Review

Data quality challenges in the UK social housing sector

Caroline Duvier[a,c,*], Daniel Neagu[b], Crina Oltean-Dumbrava[a], Dave Dickens[c]

[a] School of Engineering, University of Bradford, UK
[b] School of Electrical Engineering and Computer Science, University of Bradford, UK
[c] Incommunities Group Ltd., UK

ARTICLE INFO

ABSTRACT

The social housing sector has yet to realise the potential of high data quality. While other businesses, mainly in the private sector, reap the benefits of data quality, the social housing sector seems paralysed, as it is still struggling with recent government regulations and steep revenue reduction. This paper offers a succinct review of relevant literature on data quality and how it relates to social housing. The Housing and Development Board in Singapore offers a great example on how to integrate data quality initiatives in the social housing sector. Taking this example, the research presented in this paper is extrapolating cross-disciplinarily recommendations on how to implement data quality initiatives in social housing providers in the UK.

1. Introduction

1.1. Social housing – regulations

Although many researchers and organisations preach the benefits of high quality data (Strong, Lee, & Wang, 1997; Redman, 2013), the UK social housing sector has yet to join such data-driven developments, particularly in light of tightening government regulations (Leach, 2016). The government implemented a regulatory framework that requires social housing associations to comply with standards related to economic and consumer issues, and social housing associations can be downgraded when they do not comply with regulations (HCA, 2015). The Department of Environment, Transport, and the Regions (DETR) also highlights the need for better business strategies for social housing associations (DETR, 2000a). Further to this, social housing providers need to follow building standards (HMG, 2010a, 2010b, 2010c, 2010d) that respond to national targets of cutting carbon emissions (HMG, 2008), and therefore need to know exactly what kind of stock they have and need. Data quality can help in this endeavour to establish a strategy for the retrofitting of the stock that does not meet these standards.

1.2. Data quality – why is it important?

To illustrate the importance of data quality, it might be relevant to discuss effects of poor data to an organisation. Tee, Bowen, Doyle, and Rohde (2007) mention two particular cases of poor data, one involving Hudson Foods losing its biggest client, Burger King, to an Escherichia coli bacteria contamination in 1997, which led to Tyson Foods acquiring Hudson Foods as a result. Data was not properly recorded about which batches contained the bacteria, leading to 25 million pounds of meat being recalled. The second case involved the Los Angeles County's pension system, and $1.2bn in unforeseen liabilities. Two calculation errors spanning over 20 years were responsible for these liabilities, resulting in the county spending $125 m each year over the next 50 years in order to make up for the shortfall. According to estimates by the Data Warehousing Institute (TDWI), poor data cost businesses in the U.S. $611 billion yearly by 2002 (Eckerson, 2002); currently the cost is $3.1 trillion yearly in the U.S, according to an IBM estimate (Quintero et al., 2015, page 192). These cases show that poor data can have serious implications, whether it is a public or private organisation that collects and uses the data. Getting to some level of detail within an organisation, a survey by Harris Interactive in 2006 showed that 75% of information workers think their data cannot be trusted and said they made wrong decisions due to incorrect data, leading to bad business decisions, loss of productivity, and job-related stress. They also report that around 30% of their time is spent checking data for correctness (Heinrich, Klier & Kaiser, 2009).

These examples highlight how important good data is for an institution nowadays. This paper will focus on data quality in social housing. First, a critical review of data quality research in social housing is presented, followed by specific strategies that can be applied in the sector to ensure high data quality, with an example of one social housing provider in the UK. The paper concludes with recommendations for the future of social housing organisation in this era of (big) data.

* Corresponding author at: University of Bradford, Faculty of Engineering and Informatics, School of Electrical Engineering and Computer Science, Bradford, BD7 1DP, UK.
E-mail address: cduvier1@bradford.ac.uk (C. Duvier).

2. Critical review of data quality in social housing

This section reports on work to date on data quality with impact or applications in the social housing data domain, and uses it as a platform to identify current challenges that affect progress in data-driven applications.

2.1. A brief history of data quality

Data quality largely depends on those who use data. A company selling beer will have very different data quality criteria than a health care provider specialising on geriatrics, for example. However, data quality is important to both organisations. Lee, Pipino, Fund, and Wang (2006) highlight four rationales for organisations to add data quality on their priorities list, namely that high quality data can increase customer satisfaction, improve revenues and profits, can be a strategic competitive advantage, and generally, is a valuable asset.

Traditionally, data quality in its various forms has been applied to the private sector to achieve goals such as improving decision making (Shankaranarayan, Ziad & Wang, 2003), supply chain management (Hazen, Boone, Ezell, Jones-Farmer, 2014), customer relationship management (Bose, 2002), or to comply with regulatory requirements (Hüner, Schierning, Otto & Österle, 2011). The health care sector is increasingly making use of research into data quality, for example to improve patient care (Mikkelsen & Aasly, 2005; Fung, Lim, Mattke, Damberg & Shekelle, 2008, Rizi & Roudsari, 2013), and validate the quality of modelling techniques in toxicology (Fu, Wojak, Neagu, Ridley & Travis, 2011; Yang, Neagu, Cronin, Hewitt & Enoch, 2013; Palczewska et al., 2013) and in engineering (Oltean-Dumbrava, Watts & Miah, 2014 and Oltean-Dumbrava, Watts & Miah, 2016). International standards for data quality have also been introduced in 2015, like the ISO 8000: 8–2015 standard (ISO, 2015).

According to Redman (2013), the most successful organisations define data quality as "meeting the most important needs of the most important customers", which implies an understanding of both the customer's needs, and the customers themselves. Data quality is usually considered in three approaches: organisational, architectural, and computational, and all three aspects should be considered for a successful data quality project (Sadiq, 2013). The organisational aspect involves business analytics, where objectives, strategies, processes, roles, policies, and standards for an organisation are established. The architectural aspect focuses on building how and where data is stored and analysed, which can be understood as the technology behind the policies and standards established through the organisational data quality aspect. And finally, the computational aspect offers the solutions to data questions and provides the statistical and analytical capabilities needed to evidence the organisational data quality aspects.

Data quality crops up in a large number of disciplines, making an overarching theme difficult (Sadiq, 2013). Different frameworks have been established, one of the most notable ones by Wang and Strong (1996), with more than 3000 citations to date in the academic community, outlining criteria important to establishing data quality. Wang and Strong (1996) argue that data quality criteria should be established by the data consumers themselves, and they established the first framework created through factor analysis. Their 20 data quality criteria are organised in four categories: accessibility of data quality (accessibility, access security), contextual data quality (value-added, relevancy, timeliness, completeness, appropriate amount of data), intrinsic data quality (believability, accuracy, objectivity, reputation), and representational data quality (interpretability, ease of understanding, representational consistency, concise representation).

A further overview of the data quality research arena characterises data quality as a two-dimensional framework with topics and methods Madnick, Lee, Wang & Zhu (2009). Topics encompass four major themes, each with a number of subthemes: (1) data quality impact, (2) database related technical solutions for data quality, (3) data quality in the context of computer science and IT, and (4) data quality in curation. Methods encompass both quantitative and qualitative research methods, such as case studies, mathematical modelling, system design, and data mining. Madnick et al. (2009) suggest employing more than one method, and using both qualitative and quantitative methods.

2.2. A succinct review of data quality in social housing

This section reports results of an academic literature search that was conducted on data quality in social housing. The search included the key words "data quality"; "housing"; "social housing"; "data management"; and was carried out using google scholar to get the broadest reach. Several case studies were found on data warehousing; knowledge management; business process reengineering; and customer relationship management at the Housing and Development Board (HDB) in Singapore (Ang & Teo, 2000; Thong, Yap, & Seah, 2000; Teo, 2005; Teo, Devadoss & Pan, 2006); and a comparison between Dutch and English housing associations on asset management strategies (Gruis, Nieboer & Thomas, 2004). However; apart from asset management; academic studies on data quality in UK social housing providers are still lacking and demonstrate that there is a current knowledge need of closing the gap with general progress in data quality research for businesses in this domain.

2.3. Issues with data in social housing

Leach (2016) discusses the difficulty of analysing data of UK social housing providers due to the fragmentation of data into various databases and spreadsheets, lack of consistent collection strategies across organisations, or the sheer lack of data. The detailed case studies of the Housing and Development Board (HDB) in Singapore give great insight into the workings of this housing provider. Ang and Teo (2000) discussed lessons learned by implementing a data warehouse at the HDB, and note that the biggest issues facing the organisation were highly fragmented data, no commitment to data quality by management, and individuals guarding data they owned. Prior to changing to a single data warehouse, the HDB had over 120 different databases, plus a large number of spreadsheets used to store data. This also seems to be the issue with UK housing providers (Leach, 2016), and a large project analysing big data is currently being conducted by the Housing Associations' Charitable Trust (HACT) on a number of different housing providers in the UK (Vine, 2016). The aim of the project is to drive effectiveness and efficiency of data analysis in social housing providers.

Studies on data quality have largely focused on the private market, and less so on the public market, i.e. social housing providers (Gruis et al., 2004). One of the reasons is the traditional task-focused approach employed by social housing providers due to bureaucratic government regulations and subsidies, as opposed to the market-oriented approach to improve 'fit' employed by most private organisations. However, Gruis et al. (2004) argued that the detachment of housing associations from government, the operation in deprived areas, and stock transfer from local authorities to housing associations led housing associations over time to develop business plans. It becomes more important to understand the market, especially for making future projections.

This sparse review on data issues in social housing highlights that there are problems with data quality in social housing providers, internationally as well as nationally, but it also highlights an even bigger concern: the lack of research within the social housing sector related to their data quality. The few studies reviewed each highlight a different approach to success through data, be it through data warehousing, business process reengineering, or process management. Different strategies exist to improve data, and in the following sections, those strategies most beneficial to the social housing sector will be introduced, followed by recommendations on steps social housing providers can take to improve the data they collect and keep.