASSIST the IT professions union, a number of United Kingdom (UK) health informatics groups and the IMIA; the world body for health and medical informatics. Most of the participants were from the UK whilst others came from Europe, Australia, South Africa and the USA. The conference aimed to capture all the elements of the discipline of health informatics and also the broad themes or subject areas into which these elements could be grouped. Within small groups, participants listed the main subject areas or themes from their own curricula, knowledge and experience. Then again within small groups they identified smaller elements of the subject areas. Finally, in a whole group activity, participants assigned each element to a subject area and a level from Bloom's cognitive domain (Forehand 2005) where possible. The discussions resulted in a first data set comprising 221 elements, grouped into 13 themes that varied in size, with the smallest containing six elements and the largest 37. It was recognised that the largest theme, the 'Toolkit', which consists of IT skills and knowledge of IT processes, would likely be divided following further discussions, which subsequently happened during a 24 hour workshop in Belfast.

This consensus conference therefore used group activities as its research methodology to produce lists of elements grouped into themes. There were six researchers involved in facilitating the group and plenary activities, thus adding investigator triangulation to reduce a single researcher bias.

Workshops to verify international interpretation

Workshops were conducted in 2005 at two major health informatics conferences, the European Federation for Medical Informatics (MIE 2005) in Geneva and the American Medical Informatics Association (AMIA 2005) in Washington DC. They were short workshops and hence only explored

the overall concept and the clinical informatics theme. Participants commented that there were no major issues with either the methodology used in phase one or the initial outcomes that should modify the direction of the project. These workshops therefore used investigator triangulation in that three of the original six investigators were present at the European workshop and two at the American workshop. The investigators were therefore a subset of the original research team employing both data triangulation in that data was gathered using a different sampling strategy, in other words, those international conference participants who chose to attend the workshop and methodological triangulation as the method here was not to create themes and elements but rather take that data and refine it through smaller and shorter validation workshops. Another workshop to validate the outputs was held in Belfast in 2007 after the January 2007 workshop in London highlighted the size of the toolkit. This meeting focused on refining the technical and computing themes previously developed in phase one and successfully affirmed the two technical themes 'Computer Science for Health Informatics (ICT for Health) and Computer Systems Applications in Health (toolkit)'. Thus the large toolkit theme was logically separated and participants from computer science who had expressed concern that the single large theme did not reflect the computer science heading system were the main re-shapers of the two new themes. The resulting themes are:

- computer science for health informatics (ICT for Health)
- health and social care processes
- health (care) records
- health and social care industry
- health informatics standards
- knowledge domains and knowledge discovery
- legal and ethical
- people in organisations
- politics and policy
- technologies for health
- terminology, classification and grouping
- uses of clinical information
- using informatics to support clinical healthcare governance
- computer systems applications in health (Toolkit).

Extraction of keywords from the available published index papers on health informatics using computer software packages and techniques

Scopus is the largest abstract and citation database of research literature and quality web sources with smart tools to track, analyse and visualise research. A search of Scopus was undertaken using a set of keywords that are descriptors of Informatics. The project's International Advisory Board agreed that the following key words should be used:

- health informatics
- medical informatics
- clinical informatics
- nursing informatics
- pharmacy informatics
- dental informatics.

The keywords within each article of the Reference Manager 11 database were exported as a series of files and then imported one at a time to an Excel spreadsheet as in the raw data format the total number of keywords extracted exceeded the number of rows available in an Excel worksheet.

After processing the data to count the number of occurrences of each keyword a master list of some 10 000 different keywords were identified, many of which were just English terms rather than health informatics specific, for example the authors place of abode and conference venue or country of study. The use of keywords in many publications depends on author choice and often reflects the wish to have the article seen as being in a particular theme or subject area. This is particularly so with those conferences that identify themes for the submission of papers.

This activity produced a new set of data and so triangulated with phase one of the project that also produced raw data. In itself it was preparatory work for the next two phases of the project.

Workshop to examine and exclude keywords

The next phase of the project refined and reduced the raw data by removing keywords not directly associated with health informatics. The lists of keywords were given to information specialists, grouped into teams of three, at a workshop in London, UK in January 2007. The groups considered each word and excluded any that were not thought to be a health informatics term. Each word was tagged with the number of occurrences it had in the search. At the same time, keywords were assessed to see if they would fit into the existing cognitive map from the phase one workshop (Table 1).

The participants in the workshop reduced the list of 10 000 words to 444. The number of occurrences found in the literature search ranked each keyword on the spreadsheet and small focus groups excluded words unconnected with health informatics. The remaining 444 words appeared to be connected with areas of health informatics as opposed to being just English words and phrases used to describe the content of the papers.

Voting in of keywords by international volunteers using a voting system based in an Excel spreadsheet

An Excel spreadsheet was constructed with a list of the keywords from which participants were invited to chose (vote in) those that were associated with health informatics. The complete spreadsheet together with the instructions and examples of how to vote was emailed to the International Advisory Board, the IMIA working groups, the BCS specialist

groups, and the European Federation for Medical Informatics (EFMI) working groups.

The voting was conducted with all of the keywords listed on the spreadsheet and a choice box next to each. The 444 keywords were divided into groups and each group was given a range of letters, A to G, H to M, N to R, and S to Z. Participants were asked to complete the group that contained the initial letter of their surname. Thus, as an example, Heather Carter voted on the columns A to G and Peter Ross voted on columns N to R.

Participants voted for about 100 words in their group. They were asked to vote for the keywords they thought were health informatics terms and classify them according to which phase one theme they thought the keyword belonged with by putting the number of the theme next to the word on the spreadsheet. Keywords that were consistently chosen were added to the original phase one cognitive map. These final two phases used methodological triangulation to refine the data and match it with the output of the first two phases: the phase one workshop and the international interpretation workshops.

Results

The final spreadsheet, which forms the basis of the IMIA Knowledge Base, was constructed from the outcomes of the original phase one workshop, the subsequent phase to check international interpretation, a review and content analysis of the literature, and a two-phase refinement following the extraction of keywords from the entire electronic published papers on health informatics. The different phases to the project in all took:

- data from different sources (people and electronic papers) – data triangulation
- used different research methodologies (workshops, electronic searches, electronic analysis, electronic voting) – methodological triangulation
- information from different investigators (one primary investigator, with five secondary investigators) – investigator triangulation
- from different theoretical positions (grounded theory, educational theory) – theoretical triangulation.

Conclusion

Through using mixed modes of research within and between the different phases of the project the investigators and subsequently the IMIA Board and General Assembly can be confident in the confirmation and completeness of the data through cross confirmation and validation from more than one data source

TABLE 1: Illustrating how keywords fit into the 'theme' and 'element' framework and the number of occurrences of each keyword in the literature.

Theme	Uses of clinical information	Number of times tagged in search
Element	Data analysis and statistical presentation	-
Keyword	Automatic data processing	78
Keyword	Analysis	635

Triangulation strengthened the project and ensured the validity and reliability of the project outcomes. The endorsement of the 'IMIA Knowledge Base' took place at the IMIA Board and General Assembly meetings of IMIA in July 2010. The final report and spreadsheet are available on the IMIA website in the section on IMIA Endorsed Documents (Wright 2009).

The initial outputs from phase one have been used in a number of ways including to help formulate an undergraduate biomedical informatics degree programme (Pritchard-Copley *et al.* 2006) and as a framework to classify scientific papers for the European Federation for Medical Informatics (EFMI) conferences.

Acknowledgements

The author wishes to acknowledge the invaluable contributions of the CHIRAD health informatics team and in particular Dr Helen Betts and Dr Peter Murray, the BCS, the IMIA, and the many colleagues in the international health and medical informatics communities who contributed to the various stages of the project.

Author competing interests

I declare that I have no financial or personal relationship(s) which may have inappropriately influenced me in writing this paper.

References

Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., *et al.*, 2001, *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*, abridged edn., p. 302, Longman, New York.

- Begley, C., 1996, 'Using triangulation in nursing research', *Journal of Advanced Nursing* 24, 122–128. doi:10.1046/j.1365-2648.1996.15217.x, PMID:8807387
- Bloom, B., Masia, B. & Krathwohl, D., 1984, *Taxonomy of educational objectives*, Longman, London.
- Bryman, A., 2003, 'Triangulation', *The Sage encyclopedia of social science research methods*, Sage, Thousand Oaks, CA.
- Cohen, L., Manion, L. & Morrison, K., 1994, *Research methods in education*, Routledge, London/New York.
- Denzin, N., 1970, *The research act in sociology: A theoretical introduction to sociological methods*, Butterworths, London.
- Eden, C. & Ackermann, F., 2004, 'Cognitive mapping expert views for policy analysis in the public sector', *European Journal of Operational Research* 152, 615–630. doi:10.1016/S0377-2217(03)00061-4
- Forehand, M., 2005, *Bloom's taxonomy: Original and revised*. Emerging perspectives on learning, teaching, and technology, 1–9.
- Furst, E., 1981, 'Bloom's taxonomy of educational objectives for the cognitive domain: Philosophical and educational issues', *Review of educational research* 51, 441. doi:10.2307/1170361, doi:10.3102/00346543051004441
- Lorenzi, N., 2007, 'Towards IMIA 2015-the IMIA Strategic Plan', *Yearbook of medical informatics*, 1–5.
- Murray, P., 2008, 'The IMIA strategic plan-towards IMIA 2015', *Yearbook of medical informatics*, 7–15.
- O'Donoghue, T. & Punch, K., 2003, *Qualitative educational research in action: Doing and reflecting*, Routledge, Abingdon, Oxon, UK.
- Patton, M., 2002, *Qualitative research and evaluation methods*, Sage, Newbury Park, CA.
- Patton, M., 1990, *Qualitative evaluation and research methods*. 2nd edn., Sage, Newbury Park, CA.
- Pritchard-Copley, A., De Lusignan, S., Rapley, A. & Robinson, J., 2006, 'Towards a benchmarking statement for biomedical informatics', *Current Perspectives in Healthcare Computing* 20–22.
- Seddon, G., 1978, 'The properties of Bloom's taxonomy of educational objectives for the cognitive domain', *Review of educational research* 48, 303. doi:10.2307/1170087, doi:10.3102/00346543048002303
- Strauss, A. & Corbin, J., 1998, *Basics of qualitative research: Techniques and procedures for developing grounded theory*, Sage Publications, Inc., Thousand Oaks, CA.
- Wolf, F., 2010, 'Enlightened Eclecticism or Hazardous Hotchpotch? Mixed Methods and Triangulation Strategies in Comparative Public Policy Research', *Journal of Mixed Methods Research* 4, 144. doi:10.1177/1558689810364987
- Wright, G., 2009, 'IMIA Knowledge Base' in *Imia knowledge base final 09* (ed.), viewed 03 February 2011 from, http://imia.org/endorsed/IMIA_Knowledge_Base-final09.xls