



UNDERSTANDING THE PERSISTENCE OF UNSTRUCTURED DATA IN AEC THROUGH THE LENS OF IRISH PRACTITIONERS' PERSPECTIVES: A QUALITATIVE INVESTIGATION

Hamza Tariq, Ciaran McNally, and James O' Donnell
University College Dublin

Abstract

The AEC industry generates vast unstructured data, primarily in the form of natural language communications, complicating machine-interpretable information retrieval. While the focus of existing research has been on transforming unstructured data into structured forms, few studies address the root causes of unstructured data generation and the complexity surrounding it. Through semi-structured interviews and thematic analysis, this study identifies systematic, not technical, issues as primary contributors. It highlights confusion in understanding structured data, structured data-sharing processes, and BIM's purpose and definition. The study emphasizes the need for education on data science foundations and a universally accepted understanding of BIM.

Introduction

The construction industry generates more data than ever before, marking the industry's entry into the era of Big Data (Li *et al.*, 2023). This is due to widespread usage of different models and tools (Soman and Whyte, 2020), as well as information technologies such as BIM, GIS, internet of things, and wireless sensor networks (Huang *et al.*, 2021). These technologies can bring value to the sector through supporting automated decision making, cost saving, enhanced productivity and collaboration, safety management, resource optimization, real time evaluation and more (Cakmak and Akturk, 2024). Yet despite this potential, the AEC industry remains reluctant to take advantage of these developments.

Snyder *et al.*, (2018) estimate that about 95% of data generated in the AEC industry remain untapped. This is because much of the information generated is unstructured, meaning such data cannot be interpreted by machines and is stored in various formats such as PDFs (natural language text documents), images and videos (Soibelman *et al.*, 2008), with textual data comprising over 80% of the total (Wu *et al.*, 2022). To address this challenge and support the extraction of value from unstructured data, researchers are employing various techniques to convert them into structured formats. These include natural language processing and text mining

(Shamshiri *et al.*, 2024), knowledge graphs, linked data and ontologies (Bazuin *et al.*, 2023; Gudnason and Pauwels, 2016; Hong *et al.*, 2022), machine learning (Jung *et al.*, 2021) and semantic concept networks (Sobhkhiz and El-Diraby, 2023).

The ongoing creation of unstructured data stems from a combination of factors, including organizational and cultural divisions. Soman and Whyte, (2020) noted that these divisions, alongside technical reasons, drive fragmented practices that limit the effective use of digital tools. Such practices often rely on hybrid communication channels, like emails and undocumented meetings, causing data silos, repetition, and inconsistencies. Although these codification challenges are well-documented, **there is limited research exploring how institutional and organizational factors influence unstructured data generation. Furthermore, there is limited research on how unstructured data manifests across project stages and how their purpose evolves over time.** This is the gap that this research aims to address.

The aim of this study is to investigate the organizational and institutional factors influencing unstructured data generation in construction projects. By analyzing industry professionals' perspectives, the research explores the underlying causes of unstructured data and their implications for collaboration and machine interpretability. A key aspect of this study involves identifying how unstructured data themes evolve across different project stages, revealing patterns that emerge during project delivery. A starting assumption of this study is that professionals understand what constitutes unstructured data, and that excessive documentation introduces challenges to collaboration and machine interpretability. Furthermore, identifying the root causes of unstructured data generation can help mitigate the prevalence of such data before applying information retrieval techniques. Building on these assumptions, the research intends to answer the following question: **How do organizational and institutional factors influence the generation of unstructured data in AEC industry?** Semi-structured interviews are chosen for this study as they enable the adoption of a conversational format with open-ended questions for in-depth discussion (Schwartz-Shea

and Yanow, 2015); this allows some leeway to follow-up on whatever is deemed important by the researcher (Brinkmann, 2020). Additionally, this study draws conclusions from insights provided by professionals representing diverse backgrounds.

The remainder of this paper consists of four further sections. The next section presents the methodology employed in the research and is followed by the results, featuring interview analysis as thematic discussions. The limitations of the study are then discussed, after which the conclusions are presented alongside potential avenues for future research.

Methodology

The paper adopted the four stepped approach to qualitative descriptive study (Figure 1) as recommended by (Villamin *et al.*, 2024). The first step involved participant sampling, followed by data collection. The data collected was then analyzed and finally findings were subsequently drawn. These steps are detailed in the following section.

Sampling

The qualitative sampling process was guided by (Robinson, 2014) four point framework. The first step was about defining the sample universe. For this study, construction companies within the Ireland were targeted. To focus on the study and meet the objective, inclusion criteria were applied, limiting participants to those with

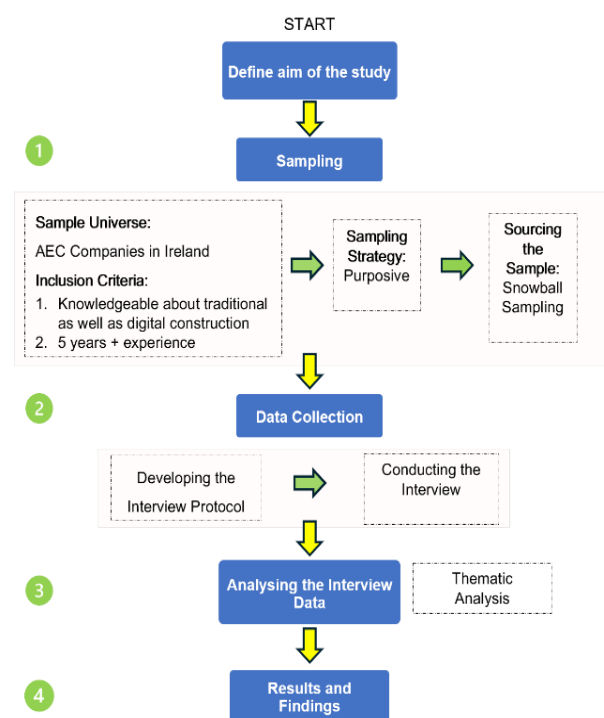


Figure 1: The four methodological steps of this study: sampling participants, collecting data, analyzing interviews and presenting findings

5+ years of experience and knowledge both in traditional construction methods as well as digital technologies within the construction industry. The second step was choosing an appropriate sample size. The study adopted flexible sample size where data would be collected until saturation was achieved. The next step was choosing an appropriate sampling strategy. Since the focus of this study was gaining a holistic understanding of how the construction sector operates, a purposive sampling strategy was adopted to recruit a mixture of company sizes, drawn from architects, designers and contractors. The final step was to source the sample. Since the study focused on interviewing professionals with specific requirements, the snowball sampling (also called referral sampling) was adopted to reach them.

Data Collection

Data was collected in the following two phases:

Phase 1: Developing the Interview Protocol

The interview protocol was developed based on the framework provided by (Kallio *et al.*, 2016), which includes identifying the prerequisites for using semi-structured interviews, conducting a literature review (as discussed in the introduction), formulating a preliminary interview guide, seeking expert assessment, and finalizing the complete guide. This process resulted in the formulation of primary questions centered on two key themes: unstructured data sources and organizational factors, ensuring the study's focus remained clear. Since the interviews were semi-structured, the questions were designed to align with the research objective while allowing flexibility for follow-up questions based on the interviewees' responses. An information sheet was designed to give participants an overview of the research study. The distinction between unstructured data and structured data was explained in the information sheet by providing the following definitions (Soibelman *et al.*, 2008)

- Structured Data: Data that follow a particular format or order, are machine interpretable and, exist in formats such as BIM files and spreadsheets.
- Unstructured Data: Data that do not follow a particular format and are not machine interpretable, such as textual content in emails, documents, audio, and video files.

Phase 2: Conducting the Interview

Participants were recruited through referrals based on the inclusion criteria. Initially, 13 individuals were contacted, of whom 5 agreed to participate. To ensure data saturation (Guest *et al.*, 2006) a sixth interview was conducted after analysing the first 5 interviews. Their profiles are summarized in Table 1. In the third interview, two professionals from the same company participated, so they were treated as single participants. All interviews were conducted via Zoom, with informed consent obtained from each interviewee for audio recording. Each interview lasted approximately one hour, and the

recordings were transcribed and securely stored in a cloud-based repository.

Table 1: Interviewee Profile: Role, Scale, and Associated Practice

#	Interviewee Code	Practice	Scale	Role
1	P1	Architectural	Micro	Director
2	P2	Multi-Disciplinary	Large	Associate Director - Mechanical Lead
3	P3	Architectural	SME	Data Manager and Senior Architect
4	P4	Architectural	SME	Director
5	P5	Contractor	SME	BIM Coordinator
6	P6	Contractor	Large	BIM Manager

Analyzing the Interview Data

For this study, thematic analysis from (Braun and Clarke, 2006) had been adopted. This method is well established in the literature and was used to identify, analyze and report patterns within the data. Transcripts were extracted from the recordings, and thematic analysis was conducted using NVivo to identify patterns and represent quotes. Data analysis followed a six-step process (Braun and Clarke, 2022). The initial codes were developed after understanding the data. These codes were then critically analyzed to generate themes. The next step involved examining and refining the themes. Finally, the themes were incorporated to create a coherent and compelling narrative of the data.

Results and Findings

This section describes the themes that emerged from the analysis of the interviews. The themes were grouped and can be considered to fall into two broad categories: (i) organizational issues and (ii) unstructured data across project stages.

Category 1: Organizational Issues

Within this category we have a number of themes identified, and these are discussed below.

Theme 1: Confusion in the understanding of unstructured data, structured data, and structured data-sharing processes

The study began with the assumption that participants had a clear understanding of what unstructured data actually are. The responses have however shown that this is not

necessarily the case. Our initial assumption proved to be incorrect, as there was notable confusion regarding the concepts of unstructured data, structured data, and the sharing of data through structured methods.

This confusion persisted despite the definitions being explained both in the information sheet and at the start of the interview. For instance, three participants explicitly expressed their lack of understanding about what constitutes unstructured data.

“I’m not sure what you mean by unstructured data though, what would you mean by that?” (P1)

“What would you consider unstructured data source?” (P3)

“I’m just not entirely sure how I’m going to differentiate between the structured and unstructured results in this one, because I’m getting all of them.” (P3)

“How much time we are spending on searching the unstructured data. ... I’m not entirely sure what part of the ... day-to-day search is relating to the unstructured data, we are just trying to minimize that time.” (P3)

“So, when you say the unstructured, that’s the likes of our Pdf documents, our minutes. things that aren’t residing in models?” (P2)

Interestingly, P5 initially provided a clear explanation of what unstructured data are, but later suggested that information shared through a common data environment (CDE), when accompanied by appropriate file naming, is transformed into structured data.

“We consider unstructured data information that is not predefined in a data model or organizational framework. such as PDFs, audio files, video images and social media posts.”

“Within the ISO 19650 we address those problems through naming suitability and provision folder structure and the lifecycle of the assets. and through an internal folder structure. And so, ... this is the structure that we follow to name every single document, regardless of what it is, regardless of whether it’s structured data or unstructured. Through this way an unstructured data becomes structured through the life cycle of the project.”

A similar confusion was also expressed by P3.

“They usually impose some sort of a common data environment which is the highest structured kind of information that we can possibly share.”

When P6 was interviewed at a later stage, this question was posed to him directly, and in his response, he conveyed a similar sense of confusion.

“So do you suppose adding appropriate file naming conventions make the data structured.” (Researcher)

“Yeah, yeah, exactly.” (P6)

In summary, confusion persists regarding the basic understanding of structured data, un-structured data and structured ways of sharing data.

Theme 2: Lack of Understanding about BIM

One of the key organizational challenges identified is the confusion and inconsistency in how BIM is understood and applied across different stakeholders. Participants often distinguished project documents and templates from BIM, recognizing only 3D models as part of the BIM process. This indicates a narrow view of BIM, which

limits BIM's potential integration into the broader project delivery framework. As P1 noted:

"As I mentioned in some of the correspondence, we don't actually use Revit, BIM as such in the traditional sense of BIM"

Furthermore, BIM is often misunderstood or underutilized, with many professionals not fully grasping the value beyond tools such as Revit. For example, P3 explained that BIM deliverables are outsourced to technicians who are not necessarily part of the design team, and the model is treated as an afterthought rather than a central tool for coordination:

"They kind of outsource their BIM deliverables to like a technician or somebody who's not necessarily the designer. And so, I think that's how they would overcome that step if they were, you know, heavily involved in the project. But the one thing that they need to deliver at the end is the BIM model. It kind of comes almost as an afterthought, like something that they produce rather than like a design tool. And then that's less useful for us because we don't have it as part of our kind of coordination sequence."

This reflects a deeper misunderstanding of BIM's role and potential for collaboration, which results in fragmented workflows and unstructured data.

Additionally, reliance on traditional document formats such as PDFs, instead of BIM-based workflows, persists in many organizations. This created difficulties for integrating data and making the data machine interpretable. As P4 pointed out:

"We're talking about BIM. So within their philosophy, or like nowadays in many companies, BIM is Revit." highlighting a narrow interpretation of BIM technology that limits broader application. This confusion extends to stakeholders who may not fully understand BIM or the intended outcomes, as expressed by P4.

"If I might say so, sometimes even people involved in the process, they're not fully aware of what it is. You know the process or what the outcome should be and we're dealing with contractor and design team that are learning"

Although P6 had an excellent understanding of BIM, he explained that many of his colleagues in the company still equate BIM solely with models. He remarked:

"It's a culture change within a very large company. It's very hard to change that full mindset. So, there's lots of people within the company that still only consider it as the models. But there are also a lot of key players, realizing more and more that it's about the data."

This lack of clarity and understanding within organizations contributes to the persistence of unstructured data, as traditional, unstructured communication channels continue to be used in place of more structured, BIM-integrated approaches.

Theme 3: Skills Gap

Another organizational factor contributing to the generation of unstructured data is the skills gap in delivering structured information, particularly in relation to BIM. Participants highlighted that effective use of BIM tools requires a high level of proficiency, and without a

skilled team, the potential benefits of BIM such as reducing construction costs or minimizing errors are not fully realized. As P1 stated:

"I don't see why using a Revit model is going to result in a reduced construction cost and arguably it might reduce the opportunity for errors, and therefore cost, arguably, but that would rely on a whole team being extremely proficient on that platform."

Furthermore, access to software and the ability to use such software effectively were also identified as constraints. This was expressed by P2:

"Probably the main constraint in terms of sharing BIM files is access to the software and user ability to use that software."

He also highlighted:

"It's a knowledge gap and an expertise gap as opposed to any technical reason."

The lack of expertise extends beyond just software usage, shaping the way models are created and shared. P2 pointed out that the way models are drawn by architects can impact their ability to be read by other software.

"It may not be a software issue; it might be down to how the model is physically drawn by the architect. And so there might be a very specific way to draw the model, so that the thermal simulation software can read it. And if the architect is not drawing it that way, then the model won't read it."

Additionally, even contractors working with BIM models often lack knowledge of key processes, such as managing asset registers or working with COBie data standards. As P4 shared:

"Even the contractors that we've been working with, they had to go in training. They didn't have any knowledge of asset register or COBie."

Lastly, P6 shared an extreme example of the skills gap, describing how a small civil company still produces their drawings manually using a drawing board. He remarked,

"When I was trying to onboard this person in there, I could tell he wasn't understanding what I was talking about. So, I tried to look at it from a different angle, and I said: Look, how do you produce your drawings, your fabrication drawings, so that your team know how long to cut a piece of pipe and what angle it should be used. And he said, well, typically, I'll take the Pdf that's given to me by the designers and then I'll sit down on the drawing board with a T-square and a set square and a pencil, and I'll draw out my drawing. And I couldn't believe it, that company was still using pen and paper to get their drawings together."

These skills gaps, combined with inconsistent application of BIM, lead to the creation of unstructured data, as the required expertise to deliver structured, interoperable information is often lacking across the project team.

Category 2: Unstructured Data across Project Stages

The responses from participants reveal a wide variety of unstructured data generated during construction projects, with distinct purposes at each stage. These stages are divided into design development, tender, project execution and statutory approvals.

Theme 1: Design Development

During design development, unstructured data primarily help convey ideas, align stakeholders, and secure approvals. Participants described the extensive use of visuals, reports, and sketches. P3 mentioned:

“A lot of presentation materials at the early stages that might be quite well developed, like visuals and things that are going to sell the project and based on images.”

P2 explained:

“At stage one, we would usually just be providing reports, maybe with some sketches to support those. They would be in PDF format.”

Manual methods also persist during this phase, as P3 shared:

“We still have people who are drawing manually, so sketches or watercolors will be generated at that stage as well, usually scanned and digitalized.”

P4 highlighted the abundance of scanned sketches:

“You know, there are lots of sketches there, scanned in and put into the server.”

Theme 2: Tender Stage

In the tender stage, unstructured data largely consist of specifications, reports, and deliverables needed for contractor engagement and approvals and they are in pdf format.

“a suite of drawings, mostly for the purposes of pricing and tendering the built projects, of obtaining costs for the built project and those drawings are supplemented with written specifications for the most part.” (P1)

“If we're talking about tender documentation, then there'd be a huge amount that would just be in specification documents.” (P2)

“But those schedules are still going to be produced in a PDF Format that is going to get issued in a tender package to a client.” (P3)

During this stage, proposals are also submitted outlining the information management procedures to be followed. As P6 stated:

“A lot of our projects, they won't specify a CDE. So, in our tender submission, we would propose our CDE, and this is done by the tendering department. I don't have great visibility of it, but they would. There's a whole series of documents outlining the naming convention, how we use it, the security, the where the information is stored online, what servers is used etc. So, we would propose all of that at tender.”

Theme 3: Statuary Approvals

Significant amounts of unstructured data are generated throughout various project stages, driven by regulatory requirements and compliance processes. In the planning phase, this includes planning documentation and written reports, as P1 explained:

“We would also then later produce planning documentation, which are drawings submitted obviously for securing planning permission for the project, for statutory approvals, fire safety certificates, disabled access certificates. So, therefore, achieving statutory approvals. And alongside those we would prepare written reports for achieving statutory approvals.”

During the design and construction stages, professionals must prepare ancillary certificates for compliance. As highlighted by P2:

“There's an ancillary certificate of design commencement, ancillary certificate of design completion, and then an ancillary certificate of design inspection. So, there's 3 certs that we have to do. But each individual engineering discipline will have to produce those, and then the contractor will have to produce. There're separate certs, the contractor has to submit. They all are packaged together and given to the assigned certifier, who then produces his overall cert, which is given to the local authority to say that you know the certified of the building has followed the building regulations.”

P2 further emphasize how these certificates are compiled into commissioning documents at the handover stage:

“For example, at the end of the life, safety systems, emergency lighting and fire detection and alarm systems, we would have to provide a design certificate at the end of the project to certify the design of the system, which then gets packaged up with the commissioning cert as part of the Handover certification document.”

The BCAR (Building Control Amendments Regulations) process further necessitates the creation of extensive documentation throughout various project stages. As P4 described:

“they're a standard PDF template that's produced...All those certs get uploaded to what's known as the building certificate management systems.”

All these pdfs are a requirement to fulfil regulations requirement.

“BCAR process. You know there is a large number of documents that need to be produced throughout the construction stage, for whatever you know, stage applies of a project, and then at completion.” (P4)

This highlights the fact that the generation of unstructured data is inherently linked to the current statutory compliance process, which requires PDF based statutory approvals.

Theme 4: Project Execution Phase

In this phase, unstructured data are generated to monitor, report, and track progress, often driven by the need for documentation and proper record-keeping. This includes meeting minutes, client meeting minutes, and site visit records.

“Meeting minutes, generally, the site meeting minutes, client meeting minutes, site visit records. So, when visiting site, observing any issues to do with quality control or progress on site for the purposes of monitoring the work on site.” (P1)

“The date for the pouring of the concrete is recorded through a photo, from my booklet on site.” (P3)

P2 mentioned preparing formal deliverables such as reports, certifications, and snag lists.

“Our formal deliverables during construction and during handover would be reports, certification, and snag list things like that, usually our... updates to our design deliverables. But no new deliverables are

produced at that stage, and it's really just reporting and snagging and close out."

Additionally, issue trackers and spreadsheets are used for day-to-day collaboration.

"We would have issue trackers, both for internal use, for ourselves, or they could be shared with other consultants. We would also consider spreadsheets generally a text document as well, because these are not calculation spreadsheets as such. They contain a lot of material like written material." (P3)

Documents produced as a part of ISO 19650 are also a part of this.

"Our main ones really are BEP, TIDP, MIDP" (P6)

These documents, shared in PDF format, form a critical part of the execution phase documentation, helping to ensure quality control, safety, and smooth project progress throughout the execution phase.

"Usually, we issue all of them in as a PDF format, and that's the way also the department like to receive them with the correct file naming." (P4)

In summary, unstructured data such as meeting minutes, site visit records, photos, reports, and issue trackers play a crucial role in documenting and tracking progress during the execution phase.

Discussion

As described at the outset, most research has focused on technical solutions for extracting value from unstructured data while neglecting systematic issues related to unstructured data generation. This study addresses this knowledge gap concerning organizational and institutional factors behind unstructured data generation. The findings reveal that a lack of clarity in distinguishing structured from unstructured data, combined with a fragmented understanding of BIM, contributes to persistent unstructured data practices. Consistent with prior research (Soman and Whyte, 2020; Wu *et al.*, 2022), professionals often rely on traditional documentation formats such as PDFs, limiting the potential for machine-interpretable information. Skills gaps in structured data processes and BIM workflows further compound these challenges, leading to inefficient data management across project stages.

This study has several limitations that should be acknowledged. First, as a qualitative study based on interviews, the findings may not be fully generalizable to the broader Irish AEC sector. A larger quantitative study could provide greater generalizability and validate these results. Second, the study did not include interviews with professionals representing clients. Including client perspectives in future research could provide valuable insights into their understanding of unstructured data and its challenges. Lastly, while the study identified key organizational and institutional factors contributing to unstructured data generation, addressing how these challenges can be mitigated falls outside the scope of this research.

Conclusions

This qualitative study employed semi-structured interviews and thematic analysis to address the knowledge gap concerning the influence of organizational and institutional factors on the generation of unstructured data. Three major themes were identified and discussed. The research also examined how unstructured data manifests across project stages and evolves in purpose over time, revealing four distinct themes. The analysis indicated that the organizational challenges contributing to unstructured data generation are primarily linked to two factors:

- 1) Education and training, leading to poor understanding of structured data, unstructured data, and structured data-sharing processes, as well as misconceptions about BIM.
- 2) A lack of expertise in delivering structured information.

These findings suggest that systematic issues, rather than technical limitations, are the primary drivers of unstructured data proliferation. For instance, visuals, sketches, and reports are commonly used to convey ideas during design development. In Ireland, tendering and statutory approval processes heavily rely on PDFs, while unstructured data such as meeting minutes and progress reports play a critical role in record-keeping during project execution.

A key question then is how can these systematic challenges be addressed to enable greater adoption of machine-interpretable data in construction? Evidence from this study highlights the need for targeted education to build a foundational understanding of data science and promote a universally accepted interpretation of BIM across the sector. Additionally, policy and practice must evolve to enable machine-interpretable statutory approvals in Ireland. Future research can also focus on developing structured data tendering methods to replace current unstructured approaches, fostering a more data-driven and efficient construction industry.

Acknowledgements

The research team would like to thank all the subject matter experts for participating in this research and sharing their insights. The authors would like to acknowledge the support from the Build Digital Project, funded by the Department of Public Expenditure, National Development Plan Delivery and Reform, Ireland.

References

- Bazuin, B., Stolk, S., Stevens, M. (2023) Towards Recommendations for Knowledge-Graph-Based Requirements Management in Construction: A Report on the EU DigiChecks Project. In International Conference on Advanced Information Systems Engineering. Springer, Cham, pp. 80–91.
- Braun, V., Clarke, V. (2022) *Thematic analysis: a practical guide*. London ; Thousand Oaks, California: SAGE.

- Braun, V., Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*. 3(2), 77–101.
- Brinkmann, S. (2020) Unstructured and Semistructured Interviewing. In P. Leavy, ed. *The Oxford Handbook of Qualitative Research*. Oxford University Press, pp. 277–299.
- Cakmak, P.I., Akturk, B. (2024) An in-depth investigation of digital construction technologies from a building economics perspective. *Construction Economics and Building*. 24(3), 40–75.
- Gudnason, G., Pauwels, P. (2016) SemCat: Publishing and accessing building product information as linked data. In *eWork and eBusiness in Architecture, Engineering and Construction: ECPPM 2016*. CRC Press, pp. 659–666.
- Guest, G., Bunce, A., Johnson, L. (2006) How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*. 18(1), 59–82.
- Hong, Y., Xie, H., Bhumbra, G., Brilakis, I. (2022) Improving the accuracy of schedule information communication between humans and data. *Advanced Engineering Informatics*. 53, 101645.
- Huang, Y., Shi, Q., Zuo, J., Pena-Mora, F., Chen, J. (2021) Research Status and Challenges of Data-Driven Construction Project Management in the Big Data Context. *Advances in Civil Engineering*. 2021, 1–19.
- Jung, Y., Fouad, A., Golparver-Fard, M. (2021) A systematic review on the requirements on BIM maturity and formal representation of sequencing knowledge for automated construction scheduling. In *Proceedings of the 38th International Conference of CIB, Luxembourg*. pp. 11–15.
- Kallio, H., Pietilä, A.-M., Johnson, M., Kangasniemi, M. (2016) Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*. 72(12), 2954–2965.
- Li, F., Laili, Y., Chen, X., Lou, Y., Wang, C., Yang, H., Gao, X., Han, H. (2023) Towards big data driven construction industry. *Journal of Industrial Information Integration*. 35, 100483.
- Robinson, O.C. (2014) Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide. *Qualitative Research in Psychology*. 11(1), 25–41.
- Schwartz-Shea, P., Yanow, D. eds. (2015) *Interpretation and method: empirical research methods and the interpretive turn*. Second edition. London New York: Routledge.
- Shamshiri, A., Ryu, K.R., Park, J.Y. (2024) Text mining and natural language processing in construction. *Automation in Construction*. 158, 105200.
- Snyder, J., Menard, A., Spare, N. (2018) Big data= big questions for the engineering and construction industry. *White Paper. First Myanmar Investment (FMI). Raleigh, US*.
- Sobhkhiz, S., El-Diraby, T. (2023) Dynamic integration of unstructured data with BIM using a no-model approach based on machine learning and concept networks. *Automation in Construction*. 150, 104859.
- Soibelman, L., Wu, J., Caldas, C., Brilakis, I., Lin, K.-Y. (2008) Management and analysis of unstructured construction data types. *Advanced Engineering Informatics*. 22(1), 15–27.
- Soman, R., Whyte, J. (2020) Codification Challenges for Data Science in Construction. *Journal of Construction Engineering and Management*. 146(7), 04020072.
- Villamin, P., Lopez, V., Thapa, D.K., Cleary, M. (2024) A Worked Example of Qualitative Descriptive Design: A Step-by-Step Guide for Novice and Early Career Researchers. *Journal of Advanced Nursing*.
- Wu, C., Li, X., Guo, Y., Wang, J., Ren, Z., Wang, M., Yang, Z. (2022) Natural language processing for smart construction: Current status and future directions. *Automation in Construction*. 134, 104059.