Information Systems: Transforming the Future

24th Australasian Conference on Information Systems, 4-6 December 2013, Melbourne

Proudly sponsored by

nab
RMIT University

Citrix
GS1 Australia
ACS Australian Computer Society
ACS Foundation
ACPHIS
AAIS
Data Quality in Community Health: An Indonesian Study

Maulidyati Aisyah
Sophie Cockcroft
UQ Business School
University of Queensland
Australia
Email: sophie@business.uq.edu.au

Abstract

We present the results of an exploratory study of data quality within a health centre (HC) in rural Indonesia. This study is carried out via two channels, first, assessing data quality using two well-known sets of qualitative data quality measures and second via content analysis of interview transcripts using Leximancer. Data is collected via semi-structured interviews based on the AIMQ and PRISM methods. PRISM has been used in a number of comparable studies in developing countries. A number of themes are identified. We document the challenges faced by the front line staff as a result of data quality issues and provide a snapshot of the state of play in a typical rural health facility in Indonesia. Suggestions are put forward for development action both locally and nationally.

Keywords
Health Information Systems, developing countries, PRISM, AIMQ, Leximancer, Case study

INTRODUCTION

Despite much local development work, health care services in Indonesia have been assessed as “poor” by international standards (OECD 2010). Previous studies (Husada and Nguyen 2012) have identified low health literacy, patient doctor ratio, incorrect diagnosis and long ordering times as motivation for providing better health information systems.

Health Service delivery has become a major concern for the Indonesian government. The Indonesian Ministry of Health (MoH) has established almost ten thousand Pusat Kesehatan Masyarakat/primary health centres (PUSKESMAS/PHC) in various districts all over the country (Husada and Nguyen, 2012). The main purpose of these PHCs is to help rural people to access health services. In 2002 the ministry enacted an order which instructed PHCs to adopt information systems called Sistem Informasi Kesehatan/Health Information Systems (SIK/HIS). Since then, various software applications have been developed and implemented, for instance simpus, e-puskesmas, and simkes (Liebscher and Hui 2007, e-Puskesmas 2011). The current study examines a health facility that is complementary to the PHCs, and is privately funded. The funding body, Badan Amil Zakat Nasional (BAZNAS), based on Islamic Philanthropy, and endorsed by the Indonesian government, runs five health centres (HC) around Indonesia.

The focus of BAZNAS has been to provide programs to meet life essentials such as health and education. HC, as one of fund distribution channels, delivers free health services using a membership system and is known as no-cashier healthcare. HC also provides an “outdoor”, or roving health services. The focus of the initiative has been management of chronic non-communicable disease, and child and maternal health including immunisation and family planning. For BAZNAS, in delivering these programs, information systems issues are secondary to those of health service delivery. The research questions examined here are

RQ1 “What is the current situation with respect to data quality in a typical rural health care setting in Indonesia?”

RQ2 “What specific operational issues arise from data quality problems?” and

RQ3 “What are the possible remedies both locally and nationally from a data management and policy viewpoint?”

It should be noted that whilst the items described in Table 1 provide a snapshot, it is the richer qualitative data captured through the interview data that enabled us to address RQ2 and RQ3.
QUALITY MEASURES

As a means of assessing HIS in developing countries, previous studies have used the Performance of Routine Information System Management (PRISM) framework (Aqil, Lippeveld and Hozumi 2009; Krishnan et al., 2010). PRISM claims data quality as one of its emphases and provides a basic set of qualitative measures in the form of questions with largely yes/no responses. Aqil, Lippeveld and Hozumi (2009) summarized studies applying PRISM framework and its tools in Uganda, Pakistan, China, and Mexico. It is argued that all of the results are consistent among these countries where low data accuracy was explained by higher perceived confidence level in checking data accuracy, yet much lower knowledge of the methods available to check the data. Meanwhile, Krishnan et al. (2010) conducted a study in India in which the PRISM framework was adopted in order to categorise HIS evaluation into input, process, and output stages. To expand the data quality framework we also use measures from the AIMQ framework (Lee et al 2002). AIMQ is a validated, multidimensional information quality assessment framework that has found it application largely in mainstream management information systems studies.

Both PRISM and the AIMQ methodology have derived a set of validated measurement items. The use of these items in our interview protocol adds to the body of knowledge from studies using these measures. The aggregated set is illustrated in Table 1

Table 1 Measurement items drawn from PRISM and AIMQ (AIMQ items generally require True/False responses)

<table>
<thead>
<tr>
<th>Concept</th>
<th>Measurement Item</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness</td>
<td>CP1 How many data items does the facility need to report on in the Rural Health Information Systems (RHIS) monthly report?</td>
<td>PRISM</td>
</tr>
<tr>
<td></td>
<td>CP2 Is an official record of management meetings maintained?</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>CA1 This information includes all necessary values</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>CA2 This information is incomplete.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>CA3 This information is complete.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>CA4 This information is sufficiently complete for our needs.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>CA5 This information covers the needs of our tasks.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>CA6 This information has sufficient breadth and depth for our task.</td>
<td>AIMQ</td>
</tr>
<tr>
<td>Relevancy</td>
<td>RP2 Does the facility's database provide comparisons among different service types?</td>
<td>PRISM</td>
</tr>
<tr>
<td></td>
<td>RP3 During the last three months, did the facility receive any feedback report from higher authority on their performance</td>
<td>PRISM</td>
</tr>
<tr>
<td></td>
<td>RP4 Have any action-oriented decisions been made based on management reports?</td>
<td>PRISM</td>
</tr>
<tr>
<td></td>
<td>RP5 Does the facility have routine meetings for reviewing managerial or administrative matters</td>
<td>PRISM</td>
</tr>
<tr>
<td></td>
<td>RP6 Does facility receive annual/monthly planned targets based on the report</td>
<td>PRISM</td>
</tr>
<tr>
<td></td>
<td>RA1 This information is useful to our work.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>RA2 This information is relevant to our work.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>RA3 This information is appropriate for our work.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>RA4 This information is applicable to our work.</td>
<td>AIMQ</td>
</tr>
<tr>
<td>Timeliness</td>
<td>TP1 Observe whether report is submitted by the deadline</td>
<td>PRISM</td>
</tr>
<tr>
<td></td>
<td>TA1 This information is sufficiently current for our work.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>TA2 This information is not sufficiently timely.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>TA3 This information is not sufficiently current for our work.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>TA4 This information is sufficiently timely.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>TA5 This information is sufficiently up-to-date for our work.</td>
<td>AIMQ</td>
</tr>
<tr>
<td>Accuracy</td>
<td>AP1 Did you receive a directive in the last three months from the management to check the accuracy of data at least once in three months?</td>
<td>PRISM</td>
</tr>
<tr>
<td></td>
<td>AA1 This information is correct.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>AA2 This information is incorrect.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>AA3 This information is accurate.</td>
<td>AIMQ</td>
</tr>
<tr>
<td></td>
<td>AA4 This information is reliable.</td>
<td>AIMQ</td>
</tr>
</tbody>
</table>
IS MATURITY

Considering the whole organisation and in respect to RQ2 is it likely that operational issues may arise through aspects other than data/information quality, significantly, as a result of process maturity within the organisation. The notion of IS maturity in system development began in the 1970s (Nolan 1973) when it was recognised that having a stable and repeatable software development process was desirable in terms of developing quality software and as a measure of the professionalism of the developers. “Maturity” as a concept has developed beyond the software engineering field and is applied to e-business (Zumpe and Ihme 2006), and ERP (Niels and Mouritsen 2005) as well as more recently in HIS settings. Pare and Sicotte (2001) developed a model of Information technology sophistication for the hospital environment that measured sophistication in administrative, Patient management and clinical support activities. Van de Wetering and Batenburg (2009) developed a model of maturity specifically relating to Picture archiving and communication systems in hospitals. Another e-healthcare maturity model was introduced by Liu, Hwang and Chang (2011). They proposed multidimensional maturity measurement model which comprised six associated dimensions: IT infrastructure, application portfolio, integration, IS planning and control, user awareness, and IS staff. Whilst it is beyond the scope of this study to develop a maturity model specific to Primary care/ General Practice or healthcare settings we can draw from these earlier studies to recognise that a enterprises at all levels go through phases of maturity with respect to Information Systems and their strategic use specifically starting with automating existing process and leading to mature data management, integration with internal and external systems and the ability to properly meet the information needs of the organisation.

RESEARCH METHODOLOGY

This study focuses on the dimensions commonly referred on data quality research (Wand and Wang 1996; Pipino, Lee and Wang 2002) these are including completeness, accuracy, relevancy and timeliness. Data and information are often used synonymously.

In accordance with data quality studies, previous research in health information systems described data quality into similar dimensions. Using PRISM led to an emphasis on routine health information system (RHIS) performance which operationally defined as improved data quality and continuous use of information (Agil, Lippeveld and Hozumi 2009). Within PRISM, several tools have been introduced to evaluate routine health information systems comprehensively. Prior studies which adopted PRISM show an assessment on national level of HIS which each health centre is on a hierarchy with other health services. This hierarchy commonly put primary health centre in district level, those which are responsible to a public hospital in regional level, province level and centralized nationally under the Ministry of Health (Hotchkiss et. al. 2010; Krishman et al., 2010; MEASURE, 2010a).

Site, Sample, and Data Collection

A survey in a form semi-structured interview is preferred as data collection method in order to gain a deeper understanding, and provide a richer set of data for analysis. This approach follows Dul and Hak (2008)

The interviews were performed in the native language of the participants. Nine interview participants were identified through the organisational gatekeeper. These individuals were individually approached in person while on the job and asked to voluntarily participate in an interview about their perception on data quality. The participants are the staff of the health care centre which consists of manager, five (5) nurses, two (2) front officers, and a surveyor. No one of the participants have an IS background and only few who have been experienced in operating HIS as end-user.

For a the full validated questionnaires the reader is referred to PRISM (Measure, 2010b) and AIMQ (Lee, et.al., 2002) Table 1 shows the data items under review. In order to cover the dimensions addressed by the two aforementioned studies, but still attain the richness of a semi-structured interview A list of seed questions were developed to encourage participants to elaborate whilst at the same time addressing the measured items.

Each interview was recorded digitally and transcribed. These interview transcripts were independently translated by two native speakers and then checked for commonality. Each interview was placed in an individual file, and then the files were grouped in folders according to the four roles within the health centre (Surveyor, Nurse, Front Office Staff, Manager, Surveyor). User tags were added to make it easier to see which remarks related to which seeding questions, specifically tags were added to identify Participants. The text was analysed using Leximancer (Smith 2005).

The underlying theory of Leximancer is content analysis, it is a research tool used for determining the presence of words or concepts in collections of textual documents (Smith 2005). It is further argued that one of Leximancer strength is its ability to conduct both conceptual and relational analysis. In conceptual analysis, Leximancer was programmed to obtain specific terms for and set dictionary from these terms. In other words,
Leximancer has the ability to make an inference from the concepts within the text. Meanwhile, in relational analysis concepts which tend to appear together in the text are extracted. A visualisation will then display the relationship between these concepts. Terms extracted in Leximancer are weighted based on the frequency of occurrence in sentences. Once the weight reaches a threshold, the terms will be defined as a concept. Besides providing quantitative means for analysing text, Leximancer also allows the information drawn to be presented in a Concept Map. It is a map where co-occurrence concepts will be put near to each other and clustered into higher-level “themes”. As part of the data preparation certain modifications were made as per table 2.

Table 2: Modifications made to Leximancer default settings to automatically analyse interviews with Health Centre employees.

<table>
<thead>
<tr>
<th>Modification</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatically identified concepts manually removed</td>
<td>Example, Further, Provide</td>
</tr>
<tr>
<td>Concepts merged</td>
<td>Patient, patients</td>
</tr>
<tr>
<td></td>
<td>Nurse, nurses</td>
</tr>
<tr>
<td>User Defined Tags Added</td>
<td>Complete/ Completeness, Relevancy, Timeliness, and Accuracy, Medical Record*</td>
</tr>
</tbody>
</table>

In the initial analysis it was found that Leximancer was not distinguishing adequately between the noun and the verb use of the word “Record”. Thus all occurrences of medical record were concatenated in the original file as medical_record.

RESULTS

The results of analysing interview data are presented first via a leximancer analysis of the raw interview data and second using the qualitative assessment measures (AIMQ and PRISM).

It emerged from discussions with key personnel that there is an implicit data structure which has evolved organically and with little input from Information Systems literate staff. This structure is illustrated in Figure 1.

Figure 1: Data Structure (Table titles as given in HC spread sheets)
The focus of the Health centre and the wider mandate of BAZNAS has been to provide health care to less advantaged people in rural communities in Indonesia. This service includes and “outdoor” or on the road service. The focus on IT has only stretched as far as getting the appointment records into digital form and maintaining consistency between paper and electronic records. We see this is remarks such as

“...what written manually is exactly the same as what we input into computer. Our first process is we search patient Medical record on the cabinet then we input name and address manually. If we compare to hospital, once we input Medical record number we can retrieve all patient’s data. Unfortunately we still input the data manually because there is no good system running so far.” FrontOfficer1

In terms of IS maturity this would be defined as automation (Zumpe and Ihme, 2006). The record that HC stores, has evolved in an ad hoc fashion around three spreadsheets. About 1000 visits are recorded per month.

This structure, which has served the health centre for the first two of years of operation is starting to cause problems. This is revealed in the interview data. Specifically the spreadsheets are used to store Age instead of date of birth, so this has to be recalculated on every visit, and is inaccurate for reports. The membership card is fraught with difficulty since there are many family members on one card, the card does not have a photograph, and procedurally there is no way of updating a member’s marital or change in name, or the fact that they have become adult. Medical record number has to be reconciled across two spreadsheets, the gender of children is not stored.

LEXIMANCER RESULTS

Figure 2 shows the breakdown of themes showing speaker tags only. File and Folder tags are omitted because they cause unnecessary crowding. There is a clear division of concerns between the front office staff, surveyor and managers, tagged in the top of the diagram and the nursing staff tagged to the bottom. Themes are heat-mapped hot colours denote more important themes (in this case health, data, report) and green or blue shades the less important ones - in Figure 2 the terms meeting, computer, and case take less prominence.
Table 3 Summary of Leximancer themes and major concerns emerging from interview data

<table>
<thead>
<tr>
<th>Related concepts</th>
<th>Most frequent Role (Figure 2)</th>
<th>Major concern</th>
<th>Sample Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>data, patient record</td>
<td>All</td>
<td>Various, from membership validity at the higher levels (Manager and Front Officer), to problems of health record numbers in relation to service provision at the nursing level</td>
</tr>
<tr>
<td>Report</td>
<td>report, monthly, age, visit management</td>
<td>Nurse 5</td>
<td>Making sure data is accurate to produce necessary reports particularly with respect to age, needing to make monthly reports to monitor specific diseases such as diabetes and blood pressure</td>
</tr>
<tr>
<td>Health</td>
<td>health, family, name, information</td>
<td>Nurse 5</td>
<td>Ensuring data allows smooth healthcare delivery</td>
</tr>
<tr>
<td>Member</td>
<td>Member, card</td>
<td>Front Officer 1, Surveyor</td>
<td>Managing membership data Problems of incorrect or missing member numbers, accuracy of data on card</td>
</tr>
<tr>
<td>Nurse</td>
<td>Nurse</td>
<td>Nurse 2, Nurse 3</td>
<td>Anomalies that occur between the nursing station and front desk. For example a person asks to the nurse to see their child as well</td>
</tr>
<tr>
<td>Services</td>
<td>Services</td>
<td>HCManager, Surveyor</td>
<td>Making sure the right people get access to services and that the correct information is stored about them</td>
</tr>
<tr>
<td>Computer</td>
<td>Computer</td>
<td>Nurses 2,3 5 and Front office 1</td>
<td>How electronic and paper data are kept current, back up procedures. In particular in relation to a patient visit booking, and family planning services</td>
</tr>
<tr>
<td>Meeting</td>
<td>Meeting</td>
<td>All</td>
<td>Discussions of what goes on in monthly evaluation meeting. Meeting performance targets</td>
</tr>
<tr>
<td>Case</td>
<td>Case</td>
<td>Front officer 1 and 2</td>
<td>Discussing cases where accuracy had been compromised. Suspicions were raised in the “case” of someone being married but still operating on a parent’s card, or someone arriving at the clinic in a car.</td>
</tr>
</tbody>
</table>
Table 3 gives a breakdown of the type of issues discovered in relation to each to the Leximancer themes.

QUALITY MEASURE RESULTS

The data items given in table 1 are illustrated in Figure 3. Where a bar appears the survey group has said “yes” to the question. Three of the survey items CA2, TA2, and AA2 were stated negatively and are thus illustrated negatively in the figure. It can be seen that participants were unanimous in three of the responses TP1 - the fact that reports are submitted on time, CP2 – an official record of meetings is kept, CA2 – that information is incomplete and AA2 - that information is incorrect. In all the other dimensions there was some difference on opinion and in many cases only one group responded positively. The items RA4, TA3 and AA3 concerning applicability, timeliness, and reliability of data were all answered in the negative. The responses to RA4 are given below for illustrative purposes, in this case only the surveyor found the data reliable.

Reliability (RA4)

*it is still not as expected. Since one of our obstacle is that the system is offline, manually input the data, counting changes on age will be a challenge. I believe that not all the nurse pay attention on the date of birth and age changes.* Manager

*Surveyor data is reliable in supporting health care program. It uses membership data I managed, for instance program for elderly people, program on certain region and other certain program* Surveyor

*Yes, but in the examination we will check medical history through health record to deliver correct drugs to the patient.* Nurse

*There is no patient who informs us voluntarily if for example his child has married or worked which effect their membership status. I will be suspicious when a patient within marriage age but he still include in his parents card. Then I will confirm about his marital status and often the patient admit that he is married.* Front Officer

DISCUSSION

It is concluded that the data quality in HC is still far from complying with the items measured in Table 1. The results are consistent with a previous PRISM study Aqil, Lippeveld and Hozumi (2009) in the sense that there is an element of overconfidence in the data, and a lack of knowledge on how to detect problems. The data quality issues are likely also explained by status of IS maturity level for HC. The utilization of IS within HC has reached the point where manual record is transferred into digital form. However, staff still need to maintain the consistency between paper and digital records. As stated by Zumpe & Ihme (2006), the stage where human component are excluded from a process to some extent is called level of automation. This is the first level of e-business IS maturity model. This status explained HC’s low data quality compliance to the measurements. Understanding current stage of maturity helps to identify strategic action to be made in order to pursue the next maturity level (Liu, Hwang, & Chang, 2011). Development of a maturity model specific to primary health care in developing countries is a possible avenue for further research.
Specific problems within the HC were identified. Several staff found difficulties in reporting due to the many manual approaches applied to softcopy data such as in calculating and categorizing age. Concerns about fraudulent membership was also revealed by the Leximancer analysis. Since many staff found irregularities in patient data, this content appeared in a number of themes in the Leximancer concept map. Repeated statements on data inaccuracy that effect on recording and reporting processes resulted in a number of concepts emerging in the “data” theme.

Moreover, the Leximancer analysis identified isolated error events specifically, when FO used the wrong medical record for a patient. This event is a significant one in a healthcare setting. Identifying such case might guide the management to put more attention to data management.

A further contribution of this study is to the practice of using Leximancer in content analysis.

Recommendations

By considering the results on current study, several recommendations RQ3 are suggested to the BAZNAS HC as follows:

- In pursuing better data quality, HC should acquire better IS. This can be done through software development. However, this development should be accompanied by caution on crucial feature for example data checking and data validation to avoid use wrong medical record and fraud.
- Alternatively, HC could buy a software package which readily available in the market. Through this acquisition, HC might not necessary to spend resources as much as in software development.
- As part of any new development HC could consider integration with data in the government based PHC schemes
- Establishing formal standard operating procedures (SOP) including data flow will help HC to encounter issues on health.
- Human resource training is necessary for staff, including management to improve literacy on IS which expected will minimize error.

Ultimately, this study presents a snapshot of the current state of data quality in small to medium sized health facilities, which might serve as a benchmark for future development –RQ1.

Limitation and future research

During the research process, we documented some limitations:

- Interviews were conducted by phone. Thus, it is possible that messages were misinterpreted during the session due to the lack of physical cues.
- It is possible that some of the team interviewed were either defensive of their system, or thought interviewer was looking for a particular answer which could potentially lead to bias.
- Interviews were transcripted and analysed in different language. This might reduce the nuance of the words chosen.
- The research was limited to one type of health care provider in Indonesia.
- Limited exploration on IS maturity model in its relation to data quality.

CONCLUSION

The work presented here represents a snapshot of data/information quality and procedures within a rural health care setting in Indonesia. It is envisaged that a full presentation of the results specifically participant responses to the quality measures would be given in a presentation.

REFERENCES


