

EXAMINING THE QUALITY AND MANAGEMENT OF NON-GEOMETRIC BUILDING INFORMATION MODELLING DATA AT PROJECT HAND-OVER

DARYN V. FITZ, NOHA SALEEB
Middlesex University, London. UK
daryn.fitz@excitech.co.uk
n.saleeb@mdx.ac.uk

Abstract. Through the exponential global increase of Building Information Modelling (BIM) adoption across the Construction industry, and the emergence of inter-connected, strategic and data-rich solutions; such as Big Data, the Internet of Things and Smart Cities, the importance associated with activities and decisions reliant on exact data input, transaction, analysis, and resulting actions becomes exponentially magnified. The supply of inaccurate BIM data may negatively impact on systems and processes that require fully assured data of appropriate quality/veracity, to support informed decision making, deliver functionality, facilitate services, or direct strategic actions within the built environment. This preliminary research intends to provide a catalyst for discussion, analysis and information retrieval relating to Building Information Modelling (BIM) processes where non-geometric data errors may; or are predicted to occur within a project environment. This may result in the delivery of data that cannot be described as representing truth or of good quality, and therefore of little value or use to the data user. The wider aspects of this research investigates specifically non-geometric data veracity & associated dimensions of data quality; in order to discover and explore future solutions to resolve current industry data quality assessment challenges. This paper provides feedback from the research focusing on the current state, presenting existing industry challenges and proposes further research areas based on initial findings.

KEYWORDS: Data, quality, management, building information modelling, BIM.

BACKGROUND AND RESEARCH RATIONALE

The construction sector is embarking on a journey of increased digitisation and the importance of providing accurate data and information through BIM associated processes will need to become more robust and go beyond current industry practice, to enable functions to not only evolve but also to play a greater role in the social and environmental impact of a digital age. This research identifies issues and challenges associated with the delivery of non-geometric BIM data at project hand-over as described in the next section.

Problem Statement

A primary element of Building Information Modelling (BIM) is information and data; this includes the generation, transaction, interpretation, and actions based on supplied data. There are many forms of instruction and legal contracts issued by Employers to their supply chain to procure data and information. However, the methods required for a Client or Employer to check that the data is exact or of good quality do not appear to be available at this time. For those suppliers of data who need to meet their contractual requirements, there is little guidance on a robust workflow that will ensure a quality assured approach to the validation and verification of required data intended for use beyond practical completion. This is considered by the researcher as an industry process and workflow gap which requires investigation, including the identification of current risks in the process.

Inaccurate data can mislead the user, affect decision-making processes and may result in incomplete data sets being disregarded or discredited due to isolated or multiple errors. When we consider the transaction of non-geometric data as part of the hand-over process to an Employer, for example data associated with maintainable building assets to support and inform planned and reactive maintenance tasks, it is possible that data sets provided are often inaccurate and incomplete which could result in a cultural environment where the recipient may assume the supplied data set cannot be relied upon and trustworthy. Facility management teams are often used to receiving incomplete, inaccurate and late delivery of information from construction teams or do not maintain their data sets, which results in the need for asset owners to re-survey facilities to gather the required data and information despite the activity incurring additional and potentially avoidable financial expense for the building asset owner.

At the 2015 ICE BIM Conference (ICE, 2015) it was reported that all centrally funded UK Government Departments would need to have “*the capability to electronically validate BIM information delivered from the supply chain*” and will also need to be “*making progressively more use of supply chain data for key business activities.*” BIM Task Group chairman Mark Bew stated “*As we move to BIM Level 3, and beyond, the reliability of data is key*” (ICE, 2015). These statements align to this research and indicate that the construction sector requires guidance and clarity on how to practically implement processes such as validation, quality assurance, and

EXAMINING QUALITY OF NON-GEOMETRIC DATA VERACITY AT PROJECT HAND-OVER FOR BIM DATA MANAGEMENT

achieve reliability of data to advance BIM maturity. Despite the presence of many data schemas, there remains continued ambiguity, and lack of clarity relating to what data is required, a standard industry data schema for product manufacture data is just one example (Ravenscroft, 2015).

This preliminary research investigated current state by focusing on data exactness or using the term adopted within the context of this research, Veracity, meaning trustworthy, veracious, and truthful (Oxford English Dictionary, 2016) and therefore can be considered correct, this is different to synonyms such as Accuracy where the technical definition is *'the degree to which the result of a measurement, calculation, or specification conforms to the correct value or a standard'* (Oxford Living Dictionaries, 2018), and as such results are variable and not always exact. The decision to adopt the term Veracity was also informed via previous research (Levine, et al., 1999) (Hernon, 1995) and is referenced within BS1192-4:2014 (BSI, 2014), where Veracity is defined as both an instruction and indication of quality *'The information provided should match the intended or actual facility.'* Wang proposed that Data and information are often used synonymously and is commonly differentiated intuitively, and described information as data that has been processed. For consistency, this research has used the term data wherever possible to refer to both data and information to avoid switching between terms (Wang, et al., 2002). This research focuses on non-geometric data only, supplied to a building asset owner at the end of a construction project. This is to ensure data, and the interpretation of that data is the focus of this study and not geometric models and model accuracy; which has already been adopted extensively as a research topic.

LITERATURE REVIEW

Data Quality Conceptual Complexities

Peter Hernon (1995) states *"it is not enough that information is readily available; before relying on any data or information, it may be important to ascertain, for example, the veracity of the content."* The Quality of Information (Cooke, 1999) is a term commonly used rather than accuracy by library and information scientists (Alexander & Tate, 1999) although verifying the accuracy of information is commonly used as one of the methods for evaluating the quality of information; additional methods for assessing quality include, but are not limited to, accessibility, relevance, comprehensibility, and navigability of information sources. This statement supports the view that the accuracy of information within a BIM workflow should be a consideration and part of the process, but additionally, other dimensions may need to be considered; this also poses the question of whether there is a logical sequence of steps or linear process to achieve a validated set of quality information or data.

It could be argued that inaccurate information may not be an issue as long as the user can identify the information that is inaccurate (Wachbroit, 2000) or as Vinton Cerf implied the user has the ability to apply critical thinking (2015). Barbules (2001) questioned how much evidence is required, and theorised that it might be dependent on how sure we need to be of the accuracy of the information. Which leads to the issue of ownership and responsibility of delivering required evidence, it could be proposed this can only be provided by those who acquire and distribute true beliefs (Smith, 2002) and not by those who utilise it. It should be noted that a user may be misled by incomplete information as well as by inaccurate information; therefore, we may need to consider the completeness of the information (Frické, 1997). Baird stated that not all information is verifiable and that “*some information is verifiable*” and can be readily checked once it is revealed (Baird, et al., 1998). Within the context of this research data & information could potentially be classified as being verifiable or non-verifiable data, as an example data associated with planned time focused activities such as cleaning regimes may be based on approximations, median data outputs or assumptions and therefore non-verifiable regarding veracity before the event or implementation. Fallis (2003) hypothesises that increasing the verifiability of information may be more cost-effective than teaching people how to evaluate information. Each person who needs to verify the accuracy of a piece of information has to expend energy to acquire and to apply these new skills. However, only one person (e.g., the author) has to expend energy to make the information more verifiable. This observation with regards current industry practice does lead to question who should be rewarded or compensated for ensuring information is verifiable. Kristo Ivanov (1972) concluded that Accuracy might be impacted by Motivation, Variability, and Frequency (familiarity), the concept of motivation and therefore a vested interest in information veracity on the part of the data provider may apply to future research.

Dimensions of Information Quality

A polygen model was developed in 1990 (Madnick & Wang, 2009) to answer data quality questions such as “Where was the data sourced from?” which in turn led to the development of a Quality Entity-Relationship model (Wang, et al., 1993) to enable the processing of hierarchical data quality metadata (Wang, et al., 1995).

EXAMINING QUALITY OF NON-GEOMETRIC DATA VERACITY AT PROJECT HAND-OVER FOR BIM DATA MANAGEMENT

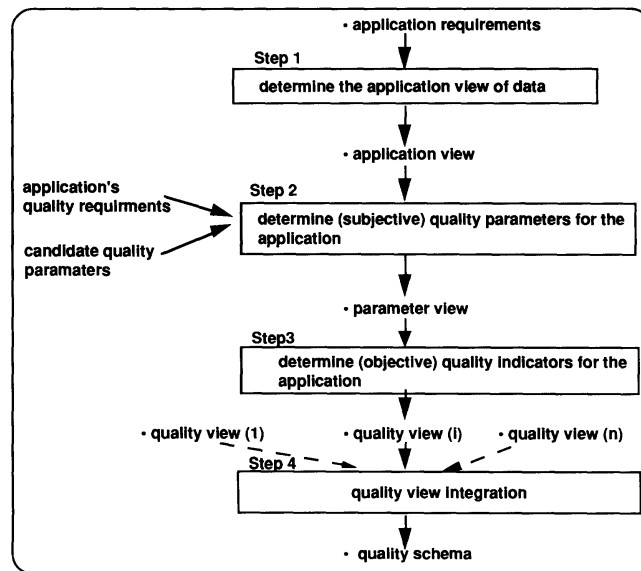


Figure 1. The process of data quality requirements analysis (Wang, et al., 1995).

In the latter of these papers it was stated that it is not always necessary to obtain zero defect data, as an example, postal services can make deliveries using a zip or post code even if the city name is incorrectly spelt. This statement is important with regards BIM, and structured data at hand-over as this asks the question “*Are we collecting too much or irrelevant data?*” The question of the actual cost to achieve zero defect data; if at all possible; was also highlighted within the research, indicating that cost to deliver may be a contributing factor to poor data quality. Relating this to the construction industry; which is cost driven and adversarial in nature; this incurred cost may also result in non-delivery. Madnick & Wang (2009) identified that organisations who collect, store and process data are faced with a number of challenges. This includes the integration of data from disparate sources, and a lack of a cohesive strategy to ensure appropriate stakeholders have the required information in a useable format, and at the right place and time. Deming (2000) proposed a Total Quality Management (TQM) framework, extended by Juran and Godfrey (1999) and developed further by Madnick and Wang through the MIT Total Data Quality Management (TDQM) program (MIT, 2002) which includes continuous data quality improvement by following the cycles of Define, Measure, Analyse and Improve (Madnick & Wang, 2009). During the TDQM program, data quality was defined from the consumer’s point of view in terms of fitness for use.

Further research developed an information quality assessment instrument (Lee, et al., 2002) for use in research as well in practice to measure data quality in

organisations. The instrument in operation separates each dimension, such as Accessibility, Timeliness, Completeness, Security as examples, into four to five measurable categories supported by functional forms to enable a method of scoring to be applied. Pipino's research also focused on assessment (Pipino, et al., 2002). Of note to this research specifically is that Pipino, like Lee, did not adopt the dimension accuracy as used by Strong (Strong, et al., 1997), but has replaced this with a term Free-of-Error further supporting a view that a consistent term for data considered as exact and truthful is required.

Lee (Lee, et al., 2002) conducted a gap analysis and identified differences between data dimensions and roles associated with data quality. In later research three major roles defined were data collectors, data custodians, and data consumers (Lee & Strong, 2004) this is relevant to this and future research as it provides roles that may be incorporated within a proposed framework. Within the Improvement stage it was hypothesised that it was more effective to change processes than actual data when quality issues were evident (Ballou, et al., 1998) (Wang, et al., 1998). In this regard, this leads this research towards new processes to avoid data change through increased transparency and ownership.

This literature review in totality to date has identified a lack of consensus relating to the actual Data Quality Dimensions to be applied, the definitions to be adopted, multiple associated complexities and a lack of specific guidance for an analytical workflow to indicate the quality of supplied or held legacy data, gaps this research intends to investigate.

RESEARCH RATIONALE AND DESCRIPTION

The objective of the research areas presented in this paper was to explore the following:

- What is the current state and potential quality of non-geometric data provided to Clients / Employers at the end of a construction project?
- What gaps may exist in current processes relating to veracity?
- What are the key issues that may need to be considered when developing a solution to the problem?

This research focused on non-geometric data veracity only, to ensure data, and the interpretation of that data was the focus of the study and not geometric models and model accuracy; which has already been adopted extensively as a research topic. Example types of non-geometric data that form the focus of this research are listed in TABLE 1 below.

EXAMINING QUALITY OF NON-GEOMETRIC DATA VERACITY AT
PROJECT HAND-OVER FOR BIM DATA MANAGEMENT

TABLE 1: Non-Geometric Data Types Examples

Data Type	Description
Space (Location)	Named location for activities such as use, inspection or maintenance including unoccupied or uninhabitable spaces.
Floor (Region)	Named spatial subdivision, including horizontal and vertical levels and areas. Example: a vertical lift shaft.
Zone	Names set of spaces (locations) sharing a specific attribute such as an activity, access, management or conditioning. Example: heating or security controlled zone.
Component Name	Named and individually scheduled physical item and features that may require management such as inspection, maintenance or replacement during operational use.
Component Type	Names specification for Components including equipment, products and materials.
System	Named set of manageable Components providing a common function. Example: an HVAC System.
Job	Named task or activity during operational phase.
Resource	Names material, equipment or skill required to exercise Jobs.
Spares	Names replaceable part associated to types.
Classification	Classification system references.
Documentation	Names external document associated to an asset.
Contact	Names person or organisation associated to the facility lifecycle.

The above listed data types are examples and depending on Asset Owner type can be extensively expanded to suit the organisations Estate Management Plans or aligned to specific data requirements of Computer Aided Facility Management (CAFM), or Computerised Maintenance Management Systems (CMMS).

During the literary investigation, the term accuracy / accurate appeared to be defined and adopted but without consistency. The US Institute of Building Documentation (USIBD) provides guidance on Level of Accuracy (LOA) (US Institute of Building Documentation, 2014), the framework adopted defines different levels of accuracy in terms of standard deviation, focused exclusively on line-work or geometric model documentation, expanding on the existing European DIN 18710

standard (DIN, 2012), but does not consider non-geometric data. However, the USIBD does state that level of accuracy specifications is something that industry has long struggled with, and if left undefined this can create problems in as-built deliverables. ISO 8000 (The International Organization for Standardization, 2009) defines data accuracy as the closeness of agreement between property value and the true value and therefore is adopting a measurement approach but provides no detail how this should be achieved. This research also needed a term that divorced itself from the word accuracy due to its common association within the construction sector as a measured tolerance, and not necessarily an undisputable truth, such as digitally scanned data being described as having an accuracy of 2mm. The term Veracity has been adopted for this research, and resulting research introduces the concept of Level of Veracity (LoV).

METHODS

From the literature review, by considering the characteristics of Data Quality Dimensions there is a strong probability that non-geometric BIM data at all project stages may be prone to error and inaccuracies, and therefore it is proposed that at project hand-over the data and information as a complete data set is unlikely to reflect the real-world environment or the building element it intends to represent. If this situation is found to be the case, then the effort and negative financial impact associated with collecting this type of data and information of low quality or veracity cannot be considered sustainable or practical, and new approaches must be researched and developed. In order to investigate current perceptions, understanding, working practices and challenges that may, or may not exist across the construction industry relating to data and information veracity, several tools were employed to collect required data. This section outlines the data collection process; the tools utilised, participants to the process and summarises the rationale for adopting approaches, such as population selection, sample size, question types, and anticipated value from the data collected.

The following data collection tools (TABLE 2) were employed to collect the required data.

EXAMINING QUALITY OF NON-GEOMETRIC DATA VERACITY AT PROJECT HAND-OVER FOR BIM DATA MANAGEMENT

TABLE 2: Data Collection Tools

Data Collection Tool	Participants	Total Number
Survey	Industry Population - To provide an industry-wide perspective.	111
Interview	BIM Experts / Practitioners - To provide expert opinion from extensive industry experience.	14
Interview	Software Developers - To provide an opinion from a technology and data management perspective	7

To capture data from the widest and most varied sample population; the use of an online survey questionnaire was selected as the tool to capture primary data from the AEC industry population. Of the two hundred and two participants who started the survey one hundred and eleven completing the entire set of twenty-six questions. Only fully completed surveys have been used to support this research. Following the collection of demographics data the objective was to collect data relating to the primary research questions and to investigate the following ancillary topics to aid this and future research:

- What the population considered current state; with regards the delivery of non-geometric BIM information and in particular the veracity of that information.
- If appropriate guidance existed to deliver and confirm quality assured data.
- If there was a direct relationship between Data and Information Quality versus Time & Cost.
- Identifying where errors or omissions were likely to occur using current workflows.
- Identifying primary focus areas of improvement for the supply of quality data.

Two sets of interviews were facilitated using Quantitative data collection, to expose new knowledge associated with this under-researched area of study and Qualitative data collection, allowing the opportunity to create a theory or new hypotheses (Jensen & Laurie, 2016). Interview questions progressed with a quantitative question, which was then followed by an open question to introduce a qualitative data collection process to provide the greatest opportunity for interviewee's to provide additional feedback and insight. Using a mixed methods approach within this study allowed the greatest opportunity to determine current state and begin to develop new theories for further research by using both Quantitative and Qualitative methods (Johnson & Onwuegbuzie, 2004).

Seven software developers were identified to support this research via interview. This list included established software solutions and also technologies that had not been released to market during the research period. The selection was made through the identification of technologies that had embedded functionality to support non-geometric data management. The intention of this section of research was to; investigate current views and general software processes with regards to data and information veracity. The data collection was not structured to provide a comparison functionality list between products, or claim that the selected products are the best solutions available to industry. It is anticipated that similar research in this area would be likely to use a different selection of technologies as there is an extensive sample to choose from, and new technologies are constantly being released to market or existing ones further developed.

The industry population selection process for the interviews adopted homogenous sampling (Miles & Huberman, 1994) to ensure the interview population consisted of industry practitioners interested in the area of research. Individuals were selected from a wide range of organisation types and professional backgrounds to provide an opportunity to collect different viewpoints and perspectives. This approach added value to the research, by providing an opportunity to explore an industry-wide viewpoint from a diverse range of experience, knowledge, and opinion provided by interviewees and also avoided convenience sampling.

ANALYSIS AND DISCUSSION OF RESULTS

From survey data, 74% of respondents confirmed the view that non-geometric information should be 100% correct at project hand-over. Additional feedback via survey comments implied a common opinion that 100% correct was not a realistic possibility which aligns with Baird's observation that not all data is verifiable and that "some information is verifiable" and can be readily checked once it is revealed (Baird, et al., 1998). Strong views were shared that Clients should state their requirements clearly within contracts and Employers Information Requirement's (EIR's) to support delivery. During the interviews, BIM Experts / Practitioners agreed that the information should be correct if requested by the Client or Employer and indicated that a contractual requirement might be needed to better ensure delivery of requested data at an agreed quality or veracity. This raises further questions as for how contract compliance will be evaluated, and it is therefore proposed that a method for measurement is required to support the contractual framework.

During the interviews, there was a common consensus that across the industry today, supplied data associated with real-world physical components that should reflect reality was unlikely due to complexities such as:

EXAMINING QUALITY OF NON-GEOMETRIC DATA VERACITY AT PROJECT HAND-OVER FOR BIM DATA MANAGEMENT

- Process gaps through lack of checking, verification, and validation
- Technology limitations
- The ability of those reviewing the correctness of the information - not having the required knowledge or experience

The survey indicated that 72% do not believe that Employers know how to check non-geometric information for correctness and that 64% do not believe the Employer will check the data provided. This feedback does not indicate whether an Employer should be responsible for the task of checking, but it does indicate a perceived lack of process, supporting tools or knowledge on the part of an Employer to undertake this task. However Fallis's (2003) hypothesis that increasing the verifiability of information might be more cost-effective than teaching people how to evaluate information may be valid, but this view could also be expanded to include methods for verifiability and measurements relating to the intrinsic quality of the supplied information.

Interviews with Software Developers indicated that there was a consensus that Employers would find the checking of non-geometric information challenging when compared to their supply-chain, the reasons stated included:

- The Employer was the most remote from the information process
- The Employer does not have the skills or tools required
- There was currently little interest to review received data or information
- Responsibility lies with the supplier of that information; therefore, there is an element of faith.

This feedback would begin to support Smith's proposition that the ownership and responsibility of delivering required evidence can only be provided by those who acquire and distribute true beliefs (Smith, 2002) and therefore in this context of this research not the Employer but instead the employed supply-chain. This introduces new risks due to a potential situation where an entire supply-chain is assessing or marking its own work and outputs in the knowledge that the Employer will not conduct his/her own compliance or quality audits.

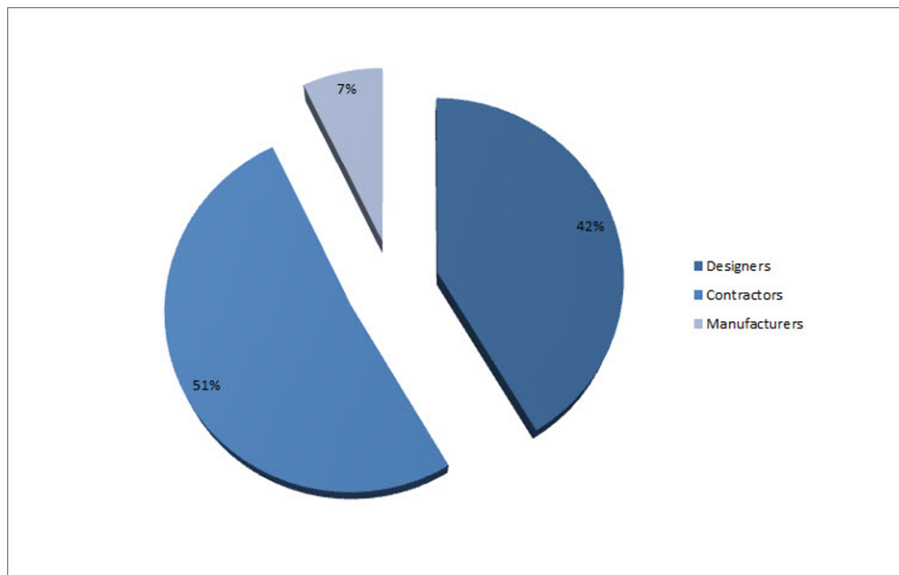
Interviews with BIM Experts/Practitioners provided some additional insights which included common opinion that:

- There was an element of apathy within Employer organisations to check received information
- It is not the Employer's role to check contractual deliverables
- Required skill sets to check would not typically exist within Employer organisations.

The opinion that a specialist team is required to check received data was a popular response and should be employed by the Employer directly or via the project management team. This role could in the future be undertaken by an Employer-side Information Manager, but this view, it is proposed in this paper, could be challenged as the Information Manager, although motivated, may not have the requisite knowledge and competencies to be successful as highlighted by Smith (2002) and without processes to evaluate Data Quality the role may revert to presence and format checking only. It is also proposed in this paper that for the UK Government BIM Level 2 mandate and associated initiatives to be successful appropriate motivation will need to be developed within Client and Employer organisations. Ivanov (1972) stated that Accuracy might be impacted by Motivation, Variability, and Frequency (familiarity) and to support improved Data Quality outputs each of these concepts has been considered within the main body of this research.

Of those surveyed, 92% did not believe traditional hand-over documentation such as Operation & Maintenance manuals and drawings were 100% accurate when supplied to the Client. This data highlights the need for this research by exposing a significant quality issue and waste of resource due to poor outcomes. BIM Experts/Practitioners further supported this hypothesis and agreed without exception that information received by Clients/Employers was never 100% accurate, and many reflected on personal experiences and observations which included comments such as; *'I have never seen a project release an accurate as-built drawing'*, *'O&M's are never complete'*, *'Nothing close to accurate just full of generic information'*.

During interviews, there were conflicting views that industry either needs to employ new roles such as 'data wranglers' or that industry needs solutions that are 'under the hood,' i.e., Simplify data collection to such an extent that individuals do not need to understand data structures such as COBie. The issue of who inputs the data and how many times it is transacted before final input was also highlighted.



EXAMINING QUALITY OF NON-GEOMETRIC DATA VERACITY AT PROJECT HAND-OVER FOR BIM DATA MANAGEMENT

Figure 2. Questionnaire - View on who has the greatest challenge producing non-geometric BIM information

The survey (Figure 2.) indicated that respondents' consider Manufacturers' having the least challenges (7%) when producing non-geometric BIM information. There was no clear consensus on whether it is more challenging for a Contractor (51%) or a Designer (42%). Interview responses from interviews supported the hypothesis that it is less challenging for Manufacturers to produce exact non-geometric BIM information when compared with Designers and Contractors.

Interviews produced the following findings (TABLE 3) when comparing Designer and Contractor challenges in delivering non-geometric data & information.

TABLE 3: Designers / Contractors challenges in delivering non-geometric data & information

Designers	Contractors
Do not work with exact information	Are contracted to deliver
Are now being asked for new and different information via BIM processes (unique ID's and classification of spaces as examples)	Have supply chain challenges, and need to collate data and information from many third parties
The design is always changing and fluid in nature	Have to model a final product
Designers do not know what is going to be procured.	Need to adapt to site conditions and change
	Skill gaps exist on-site recording site data
	The contractor is not the original author
	Contractor mistakes or omissions due to a lack of design or product understanding
	May select wrong manufacturers specification information
	Contractors have to do a lot of fact-finding
IFC is not delivering what was intended	
There is little industry agreement whether data should be in a model, database or spreadsheet	
Industry requires a seamless data transfer solution	

The comments received indicated that a method to confirm and report to the Employer that the hand-over of data is of the required quality/veracity and complete may be required, without the need for direct intervention by the Employer. This supported a related common view that the Employer should not need to directly check the data. On reflection, many of the comments relating to contractor challenges are similar to the Employers, and therefore a method for establishing data and information quality may also be required by the contractor or third party provider responsible for the collation of data and information on their behalf.

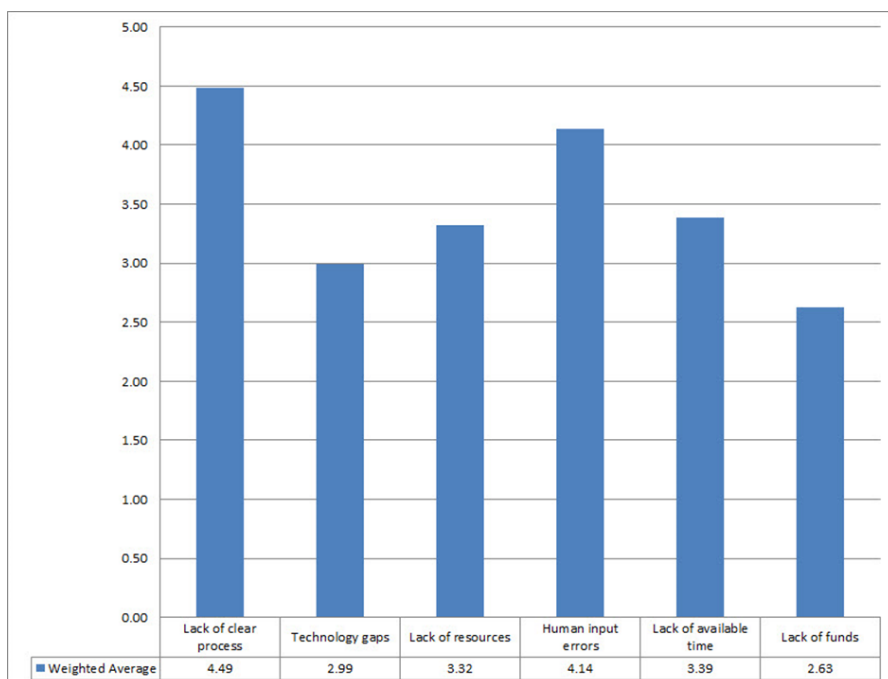


Figure 3. Questionnaire - Identification of risk to non-geometric information veracity

Survey data (*Figure 3*) ranked lack of clear process as the major contributor to risk of non-geometric information veracity, followed by human input error, technology gaps ranked only 5th, which may infer that risk to information veracity may be best mitigated through the process rather than enhancing current technologies initially. When asked what would help to deliver exact non-geometric information, survey respondents replied 23% Industry standards, 23% Technology to assist, 21% Company workflows, 18% Time for the task, 15% Specific training. Once again there was no indication that technology advancement alone would provide a viable solution and two Software Developers during interviews stated that workflows were a

EXAMINING QUALITY OF NON-GEOMETRIC DATA VERACITY AT PROJECT HAND-OVER FOR BIM DATA MANAGEMENT

prerequisite for the development of the technology. BIM Experts/Practitioners when interviewed typically stated appropriate standards would be the most helpful.

During interview Software, Developers responded that their individual systems did consider non-geometric data exactness, but the veracity of that information could not be guaranteed using the technology, because the technical content and therefore confirmation of the veracity of that data must be provided by the author.

Technology processes for checking data were identified and included:

- Checking the presence of data against employer's requirements.
- Checking data is in the expected format such as date or currency.
- Checking specified data against manufacturers' data via comparison rules.
- Checking and interrogating data using algorithms and rule sets.
- Identification of expected parameters associated with elements, such as a fire door object requiring a fire rating parameter.

The above processes would provide benefits to the structure of the data, its format and provide presence checking, but none of these processes assist in confirming the intrinsic quality of that data, whether it is trustworthy or reflects the real-world as examples. This is where the concept of Data Quality Dimensions proposed by Wang (Wang, et al., 1993) and others begin to add another layer of evaluation towards veracity, by exploring qualities such a provenance, accuracy, timeliness as examples.

When asked what has been the greatest challenge conceptually or during software development with regards data veracity; the responses received from all interviewed were very similar; the challenges identified included:

- Not receiving data in the first place
- Not being able to trust the data
- Not receiving data that was structured
- Project teams are not specifying requirements
- Lack of understanding within design teams and a need to change working practices
- Classification systems are required and need to be issued to software developers
- Misalignment between construction specialists and a data management specialists (individual knowledge and experience need to be aligned)

The above feedback indicates many challenges for the Construction sector in a global economy where digital transformation is considered the future for a sustainable business. The fact that not receiving data, or not being able to trust the received data was specifically cited when asked about conceptual and software development challenges validates the need for this research, and this is discussed in the next section.

DISCUSSION

This research presented evidence that multiple challenges currently exist in the procurement and delivery of non-geometric data in both quality and the veracity of that data. This has been achieved by collecting and analysing industry data via a survey and BIM experts, practitioners and software developer's points of view and where possible identifying consensus of opinion and experiences. This research has identified a number of key findings which is summarised below:

The first research question was. What is the current state and potential quality of non-geometric data provided to Clients/Employers at the end of a construction project? This research has indicated a long-standing acceptance that data provided to the Client/Employer is of a poor quality, accepted on faith and that the intrinsic quality of that data is unknown beyond basic presence checking. When considering what gaps may exist in current processes relating to veracity? This research has identified a lack of defined standards, methods and procedures for data quality assessment relating to veracity. This directly results in a lack of ability of the Employer to instruct its supply chain to comply with or implement directly, a data quality assessment process focusing on intrinsic quality. This lack of process has also been identified as the primary cause for poor data quality and lack of veracity at hand-over. The third research objective posed the question. What are some of the issues that may need to be considered when developing a solution to the problem? These are summarised below.

Contracts and Standards

Current contractual requirements and legal frameworks may need to be amended to include clauses relating to the Data Quality or Veracity required by the Employer and state the method of measurement(s) for compliance. In addition, current UK BIM standards and specifications do not include specific guidance relating to expected data veracity or quality, and only make reference to validation and verification processes. Where validation is a process to check if data is sensible and reasonable, but it does not check the quality of that data, and where verification is often a manual process to check data entered matches the original source but is totally reliant on the competencies of the person undertaking the task. This final point is considered significant; Lee & Strong (2004) defined three major roles associated with data, data collectors, data custodians, and data consumers. It is proposed in this paper that currently the verification or validation processes are being ignored or partially implemented by data consumers, data custodians have little trust in the data provided and it is typically not complete and as proposed by Smith (2002) - evidence can only be provided by those who acquire and distribute true beliefs. This leads to a hypothesis that the data collectors, such as the product manufacturers or constructors are best positioned to positively influence the quality of the data provided and contribute to Data Quality assessments once the requirements have been clearly specified. There is an opportunity to expand and build upon existing BIM standards by developing supporting guidance in areas such as validation & verification and as suggested by

EXAMINING QUALITY OF NON-GEOMETRIC DATA VERACITY AT PROJECT HAND-OVER FOR BIM DATA MANAGEMENT

this research towards a state of veracity, but these standards also need clear ownership and responsibilities to be defined and made a contractual obligation.

Competency and Process Gaps

There is a consensus that data associated with a real-world physical component should reflect reality. However, the realities of this being achieved currently across the industry were considered unlikely due to multiple complexities which included process gaps through lack of checking, verification and validation, current technology limitations and those individuals who are tasked to review non-geometric data do not currently have the required competencies.

There are strong indications to suggest that suppliers do not believe the Employer will, or has the capability to check supplied data and that there is an element of apathy within Employer organisations to check received information.

The following contributing reasons were identified:

- The Employer was the most remote from the process, therefore, the information
- They do not have the skills or tools required
- There was currently little interest to review received data.
- Responsibility lies with the supplier of that information; therefore, there is an element of faith pertaining to a guaranteed of veracity.

A specialist resource representing the Employer's interests may be required to support the Employer, and this could be incorporated into a revised Employer's Information Management role, however, rather than receiving data third-hand it may duties include direct data collection or auditing of that data at source, i.e. assessing the outputs from the original author or data collector.

Industry perception, as identified within this research, is that hand-over data & information is of poor quality, incomplete and rarely reflects as-constructed. This indicates a long-standing culture on the part of the Employer and its supply-chain that this situation is inevitable and has to be accepted. Via Building Information Modelling (BIM) and industry digitisation this view and cultural acceptance is being challenged but not at a Data Quality level, and if cultural change is to be achieved the supplied information and data must be suitable for purpose, and adopting just one Data Quality Dimension trustworthy. Currently, best practice is limited to presence and format checking and alignment to pre-described classification systems which provides little indication or confidence in the trustworthiness, accuracy or believability of that data set.

Technology and System Challenges

For technology providers, the changing industry standards relating to Building Information Modelling (BIM) adds extra complexity in developing future toolsets, and there is also a lack of clarity on how dispersed or centralised construction related data sets and repositories will be configured in the future to suit project data and information requirements. Existing tools are limited in their ability to confirm veracity and instead the only person who can confirm the veracity of data is the original author. It is proposed that processes relating to data delivery that might be considered one hundred percent truthful must ensure the original author provides the required provenance, knowledge and confirmation; because current technology cannot measure or confirm data veracity as human interpretation is required to put context to the data and data-sets.

It is unlikely that all data can reach a state that can be termed as exact; as an example data associated with radiator wattages will not always reflect the physical installation as that will be dependent on factors such as heat source and settings on items such as thermostatic radiator valves. When considering the use of terms veracity and accuracy associated with Building Information Modelling (BIM), veracity could be utilised to describe a measured range for non-geometric data towards 100% truth, where accuracy has already been adopted (US Institute of Building Documentation, 2014), to define different amounts of measured geometric deviation from the truth and the actual physical built world.

The Employer is not in a position to approve or confirm the veracity of data received, but can accept some dimensions of data quality such as completeness, and determine if all data requested has been received. During interviews with software developers, a number are incorporating this type of functionality within their applications. This will assist the project management and efficiency of the data collection process, but it will not confirm that the data supplied is appropriate for use or if it is of poor, medium or good quality depending on the individual Data Consumers requirements.

CONCLUSION

During this initial study, it was evident that there needs to be extensive education across the industry to explain what Data Quality is, and there is a requirement for a concise list and categorisation of Data Quality Dimensions specifically for Construction sector adoption to be defined to support further research in this area. Guidance on how to deliver and assess Data Quality to ensure it meets the Data Consumers' requirements, including future contractual arrangements is also required, with a caveat that not all data within data records will be verifiable and therefore beyond dispute.

EXAMINING QUALITY OF NON-GEOMETRIC DATA VERACITY AT PROJECT HAND-OVER FOR BIM DATA MANAGEMENT

The research from data collection indicates that secondary data entry by third parties, data presence checking via visual inspection or scripts, and checking processes to confirm that data requested by the Employer has been provided appears to reflect best current working practice at this time. All indications from the data collected suggest that all three of these processes have scope to become more efficient and improvements can be made to provide a more robust system but a new supporting framework is required with the concept of Level of Veracity (LoV) proposed.

This preliminary research has confirmed that this area of research is needed and that Data Quality and assurances related to that data will become of increasing importance as the Construction sector continues its modernisation through digital transformation, continued adoption of Building Information Modelling and associated processes including the management of legacy data.

REFERENCES

- ALEXANDER, J. E. & TATE, M. A., 1999. *Wisdom: How To Evaluate and Create Information Quality on the Web*. 1st ed. New Jersey: Laurence Erlbaum Associates.
- BAIRD, D. G., GERTNER, R. H. & PICKER, R. C., 1998. *Game theory and the law*. 2nd ed. Cambridge, Mass.: Harvard University Press.
- BALLOU, D., WANG, R., PAZER, H. & KUMAR, G. T., 1998. Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science*, 44(4).
- BSI, 2014. BS 1192-4:2014 Collaborative production of information Part 4: Fulfilling employers information exchange requirements using COBie - Code of practice, London: BSI.
- BURBULES, N. C., 2001. *Paradoxes of the Web: The Ethical Dimensions of Credibility*, Illinois: Library Trends.
- CERF, V. G., 2015. Internet Society. [Online] Available at: <http://www.isoc.org/internet/conduct/truth.shtml> [Accessed 19th October 2015].
- COOKE, A., 1999. *Neal-Schuman authoritative guide to evaluating information on the Internet*. Illustrated ed. New York: Neal-Schuman Publishers Inc.
- DEMING, W. E., 2000. *Out of the Crisis*. 2nd ed. Cambridge, Mass: MIT Press.
- DIN, 2012. DIN 18710 PART 1, GENERAL REQUIREMENTS, Berlin: Beuth.
- FALLIS, D., 2003. *On Verifying the Accuracy of Information: Philosophical Perspectives*, Arizona: University of Arizona.
- FRICKÉ, M., 1997. Information using likeness measures. *Journal of the American Society*, 48(10), pp. 882-892.
- HERNON, P., 1995. Disinformation and misinformation through the internet: Findings of an exploratory study. *Government Information Quarterly*, 12(2), p. 133.
- ICE, 2015. ICE Conferences. [Online] Available at: <http://www.ice-conferences.com/ice-bim-2015/> [Accessed 4th November 2015].
- ISO, 2011. ISO/TS 8000-1:2011 - Data quality -- Part 1: Overview, Geneva: ISO.
- IVANOV, K., 1972. *Quality-Control of Information on the concept of accuracy of information in data-banks and in management information systems*, Virginia, USA: National Technical Information Service.

- JENSEN, E. & LAURIE, C., 2016. *Doing Real Research - A Practical Guide to Social Research*. Reprint ed. London: Sage Publications Ltd.
- JOHNSON, R. B. & ONWUEGBUZIE, A. J., 2004. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33(7), pp. 14-26.
- JURAN, J. M. & GODFREY, A. B., 1999. *Juran's Quality Handbook*. 5th ed. New York: McGraw-Hill.
- LEE, Y. W. & STRONG, D. M., 2004. Knowing-Why About Data Processes and Data Quality. *Journal of Information Management Systems*, 20(3), pp. 13-39.
- LEE, Y. W., STRONG, D. M., KHAN, B. K. & WANG, R. Y., 2002. AIMQ: a methodology for information quality assessment. *Information & Management*, Volume 40, pp. 133 - 146.
- LEVINE, T. R., PARK, H. S. & MCCORNACK, S. A., 1999. *Accuracy in detecting truths and lies: Documenting the "veracity effect"*, Oxford: Communication Monographs.
- MADNICK, S. E. & WANG, R. Y., 2009. Overview and Framework for Data and Information Quality Research. *ACM Journal of Data and Information Quality*, 1(1).
- MILES, M. B. & HUBERMAN, M. A., 1994. *Qualitative data analysis: an expanded sourcebook 2nd Edition*. Thousand Oaks: Sage.
- MIT, 2002. MIT TDQM PROGRAM. [Online] Available at: <http://web.mit.edu/tdqm/www/index.shtml> [Accessed 26th October 2015].
- OXFORD ENGLISH DICTIONARY, 2016. OED Online. [Online] Available at: www.oed.com [Accessed 30th April 2016].
- OXFORD LIVING DICTIONARIES, 2018. Oxford Living Dictionaries. [Online] Available at: <https://en.oxforddictionaries.com> [Accessed 22th September 2018].
- PIPINO, L. L., LEE, Y. W. & WANG, R. Y., 2002. Data Quality Assessment. *Communications of the ACM*, 45(4).
- RAVENSROFT, T., 2015. [bimplus.co.uk](http://www.bimplus.co.uk). [Online] Available at: <http://www.bimplus.co.uk/news/new-bim-maindate-data2ils-include-data-validation/> [Accessed 4th November 2015].
- SMITH, C., 2002. Contextualism and the division of labour. *Social Epistemology: A Journal of Knowledge, Culture and Policy*, 16 Issue 1, 2002 (1), pp. 65-81.
- STRONG, D. M., YANG, L. W. & WANG, R. Y., 1997. Data Quality in Context. *Communication of the ACM*, 40(5), pp. 103,110.
- US INSTITUTE OF BUILDING DOCUMENTATION, 2014. *Guide for USIBD Document C220TM: Level of Accuracy (LOA) Specification 2014 ver. 1.0*, s.l.: USIBD.
- WACHBROIT, R., 2000. RELIANCE AND RELIABILITY: THE PROBLEM OF INFORMATION ON THE INTERNET. [Online] Available at: <http://journals.gmu.edu/PPPQ/article/view/345/273> [Accessed 20th October 2015].
- WANG, R. Y. & MADNICK, S. E., 1989. *The Inter-database instance identification problem in integrating autonomous systems*. Cambridge, MA, Massachusetts Institute of Technology.
- WANG, R. Y., REDDY, M. P. & KON, H. B., 1995. Toward quality data: an attribute based approach. *Decision Support Systems*, Issue 13, pp. 249 - 372.
- WANG, R. Y., ZIAD, M. & LEE, Y. W., 2002. *Data Quality*. 1st ed. New York: Kluwer Academia Publishers.
- WANG, R. Y., KON, H. B. & MADNICK, E. S., 1993. *Data Quality Requirements Analysis and Modeling*. Vienna, Ninth International Conference on Data Engineering.