Qualitative Research JOURNAL

Volume 5 • No. 1 /2005

Refereed article: Qualitative Research Journal, Vol. 5, No. 1, 2005, pp. 5–14

Data quality issues in practice and theory: A cross-cultural example

Lucy Firth, *Department of Information Systems, The University of Melbourne*
David Mellor, *School of Psychology, Deakin University*
Jenny Pang, *Department of Information Systems, The University of Melbourne*

ABSTRACT

Practical considerations and traditions play a substantial role in data collection exercises, often limiting the focus of study to either qualitative or quantitative issues. An industry with a particularly strong quantitative emphasis is the insurance and reinsurance industry, where actuarial decisions are based on detailed and exacting numerical analysis of data that are assumed to be reliable and valid. However, the qualitative investigation of the quality of data in one reinsurance setting reported in this paper shows that where the meanings of the questions asked and of the answers provided are subject to interpretation, the quality of data collected for entry to databases can be poor. While this can be exacerbated in cross-cultural contexts, it is also generally true. Due to the constrained nature of insurance practice, the existence of a range of techniques combining qualitative and quantitative methods is somewhat academic. Therefore, because researchers have the latitude to investigate both qualitative and quantitative factors in the industrial context, a call is made for researchers and industry to work more closely together.

INTRODUCTION

There may be two distinct understandings of the choice between qualitative and quantitative methods—one used by researcher, the other by practitioners in industry. To the researcher, the fact that the methods come with antithetical epistemological frameworks and techniques, and derive from different views about how social reality can/ought to be studied (Bryman, 1988), suggests a choice. For the practitioner in industry, the focus on numerical outcomes in profit, output, quality assurance and standards may suggest a lack of choice. A perusal of books on managerial decision making in business finds an overwhelming focus on quantitative measures, while books on management research are increasingly qualitative. The focus on quantitative methods in business decision making may not only influence the type of studies undertaken, but may also mean that valuable information is overlooked.

Increasingly, partisan approaches are yielding to the good sense of combining quantitative and qualitative methods. Bryman (1988) draws on an array of literature to derive a list of possible ways in which the two may be combined. The first of these, triangulation, extrapolates the quantitative practices of comparing different measurements of the same element, the point being that qualitative methods can be seen as another way of achieving multiple observations on quantitatively measured phenomenon. The second way of combining the two approaches uses qualitative methods to facilitate quantitative methods by establishing preliminary concepts for measurement or scale development. The third combination uses qualitative methods to fill gaps in quantitative knowledge, thereby producing a single picture in which the qualitative findings sit side by side with quantitative findings. The fourth combination explores both researchers' and participants' perspectives to provide insight into actors' motives and understandings of their actions. This may add great power to studies of actions. The fifth way of combining the two methods uses quantitative methods to provide an indication of the significance of qualitative findings in a sample or population. This strengthens the generalisability of qualitative findings. The sixth way uses qualitative methods to facilitate the understanding of observed quantitative relationships between variables, by attempting to answer the question of 'why' variables correlate. The seventh combination ties qualitative macro findings to quantitative micro findings to provide understanding of the macro findings and to enable policy. The eighth combination enables research to progress by using qualitative methods to move from hunches to operationalisable, quantitatively measurable research propositions. Finally, Bryman suggests a hybrid combination of the two approaches, with research topics having both a qualitative and quantitative elements.

Langhout (2003) urges researchers to see qualitative and quantitative methods as a continuum, each extending the other. While researchers have the freedom to choose methods and to design projects accordingly, the practitioner in industry may rely on proven methods that become part of the culture. Where practitioners are constrained in the methods that they employ, their focus may be on one aspect of the problem—either qualitative or quantitative. However, without an external source indicating that this is the case, and that it could be remedied by a different approach, industry methodology may suffer from inertia.

One industry that relies almost exclusively on quantitative data is the insurance industry, which constructs complex and extensive databases of numerical data as the basis of its actuarial decision making (Dror, 2002). With a focus on quantitative attributes of data, there may be little attention to data quality (Shanks & Darke, 1998). In reality, the quality of the data contained in databases may be very important (English, 1999; Redman, 2001; Wand & Wang, 1996). Increasingly, insurance is a global business environment that requires data gathering, manipulation and interrogation in many varied cultural contexts. Many researchers (eg Ho, 1998; Mella, 1994; Tata, 2000) argue that there is great opportunity for loss of meaning in multicultural contexts. Nevertheless, there is little published work on the quality of data in insurance databases, and none in a multicultural context (Firth, Pang, Sampson & Shanks, 2003).

This paper sets out the theoretical basis of data quality, and then considers the importance of data quality in a case study example, namely the establishment of a database for reinsurance of community-based health insurance in a remote area in the

Philippines. The qualitative techniques for assessing the quality of that data are briefly described with a particular emphasis on the cultural context in terms of accuracy issues. The reliance on research methods norms by industry (in this case quantitative methods by insurers) is then discussed in the light of the findings.

DATA QUALITY ISSUES

Data quality refers to the fitness of data for its purpose (Shanks & Darke, 1998). In assessing data quality it is important to consider both the characteristics of the data themselves, and the assessments by users of the data. To assist in this, Shanks and Darke suggest three discrete but interrelated levels of data quality: syntactic, semantic and pragmatic. Syntactic data quality is concerned with the *structure* of symbols and focuses on the form of data rather than its meaning. The goal of syntactic data quality is consistency where data values for particular data elements use a consistent symbolic representation or coding scheme (Ballou & Pazer, 1995).

Semantic data quality concerns the *meaning* of data and focuses on how symbols are used to represent things in the real world. The goals of semantic quality are completeness, accuracy and currency (Tayi & Ballou, 1998; Wang & Strong, 1996). Completeness is concerned with the extent to which there is a one-to-one correspondence between data and things in the real world system. Accuracy is concerned with how well data represents states of the real world. Currency is concerned with how up-to-date the data is (this is different to timeliness which depends on how the data are being used).

Pragmatic data quality concerns the *use* (in terms of *useability* and *usefulness)* of data by stakeholders in performing their work tasks. It is therefore a concept that depends on who uses the information, for what purpose and in what context. Usability is the degree to which each stakeholder is able to effectively access and use the data. Usefulness is the degree to which stakeholders are supported by the data in accomplishing their tasks within the social context of the organisation. These three levels of data quality were used to assess the data quality of the Social Re project in the Philippines.

SOCIAL RE PROJECT

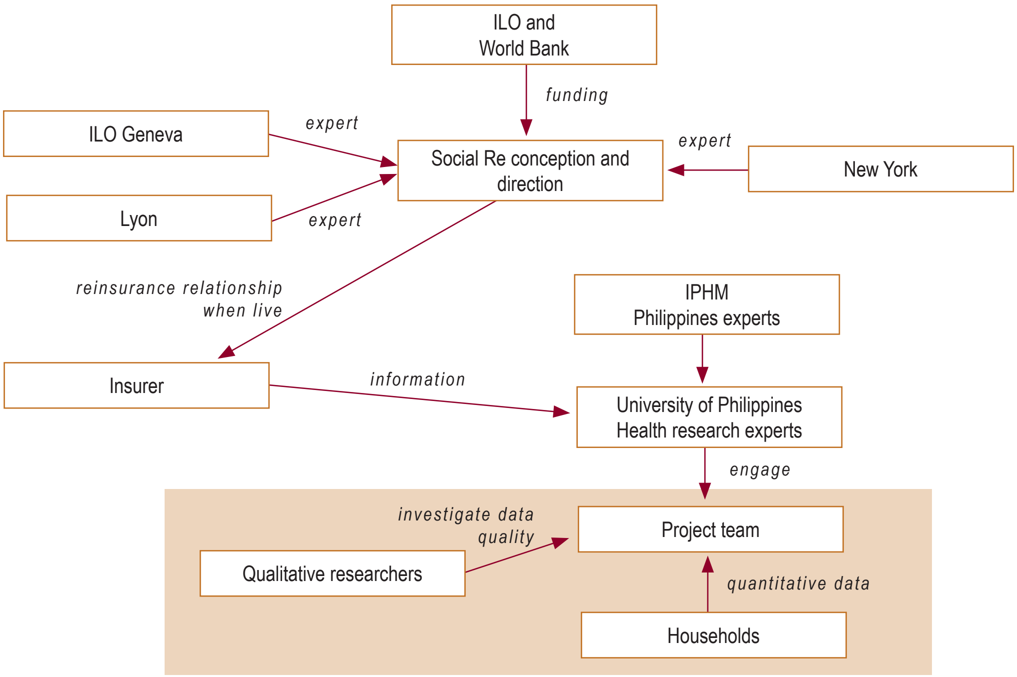
An ILO and World Bank sponsored program (Social Re) is currently piloting not-for-profit reinsurance of community-based health insurance in the Philippines. The *raison d'être* for Social Re is to break the vicious cycle in which poverty leads to ill health and to greater poverty and in turn to greater ill health. Health insurance would go some way to break this cycle, and indeed, over the last decade some communities have set up their own health insurance schemes. However, without reinsurance, the small numbers covered and the covariance between their health outcomes means that the community-based health insurance programs are financially unstable. Social Re plans to offer not-for-profit (social) reinsurance to community-based health insurance schemes in order to overcome that instability while applying insurance disciplines of actuary and underwriting (Preker, Langenbrunner & Jakab, 2002).

Preliminary to offering reinsurance, Social Re requires detailed baseline information on the community, its health histories and practices, and its capacity and willingness to pay. To offer reinsurance, Social Re requires detailed information on membership, payments received and benefits paid. This amounts to a considerable data stock in the baseline stage and data flow in the operations stage. These data must be provided by the community program in standardised format and detail if Social Re is to be able to use it.

In 2002, Social Re collected data for its baseline in five communities with health insurance schemes in the Philippines. That involved having households complete a quantitative questionnaire. The establishment of the baseline involved a complex of parties, relationships and processes, as indicated in Figure 1; some salient aspects are presented here. The data collection instrument, a 37-page questionnaire, was designed in Geneva (Switzerland) and Lyon (France) and translated from French into English. A Manila-based health policy organisation (IPHM) arranged for the questionnaire to be translated into Tagolog (the official language of the Philippines), and piloted the questionnaire at locations in and around Manila, and in some outlying stations. Subsequently, IPHM translated the questionnaire into the major language of each of the provinces in which the baseline was to be established. IPHM also arranged for a team of public health researchers at the University of the Philippines (UP) to undertake the fieldwork for the baseline including the household survey and analysis of records held by the community-based health insurance scheme (henceforth called the insurer). The UP group engaged local graduates with at least some research experience to form the Project Team and to administer the questionnaire in each community.

Because the population is largely illiterate, the questionnaires were administered verbally with the interviewer entering the responses on the questionnaire. The questionnaire covered issues on the economic status of the household, the standard of accommodation provided to the household, the names and ages of the household members, the health status of its members, their medical history and their health practices.

Figure 1: The main players and their roles and relationships in establishing the baseline



SOCIAL RE'S DATA QUALITY CASE STUDY

Social Re's quantitative focus and data collection practices are consistent with actuarial and underwriting decision making (Outreville, 2002). However, the question arose: Is the data collected by Social Re using this process of good quality? To investigate this, at the baseline stage, two independent qualitative researchers were engaged by Social Re to assess data quality issues arising from the household survey. The purpose was to focus on the semantic, syntactical and pragmatic data quality issues as perceived on the ground in the Philippines, and in the directing office in Geneva. These qualitative researchers (the first and third authors) observed the data collection process and interviewed the members of the Project Team before reporting back to Social Re on the data quality issues.

It is the case study of the investigation of the data quality of Social Re's qualitative data that forms the core of this paper. For clarity, we restate that two data collection processes were undertaken concurrently: the quantitative data collection through the household survey as the baseline for Social Re's database; and the qualitative investigation of that survey process.

Data collection for the establishment of Social Re's baseline for their database was undertaken in July 2002 in Guimaras, an island province, 3 km offshore from the city of Iloilo in the Visayas region of the Philippines. Sixty per cent of the Guimaras population of 84,870 live below the poverty level. The community health insurance scheme has 6000 member households paying a minimum of 85 pesos per annum (approximately AUD$3). For this, the members receive free medication associated with hospital stays of at least 24 hours duration up to a total cost of 4,700 pesos per annum.

The qualitative researchers (the independent investigators) travelled to Guimaras with the Project Team during the data collection stage of establishing the baseline. The qualitative researchers' objective was to assess the quality of the data being collected by the Project Team. Although quality control is often construed in a quantitative way (Juran, 1988), in this case the very concept of quality was understood to be essentially qualitative. Far from being concerned with a proportion of data entries that would be rejected by the system, as may be the case if the syntactic quality of the data was poor, we, the qualitative researchers were interested in the quality of the data from the perspective of the users. This implies its semantic qualities of meaning and accuracy and pragmatic qualities of fullness. Qualitative methods were called for because not only is the concept of quality essentially qualitative, but also the context was essential to the meaning and value of the data (Yin, 1993).

Cultural sensitivity is always called for in research (Gil, 1999; National Health and Medical Research Council, 1999). This was very true in this case where as qualitative researchers we were called upon to discuss with the Project Team the quality of the data they were collecting. One major concern was that the application of inappropriate techniques from the western paradigm of interviewing may have resulted in answers biased towards satisfying the needs of the interviewer. According to Ho (1998) the appropriate technique for collecting data in the Philippines is *Pagtanung-tanong*. *Pagtanung-tanong* is a Filipino concept of nonreactive, naturalistic, informal inquiry that complements Filipino culture. *Pagtanung-tanong* is the 'way things are done around here'. It avoids direct questioning by 'asking around'—that is, asking general questions to promote general discussion leading to a relevant issue, rather than direct questioning. Because *Pagtanung-tanong* is a general concept rather than a rigid technique, it could be modified to suit the cultural conditions of Guimaras and the experience of the Project Team. The investigators spent several days with the Guimaras Project Team, observing them while they administered the questionnaire, and subsequently 'asking around' about the questionnaire process and details about the data collected.

Utilising *Pagtanung-tanong* enabled two complementary processes to occur. It enabled the Project Team workers to reflect on their role and to reflect on the data that they were collecting from households. Reflecting and talking through the process of their data collecting and the nature of the responses from the households enabled the Project Team members to clarify niggling concerns that they had with quality issues. For instance, while all Project Team members had agreed upon a common translation of the questionnaire into the local language, many reported feelings of discomfort with the meanings and wordage to which they had agreed. Issues arising from the translation process are discussed further below.

In the informal process of talking over and talking around the base case establishment process, several data quality issues were identified. The full list of issues can be found in Firth et al. (2003). Here we present those related to accuracy (semantic data quality) that arose due to cultural differences that were not detected in the questionnaire development stage, the piloting stage or the translation stage, although each of these was done in an apparently rigorous manner.

FINDINGS RELEVANT TO THE DATA QUALITY

Accuracy issues

As indicated above, accuracy is one of the key goals of semantic data quality (Tayi & Ballou, 1998; Wang & Strong, 1996). Data accuracy is one of the main concerns to insurers and reinsurers because accumulated small inaccuracy can destroy the margins on which actuarial decisions are based (Vate & Dror, 2002).

The issues that caused lack of accuracy in the Social Re data were generally due to the multicultural nature of the Social Re project. Guimaras village life is a world away from that in Geneva. Some of the issues may be common to all of the Philippines, while some could be peculiar to regional populations, or to the population of one particular location. The Social Re team, based in Geneva, had an understanding that they were dealing with Filipinos and that the Filipino culture of Metro Manila was THE local culture. From 12,000 km away this looks reasonable. But Filipino provincial communities are culturally distinct.

While the ILO Geneva experts were keen to ensure that the questions were relevant to 'local' conditions, their inappropriate concept of 'local' threatened the integrity of the data. This occurred despite the fact that the Geneva experts and the Manila team (IPHM) made considerable efforts to include culturally relevant questions and response choices to the questions in the survey. It was only when the project team in Guimaras intervened that this threat was avoided. An example is the question 'How do members of your household dispose of their human waste?' A reflective process, as well as trialling of the questionnaire by the Project Team, enabled the following modification in terms of precoded options: 'use public toilet facilities', 'use neighbour's toilet facilities', 'wrap and throw', none of which may have been suggested from the western perspective, or even the Metro-Manila perspective on waste management. Without such modifications there would have been an unacceptably high incidence of 'other' requiring postcoding, and/or of selection of a precoded option that looked appropriate. While those modifications served well for that particular question, the following sources of inaccuracy were identified.

Translation of instrument survey

During training, the Project Team raised concerns about what it thought might lead to poor data quality. Perhaps most importantly, it indicated that the local people speak Karaya, not Ilongo into which the survey instrument had been translated. It was deemed too late to translate and print the instrument into Karaya. The solution was to discuss possible translations and to agree upon one that relayed the same meaning. It was agreed to administer an English version of the instrument, with the Project Team members relying on memory to achieve the correct translation into Karaya. At the stage of administering, the survey instrument had already been translated from French to English to Tagalog (for testing) to Ilongo, and then directly from English to Karaya on the spot. The potential for loss of meaning here was not only immense, but also difficult to identify without going through lengthy translation and etymological processes.

Conflicting definitions of terms

The insured unit is the household. But, in the 37-page instrument, the term 'household' was used apparently interchangeably with 'family'. Moreover, neither the Project Team members nor the interviewees seemed to be clear as to what was intended by either expression. This is not just an issue of translation (see above) but also of meanings in cross-cultural contexts. This became particularly apparent with the question 'Please indicate which of these illnesses members of your family have had in the past 5 years'. Having just listed the names and ages of all of those who lived in the house as household members, it was not clear if this question referred to them or to some other concept of family. Moreover, even if it is accepted that it is the household membership that is of interest, it was not clear if the question referred to the current members and the diseases they had over the past 5 years, irrespective of which household they were a member at that time; or to the people who were living in this house at the time of the disease. This is important in cultures where individuals move from household to household of extended family members to be close to school, or to employment, etc.

Culturally insensitive questions asked of the interviewee

Not only do questions vary in meaning cross-culturally, but also answers may vary according to power distance inherent in particular cultures (Hofstede, 1980). Therefore, questions and their answers vary in meaning in the context of local culture. For example, asking: 'If you ever had a complaint about your [insurance] package, do you know with whom to register your complaint?' appears to have been meaningless given the rural Guimaras cultural norm of not complaining to authorities. Moreover, Guimaras cultural norms of modesty render meaningless the responses to the question 'If I were to ask you to scale your household's level of wealth from one to 10 with one being very poor and 10 being very rich, what would be your household's scale?'

Inappropriate answer structures

In the same vein as the previous point, the use of Likert scales that ask members of agreement-prone cultures to indicate whether they strongly agree, agree, disagree, strongly disagree, are unlikely to elicit a sensible response.

DISCUSSION

In this paper we have argued that in the insurance industry the choice methods for collecting, recording and making sense of data are necessarily quantitative. However, because the quality of that data is essential to actuarial decision making, there is a clear case to assess the quality of the quantitative data. Of the three levels of data quality, two (semantic and pragmatic) are essentially qualitative, and can only be assessed in the context to which they relate. To do this, we have argued that qualitative methods are called for, particularly in the case study presented here, which demonstrates that the meaning of the data (semantic quality) can be culturally sensitive. However, industries dominated by quantitative concepts may not recognise the potential impact of qualitative issues on their operational capacity. While the insurance industry is keenly concerned with the quality of their data, their focus is at the syntactical level. Ideas of cultural meaning and context-based useability are a long way from the cut-and-dried world of the actuarial, but at least in this case are essential to the ability to make good insurance decisions.

In the case study presented here, the use of qualitative methods to assess the quality of data that is collected and analysed quantitatively is a departure from the normal practice in insurance. The insurer (Social Re) followed industry norms in establishing the baseline for their database using qualitative data collected qualitatively. Using these norms led to the development of a research instrument that was suited to the environment for which it was developed, but which proved ill equipped to capture quality data in the context of Guimaras. While it may be understood that Social Re's decision to engage the qualitative researchers reflects awareness that their norm approach may not elicit good quality data in the Guimaras context, it is not clear if Social Re would have similar concerns about data quality in developed contexts. What is clear from the case study is that introducing qualitative research to the quantitative process provided rich triangulation that not only enabled accuracy checks, but also checks on the meaning and usefulness of the data.

CONCLUSION

While there is now an abundance of literature calling for the marriage of qualitative and quantitative techniques to help us develop richer understanding of real-world phenomena, industry often has a culture of doing what has been done before, how it was done before. We suggest that industry needs concrete examples of the value of marrying qualitative and quantitative methods. Without examples that show the value of creatively and profitably bringing the two paradigms together, the tendency may be to see them as relating to different parts of the same story, with one part being far less relevant than the other. We argue that in this regard industry may require the support of researchers and academics who have become familiar with the role that qualitative data can play, while industry has not. In the case study presented here, we saw that Social Re took many steps to be culturally sensitive and to ensure the quality of its data, but nevertheless committed oversights. However, Social Re was open to the suggestion that quality in the broader sense is just as important as quantity of data to its operations and those that it hopes to help.

We propose that our finding relating to the use of qualitative methods to assess quality of data central to quantitative processes is generalisable in the sense suggested by Yin (1993). We have presented one case study, the establishment of databases for reinsurance, but our findings may be generalisable to similar contexts, such as the creation of databases for use by different stakeholders, and where the subjects to which the data relate may have different meanings for the various stakeholders, and the respondents from whom the data were collected. It remains the province of those who may wish to apply the findings and conclusion presented here to decide if the context is sufficiently similar.

References

Ballou, D. P., & Pazer, H. L. (1985). Modelling data and process quality multi-input multi-output information systems, *Management Science*, *31*, 150–162.

Bryman, A. (1988). *Quantity and quality in social research*. London: Unwin Hyman.

Dror, D. (2002). Health insurance and reinsurance at the community level, In D. Dror & A. Preker (Eds.), *Social Re-insurance: A new approach to sustaining community health financing* (pp. 103–124). Geneva: The World Bank and International Labour Organisation.

English, L. (1999). *Improving data warehouse and business information quality*. New York: Wiley Computer Publishing.

Firth, L., Pang, J., Sampson, J., & Shanks, G. (2003, June). Data quality issues in community health information systems in developing countries. *Proceedings of the IFIP Joint WG8.2/9.4 conference on IS perspectives and challenges in the context of globalisation*. Athens.

Gil, E. F. (1999). Culturally competent *research*: An ethical perspective. *Clinical Psychology Review, 19*, 45–55.

Ho, D. F. (1998). Indigenous psychologies. *Journal of Cross-Cultural Psychology*, *29*, 88–104.

Hofstede, G. (1980). *Cultures, consequences*. Beverly Hills: Sage.

Juran, J. (1988). *Juran's quality control handbook.* New York: McGraw Hill.

Langhout, R. D. (2003). Reconceptualizing quantitative and qualitative methods: A case study dealing with place as an exemplar. *American Journal of Community Psychology*, *32*, 229–244.

Mella, O. (1994). *Religion in the life of refugees and immigrants*. Stockholm: Janina Jassinska-Luterek.

National Health and Medical Research Council, (1999). *National statement on ethical conduct in research involving humans*. Canberra, Australia: National Health and Medical Research Council.

Outreville, J. (2002). Introduction to insurance and reinsurance coverage. In D. Dror & A. Preker (Eds.), *Social Re-insurance: A new approach to sustaining community health financing* (pp. 59–74). Geneva: The World Bank and International Labour Organisation.

Preker, A. S., Langenbrunner, J., & Jakab, M. (2002). Development challenges in health care financing. In D. Dror & A. Preker (Eds.), *Social Re-insurance: A new approach to sustaining community health financing* (pp. 21–36). Geneva: The World Bank and International Labour Organisation.

Redman, T. (2001). *Data quality: The field guide*. New Jersey: Digital Press.

Shanks, G., & Darke, P. (1998). Understanding data quality in data warehousing: A semiotic approach. In I. Chengilar-Smith, & L. Pipino (Eds.), *Proceedings of the. MIT conference on information quality* (pp 247–264). Boston: MIT.

Tata, J. (2000). Toward a theoretical framework of intercultural account-giving and account evaluation. *International Journal of Organizational Analysis*, *8*, 155–178.

Tayi, G. K., & Ballou, D. P. (1998). Examining data quality. *Communications of the ACM*, *41*(2), 54–57.

Vate, M., & Dror, D. (2002). To insure or not to insure? Reflections on the limits of insurance. In D. Dror & A. Preker (Eds.), *Social re-insurance: A new approach to sustaining* *community health financing* (pp. 125–152). Geneva: The World Bank and International Labour Organisation.

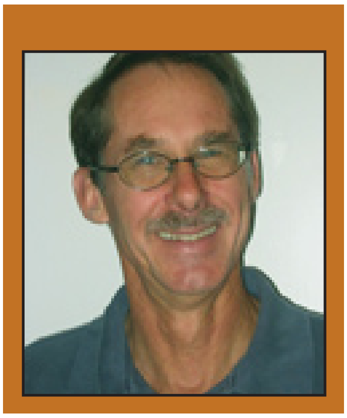
Wand, Y., & Wang, R. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86–95.

Wang, R., & Strong. D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–34.

Yin, R. (1993). *Applications of case study research*. Newbury Park, CA: Sage.

LUCY FIRTH currently lectures in oganisational change, innovation and entrepreneurship and management information systems. Previously, Lucy worked with the United Nations (Geneva) and with the Economics of Infrastructures Group (Delft). With a PhD in economics Lucy's research interests include IS/IT for development and use of IS/IT in the health industry.

Contact details:
Dr Lucy Firth
Department of Information Systems
The University of Melbourne, Victoria 3010
Australia
Phone +61 3 8344 1523
Email: lfirth@unimelb.edu.au



DAVID MELLOR currently chairs the postgraduate courses in clinical psychology at Deakin University and contributes to a variety of other courses. Having been a practising and academic psychologist for many years, David's main research interests now are the impact of the impact of the Internet on older Australians' sense of wellbeing, childhood disorders and political psychology.

Contact details:
Dr David Mellor
School of Psychology
Deakin University, Victoria 3125
Australia
Phone: +61 3 9244 3742
Email: mellor@deakin.edu.au

JENNY PANG is a recent graduate in information systems. Having completed an internship with the ILO in Geneva, Jenny is now undertaking further studies.

Contact details:
Jenny Pang
Department of Information Systems
The University of Melbourne, Victoria 3010
Australia