Interfacing and the making of data infrastructures within building construction and operation

Anushri Gupta
London School of Economics
a.gupta140@lse.ac.uk

Roser Pujadas
London School of Economics
r.pujadas140@lse.ac.uk

Will Venters
London School of Economics
w.venters@lse.ac.uk

Abstract

Data analytics is important for supporting innovation and creating value. Literature has highlighted the promising role of advanced technologies like AI, ML and big data analytics in driving operational efficiency and the optimisation of services which contribute to sustainability goals. Despite the recognition of interconnectivity, and data linking as important dimensions of data analytics, there remains a gap in our understanding of how this "interconnecting" is undertaken in practice. Our research addresses this gap by examining the practices involved in interfacing the various elements of data infrastructure that span differing physical, organisational and technological boundaries in order to achieve sustainability goals. We study this by considering the case of a large property development and management company. Using the theoretical lens of digital coordination theory, our initial findings highlight various resistances in creating data infrastructures and interfacing of systems within the property development and management organisation.

Keywords: Interfaces, Data Infrastructure, Practice perspective, Digital co-ordination theory, Sociomateriality

1.0 Introduction: Data analytics for innovation and sustainability

Data analytics is important for driving innovation and achieving the triple bottom line of sustainability - driving economic, environmental and social value (Davenport & Horton, 2006; Gholami et al., 2016; McAfee & Brynjolfsson, 2012). Literature has highlighted the promising role of advanced technologies like AI, ML and big data analytics in driving operational efficiency and the optimisation of services which contribute to sustainability goals. For example, the management and cooperation of heterogeneous sensors for public spaces monitoring; centralized operational centre of smart buildings for energy efficiency and security; efficient use of electricity within smart infrastructure and data centres; real-time applications for disaster management in urban spaces based on information collected from various entities (e.g. crowd sourcing, homes, vehicles); air quality monitoring (Fan et al., 2020; Ismagilova et al., 2019; Lex et al., 2019; Luo, 2022; Sembroiz et al., 2019).

Yet despite all its promises, recent literature in Information Systems (IS) suggests the need to better understand how organizations realise value from data analytics, and the practices involved (Monteiro & Parmiggiani, 2019; Østerlie & Monteiro, 2020). Such a practice perspective considers the necessary work to gather and prepare data so that it can be analysed (Parmiggiani et al., 2022; Tempini, 2017), including integrating or synthesizing data from different sources (Günther et al., 2017).

Literature has consolidated how meeting the above ambitions needs data to be drawn from a diverse range of sources and systems - databases, actuators, sensors, mobile phones, smart cards, smart meters, open government data and social media, to name a few, that generate vast amounts of structured and unstructured data (George et al., 2014; Gupta et al., 2020; OECD, 2016). In this context, a number of authors discuss techniques like data linking and centralisation to bring these diverse sources of data together to create value (Barns, 2018; Bright et al., 2019; Raetzsch et al., 2019).

Despite the recognition of interconnectivity (Günther et al., 2017), and data linking (Bright et al., 2019) as important dimensions of data analytics, there remains a gap in our understanding of how this "interconnecting" is undertaken in practice. Our research addresses this gap by examining the practices involved in interfacing the various elements of data infrastructure that span differing physical, organisational and technological boundaries in order to achieve sustainability goals.

We do this by empirically studying interfaces and the processes of interfacing. Our guiding research question is: *What is the role of interfacing in value creation within data infrastructures?* Adopting a practice lens, and drawing upon digital coordination theory and data infrastructure, we explore the issues associated with interconnecting and interfacing various systems of data generation and analysis within an organisation.

2.0 The need to study Data infrastructure and interfaces

Data infrastructures are conceptualised as socio-technical systems (Brous et al., 2016) that go beyond recognising the role of the 'hard infrastructure' (physical servers, and networks) (Blazquez & Domenech, 2018); to also emphasise the integral role of 'soft infrastructure' (value networks and governance aspects) (Jetzek, 2016; Suzuki & Finkelstein, 2019), in delivering data enabled products and services. A data infrastructure embodies elements of technology (hardware/software), data (e.g. reference, thematic, and metadata), governance (policies, processes and regulations), and people (decision makers, entities, agents and other data consumers) that follow shared rules (British Standards Institution, 2017; Brous et al., 2019; Dodds & Wells, 2019; Oliveira & Lóscio, 2018) in creating value from data. These shared rules comprise of data assets, like identifiers, technologies that help manage and use them, policies that govern how they are used, and the organisations that curate and maintain them. As such the notion of a data infrastructure pivots around creating value from data by taking into consideration the complex interplay between technological, institutional, organisational and economic changes.

The concept of a data infrastructure goes beyond the siloed operation of data use and analysis within the boundary of a department/organisation, and supports cross-boundary data sharing (Iryna Susha et al., 2017). Integral to facilitating this cross-boundary flow of data between departments/organisations, is the role of interfaces. In this context, literature highlights the role of technical interfaces like APIs, platforms, open (meta) data standards, identifiers, and open-source technologies to name a few (Barns, 2018; Jeong et al., 2020; Raetzsch et al., 2019).

Identifiers, for instance, are an integral interface that enable integrating data within organisations, between business partners and across sectors and industries in the city (for e.g. datasets like energy performance certificate and property prices if combined can provide useful insights)[1]. Similarly, conceptualising well defined interfaces like APIs and metadata standards has the potential of streamlining the building permit application process (as shown in the 'BRISE'[2] project in Vienna).

Such initiatives underline the need to further our understanding on interfaces in the constitution of data infrastructures. However, the conceptualisation of interfaces as just a technology is insufficient to uncover the necessary interfacing processes, which

[1] https://theodi.org/article/the-public-sector-geospatial-agreement-theres-still-more-to-do-to-unlock-the-value-of-the-uks-geospatial-data/

[2] https://www.uia-initiative.eu/en/uia-cities/vienna-call4

involve interactions between data, technology, people. Effective functioning of a data infrastructure is reliant on the interfacing between the various elements that define it. In essence, we need an understanding of not only technology that enables data to 'flow' between the various digital systems, but also the practices and processes that enable these elements to work together (Kitchin & Moore-Cherry, 2020; Meijer, 2018). To this end, we intend to extend our understanding of data infrastructures through the incorporation of practice theory and in particular Digital Coordination theory (Venters et al 2014) (itself drawing upon Pickering's (1995) Mangle of Practice).

Digital Coordination theory argues that current human practice seeks to accommodate the various resistances faced (e.g. technology, standards or differing human intentions) within a cycle of tuning, improvising, reacting, that is informed by the past (e.g. conventions, learnt practices, installed bases of software) and oriented to the future (e.g. intentions, expected achievement, emerging plans). This theoretical addition allows us to examine how a data infrastructure's hard infrastructure materially resists, how actors building the data infrastructure seek to respond to such resistance (based on an intention of achieving sustainability goals, and drawing upon existing construction practices and standards) and how this tunes (or Mangles) the resultant sociomaterial arrangement that is the data infrastructure-in-use. In this way our study will shed light on the bringing into existence and use, systems of technology, data, people and processes to drive environmental sustainability goals.

3.0 Research Design

Our ongoing empirical research involves the study of data infrastructure in the making. We draw on a qualitative study of a large property development and management company. We are studying their efforts to implement data analytics with various aims, including achieving the net zero sustainability targets. Our research studies the challenges and efforts associated with interfacing systems together to gather, curate, and analyse data.

This is research in progress and our initial findings presented in this paper draw on 10 interviews conducted over the summer 2022. Interview participants included key roles such as analytics director (LS5, LS14), data analysts (LS4), a data director (LS3), a procurement manager (LS9, LS6), and data governance consultant (LS13) within the organisation. Interviews lasted for an hour and were fully transcribed for analysis. Over the coming months, we plan to conduct further interviews, gather relevant documentation, and conduct observations (including visiting the buildings and understanding the systems used, and attending relevant meetings). Our initial analysis is interpretive undertaken through reading and discussing transcripts (Walsham, 2006). Drawing on Digital Coordination theory, we established codes to interpret our empirical data: processes (hybrid, co-working of teams, re-defining meta-data standards), technology (Information Technology, Operational Technology), data (consumer, assets).

4.0 Initial Findings

Our initial analysis highlights various resistances in creating data infrastructures and interfacing of systems within the property development and management organisation. These resistances manifested themselves along technical, organisational, and managerial dimensions, which we empirically substantiate below.

- Interfacing of ‘technical’ IT (Information Technology) and OT (Operational Technology) systems demands the tuning of ‘organisational’ interfacing by various teams co-working within the organisation.

In interfacing the various digital and technical systems within the organisation, interviewees distinguished between IT (more traditional Information Technology systems like Risk Management applications or Business Intelligence) and OT (Operational Technology systems using embedded controllers and specialist computers such as CCTV, lighting and Building Management Systems (BMS)). Such Operational Technology is often quite specialist and emerges from an installed base of engineering rather than business technology. This causes resistances as while vendors selling IT solutions “*understand data analytics because they've been working with log files, click rates and marketing analytics, they've never worked with OT... so, where is the sensor, its unique identifier, the haystack protocol or things like that.* (LS5)*”* On the other hand, vendors selling OT solutions have future intentions that create resistances – they *“want to sell closed platforms... they want to give you analytics, they want to create AI, but then they lock you in, you don't have the data.*(LS14)*”* As a result, there lies a challenge in interfacing these systems that don't *"speak to each other"*. The interviewee noted that their response is through tuning the teams internally by getting the "*IT, data and analytics team, building operations and building engineering teams to work together ... that are historically separated within organisations.* (LS5)" They made it evident that interfacing of these digital systems is not an *"IT lead conversation"* but involves co-working of teams to map the data and digital systems within the organisation to enable its interfacing.

- Interfacing data storage systems demands a hybrid approach to the technical architecture that raises critical managerial questions

In achieving their intention of net zero carbon within a building's development, our interviewees highlighted 'data' as a critical asset, that was largely broken down into corporate data, data on clients/customers occupying those buildings, and real time IoT data from assets deployed within the building. Currently in driving analytics through these datasets, the organisation however uses diverse business analytics platforms across the operations of the building that presents a material resistance to the data handling. As the interviewee notes *"current business analytics platforms are unable to handle the data streams of IoT data that amount to Pb (peta bytes). These platforms are good enough to handle corporate data that usually amounts to Gb (giga bytes).* (LS4)*"* The IT team are responding to this resistance by interfacing current corporate business analytics platforms with an IoT gateway and IoT hub tools based in the cloud with the intention of building a data lake architecture which they intend will allow easier access and use of data.

Effective functioning of the above data lake architecture in turn requires the adoption of a hybrid approach to the technical architecture (on premise IT + cloud) to respond to the various technical resistances caused by the OT technology and networking issues. As an interviewee explains *"I need the physical servers, physical IoT gateway in the building, not in the cloud because there might be challenges with the network. Say we lose the internet connectivity, ... anything locally still needs to work, the building still needs to run, the air conditioning, the lighting still needs to work. Hence, we need an on-premise infrastructure....need some local storage.*(LS9)*"*

However, as one of our interviewee notes, in working towards the above hybrid approach lies several important managerial challenges that need to be considered. For instance, *"the frequency with which data needs to be updated; determining IT infrastructure cost based on size of data etc.* (LS3)*"* As such, our findings illustrate that interfacing of systems faces not only technical resistance but also managerial resistances.

- Interfacing various asset management systems within a building demands the (re)definition of metadata standards

Building operations are supported through the various assets deployed within the building (plumbing, lighting etc). Interviewees noted that these assets are supplied by different vendors and typically work in silos with differing intentions for their products. Each of these assets have associated metadata that is defined by the vendor providing that system and is named differently: "there *is no standard out in the industry."* This poses a significant material resistance to the optimization of the centralised operations of these systems through data analytics and achieving net zero carbon emission intentions; *"unless of course, you do very complex mapping"*. This is being accommodated by interfacing the various asset systems to create a centralised asset management system that demands internally defined metadata standards for the assets involved.

Interestingly, the historical installed base of building technology and the history of technology use and practices mean legacy building development require standards to be redefined. As an interviewee notes, *"If you're developing a new building, then you have the most control, because you're using the latest systems, and they're still defining these IDs, those vendors.... But if it's an existing operational building, then we don't have any contractual control over the new things or the vendor will ask us for money. And that adds up depending on how many vendors there are. So, in the operational environment, what we do is we deploy a database server, which does the tagging for us. And there are sort of formulas applied against it, so that certain types of devices get tagged a certain name*. (LS14)" Here a new technology (the database server) is used to accommodate the installed base of existing systems standards and allow their redefinition and ongoing use.

5.0 Discussion

This is an ongoing research project that aims to expand the growing area of research on data infrastructures to achieve environmental sustainability goals. In this early paper, we have gone to the backroom (or the substrate) of the "backrooms of data science" (Parmiggiani et al., 2022) to provide evidence of the work and practices of materialisation needed to realize the value of data analytics in organisations. In doing so, and drawing upon Digital Coordination theory, we reveal the relevance of interfaces in the development of data infrastructures, and uncover the interfacing processes that enable the development and functioning of data infrastructures. The initial findings from our case study of a large property development and management company, illustrated the complexity of bringing together various teams involved in the development and management of a building to define an asset management system of a building; vendor lock in issues with the suppliers of those assets; the need to define metadata standards; and maintaining a hybrid approach to developing the IT system of building operations - which altogether reflect the sociomateriality (Venters et al., 2014) of technology and data constituting data infrastructures to meet the net zero sustainability ambitions.

Our future research direction will consider studying how the materiality and materialisation of data plays out in creating value in such contexts, and how synthetic knowing (Monteiro & Parmiggiani, 2019) in this context becomes consequential for the achievement of net zero goals. Our early findings illustrate that data isn't just available, but is material in nature, created by sensors, servers and platforms (and the latency therein), and people and processes. As empirical findings show us, common business analytics solutions present challenges in new contexts such as buildings where their focus on lightweight data (e.g. social media or marketing data) limits their ability to handle constant streams of sensor data. In our case the accommodation to this challenge was through the creation of a data lake platform that could host the organisation's corporate data as well as its real time sensor data generated from IoT assets.

Another possible research stream is understanding the evolution of interfaces and the data infrastructure, as a building transitions from construction to completion and maintenance. This can help evaluate the relationships between evolutionary change across multiple interfaces, contexts of use, and organisational goals and examine how resistances and accommodations emerge in this transition into use. Finally, we seek to develop a theory of data infrastructure evolution that builds upon the theoretical framework developed here.

Acknowledgements

This research was supported by the UK's Engineering and Physical Sciences Research Council -EPSRC (Grant EP/R006865/1).

References

Barns, S. (2018). Smart cities and urban data platforms: Designing interfaces for smart governance. *City, Culture and Society*, *12*, 5–12.

Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*, 99–113.

Bright, J., Ganesh, B., Seidelin, C., & Vogl, T. M. (2019). Data Science for Local Government. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3370217

British Standards Institution. (2017). *Smart cities – Guide to establishing a decision-making framework for sharing data and information services* (Issue October 2016, pp. 1–48).

Brous, P., Herder, P., & Janssen, M. (2016). Governing Asset Management Data Infrastructures. *Procedia - Procedia Computer Science*, *95*(95), 303–310. https://doi.org/10.1016/j.procs.2016.09.339

Brous, P., Janssen, M., & Herder, P. (2019). Next Generation Data Infrastructures: Towards an Extendable Model of the Asset Management Data Infrastructure as Complex Adaptive System. *Complexity*, *2019*(2), 1–17.

Davenport, E., & Horton, K. (2006). The Production of Service in the Digital City: A Social Informatics Inquiry. *IFIP International Federation for Information Processing, Social Informatics: An Information Society for All?: IN REMEMBRANCE OF ROB KING*, *223*, 233–242.

Dodds, L., & Wells, P. (2019). Data infrastructure. In *State of open data*.

Fan, C., Yan, D., Xiao, F., Li, A., An, J., & Kang, X. (2020). Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches. *Building Simulation 2020 14:1*, *14*(1), 3–24. https://doi.org/10.1007/S12273-020-0723-1

George, G., Haas, M., & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal*, *57*(2), 321–326.

Gholami, R., Watson, R. T., Molla, A., Hasan, H., & Bjørn-Andersen, N. (2016). Information Systems Solutions for Environmental Sustainability: How Can

We Do More? *Journal of the Association for Information Systems*, *17*(8), 2. https://doi.org/10.17705/1jais.00435
Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191–209.
Gupta, A., Panagiotopoulos, P., & Bowen, F. (2020). An orchestration approach to smart city data ecosystems. *Technological Forecasting and Social Change*, *153*, 119929. https://doi.org/10.1016/j.techfore.2020.119929
Iryna Susha, P., Verhulst, S., Janssen, M., Theresa Pardo, tudelftnl, Susha, I., Janssen, M., Verhulst, S., & Pardo, T. (2017). *Data Collaboratives: How to Create Value from Data for Public Problem Solving?* https://doi.org/10.1145/3085228.3085309
Ismagilova, E., Hughes, L., Dwivedi, Y. K., & Raman, K. R. (2019). Smart cities: Advances in research—An information systems perspective. *International Journal of Information Management*, *47*, 88–100. https://doi.org/10.1016/J.IJINFOMGT.2019.01.004
Jetzek, T. (2016). Managing complexity across multiple dimensions of liquid open data: The case of the Danish Basic Data Program. *Government Information Quarterly*, *33*(1), 89–104.
Kitchin, R., & Moore-Cherry, N. (2020). Fragmented governance, the urban data ecosystem and smart city-regions: the case of Metropolitan Boston. *Regional Studies*. https://doi.org/10.1080/00343404.2020.1735627
Lex, S. W., Calì, D., Koed Rasmussen, M., Bacher, P., Bachalarz, M., & Madsen, H. (2019). A cross-disciplinary path to healthy and energy efficient buildings. *Technological Forecasting and Social Change*, *142*, 273–284. https://doi.org/10.1016/J.TECHFORE.2018.07.023
Luo, J. (2022). A Bibliometric Review on Artificial Intelligence for Smart Buildings. *Sustainability 2022, Vol. 14, Page 10230*, *14*(16), 10230. https://doi.org/10.3390/SU141610230
McAfee, A., & Brynjolfsson, E. (2012). *Big Data: The Management Revolution*. Harvard Business Review.
Meijer, A. (2018). Datapolis: A Public Governance Perspective on “Smart Cities.” *Perspectives on Public Management and Governance*, *1*(3), 195–206.
Monteiro, E., & Parmiggiani, E. (2019). Synthetic Knowing: The Politics of the Internet of Things. *MIS Quarterly*, *43*(1), 167–184.
OECD. (2016). *Data-Driven Innovation: Big Data for Growth and Well-Being*.
Oliveira, M. I. S., & Lóscio, B. F. (2018). What is a Data Ecosystem? *ACM International Conference Proceeding Series*.
Østerlie, T., & Monteiro, E. (2020). Digital sand: The becoming of digital representations. *Information and Organization*, *30*(1), 100275.
Parmiggiani, E., Østerlie, T., & Almklov, P. G. (2022). In the Backrooms of Data Science. *Journal of the Association for Information Systems*, *23*(1), 139–164.
Pickering, Andrew. (1995). *The mangle of practice : time, agency, and science*. 281.
Raetzsch, C., Pereira, G., Vestergaard, L. S., & Brynskov, M. (2019). Weaving seams with data: Conceptualizing City APIs as elements of infrastructures. *Big Data and Society*, *6*(1). https://doi.org/10.1177/2053951719827619
Sembroiz, D., Careglio, D., Ricciardi, S., & Fiore, U. (2019). Planning and operational energy optimization solutions for smart buildings. *Information Sciences*, *476*, 439–452. https://doi.org/10.1016/J.INS.2018.06.003

Suzuki, L. R., & Finkelstein, A. (2019). Data as Infrastructure for Smart Cities. In *Computing and Networks* (Vol. 23).

Tempini, N. (2017). Till data do us part: Understanding data-based value creation in data-intensive infrastructures. *Information and Organization*, *27*(4), 191–210. https://doi.org/10.1016/j.infoandorg.2017.08.001

Venters, W., Oborn, E., & Barrett, M. (2014). A trichordal temporal approach to digital coordination: The sociomaterial mangling of the CERN grid. *MIS Quarterly: Management Information Systems*, *38*(9), 927–949.

Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems*, *15*, 320–330. https://doi.org/10.1057/palgrave.ejis.3000589