Centre for Digital Built Britain

2018/19 General Project

Analysing Systems Interdependencies Using a Digital Twin

Final Reporting

10 July 2019

Name of Project PI	Affiliation
Professor Jennifer Whyte	Imperial College London
	Email address
	j.whyte@imperial.ac.uk
Co-I name	**Affiliation**
Professor Daniel Coca	University of Sheffield
Professor John Fitzgerald	Newcastle University
Professor Martin Mayfield	University of Sheffield
Dr Ken Pierce	Newcastle University
Professor Nilay Shah	Imperial College London
Research Team name	**Affiliation**
Dr Long Chen	Imperial College London
Dr Carl Gamble	Newcastle University
Dr Cristian Genes	University of Sheffield
Dr Filip Babovic	Imperial College London
Dr Akeem Pedro	Imperial College London

Please cite as: Whyte, J., Chen, L., Gamble, C., Genes, C., Pierce, K., Fitzgerald, J., Coca, D., Mayfield, M., Babovic, F., Pedro, A., Shah, N. (2019) Analysing Systems Interdependencies Using a Digital Twin, 2018/19 General Project funded by CDBB, Final Report.

Abstract
This work provides a first step towards next-generation systems engineering by demonstrating the feasibility of using a digital twin to generate new insight into systems relationships and interdependencies. This step required substantial interdisciplinary work and industry collaboration in order to examine the potential to combine a set of relevant analytical methods (e.g. BIM query, network analyses and multi-modelling). We assembled an experienced team (Imperial College London, University of Sheffield, Newcastle University), and worked closely with and used empirical data from a major project (Thames Tideway Tunnel). This first step delivers fundamental theoretical understanding that will support the use of the digital twin for systems analyses, and a practical contribution to the identification, prioritisation and management of interdependencies. The long-term ambition is to build the tools that decision makers need in order to understand infrastructure system interdependencies within and across project boundaries.

1. Introduction / Overview of the Project

1.1: Background

The term 'digital twin' has various definitions in infrastructure and manufacturing industries. Digital twins in infrastructure are defined as: *"realistic digital representations of physical things. They unlock value by enabling improved insights that support better decisions, leading to better outcomes in the physical world"* (Bolton et al., 2018, p.10)[1]. In manufacturing, Deloitte defines the digital twin as: *"a near-real-time digital image of a physical object or process that helps optimize […] performance"* (Parrott & Warshaw, 2017, p.2). Across these definitions, attention is drawn to the use of flows of digital information on physical assets to improve decisions, outcomes and performance, with the latter definition adding that such information may be near-real-time — a mirror image. Thus, while Computer-Aided Design (CAD) focuses on geometric information and Building Information Modelling (BIM) on both geometric information and other associated asset information, the digital twin may include information on behaviours, so it may model performance, and be used to understand mathematically the existing and predicted performance of infrastructure systems.

Questions about the digital twin are the focus of a growing trajectory of research on infrastructure, with there being major investments in work on urban observatories, as well as a Data & Analytics Facility for National Infrastructure (DAFNI), through the UK Collaboratorium for Research in Infrastructure and Cities (UKCRIC) as well as the work of the Digital Twin Hub (supported by the Centre for Digital Built Britain (CDBB)). While the shift from CAD to BIM involved the addition of asset information as well as geometry, the shift to a digital twin requires the use of a broader range of sources of data, which may involve geometries, asset information and associated time series data on processes. Such data may be generated through the activities of professionals, in both production and operation, and through a range of sensing devices (including photographs, laser scans, and embedded sensors). New questions arise: how to configure digital twins for different use cases? How to abstract and use design-time models to facilitate analyses? Also, as digital twins become aggregated across organisational boundaries, can they be used to understand systems interdependencies?

This research builds upon the insight that projects are interventions into existing infrastructure (Whyte et al., 2019), so as to explore how systems interdependencies can be analysed through the use of a digital twin. It provides a first step towards next-generation systems engineering by demonstrating the

[1] This was part of the 'Gemini Principles' developed by the Digital Framework Task Group, and its Digital Twin Working Group, as part of its work to develop a National Digital Twin following the National Infrastructure Commission's "Data for the Public Good" report (NIC, 2017).

feasibility of using a digital twin to generate new insight into systems relationships and interdependencies. There is significant value in this, as infrastructure involves highly interdependent transport, water, building, power and energy systems-of-systems. Unlike the defence, aerospace and automotive sectors, in which many classic methods and tools for systems engineering have been developed, infrastructure involves systems that, to a large extent, are open and interconnected and include assets of different ages. As these become increasingly cyber-physical, the delivery of projects that modify or expand the existing infrastructure is growing in complexity.

The use of digital data and analytics can enable engineering decision makers in such open, interconnected systems to rapidly understand interdependencies within infrastructure projects and across their boundaries. This work requires substantial interdisciplinary work and industry collaboration in order to examine the potential to combine a set of relevant analytical methods (e.g. BIM query, network analyses and multi-modelling) and to identify and develop relevant, new data-driven methods.

By working with Tideway, we engage with leading practice that already makes extensive use of a range of digital modelling techniques, using them successfully in decision making over two decades of project development and delivery. With increasing interest and definition on the concept of a digital twin, this work takes a theoretical approach, drawing on research disciplines, and provides steps towards a framework through which digital approaches can be implemented and potentially integrated.

1.2: Aims and objectives

The main aim of this research was to articulate the extent to which a digital twin can be used to generate new insight into systems relationships and interdependencies. To do so, associated objectives were to:

1. Identify and rank the importance of critical interdependencies emerging in Tideway, both in the infrastructure system and in the enabling production system;
2. Develop new approaches to identifying critical interdependencies in time for decision makers on the project to make decisions by linking digital data; and
3. Articulate, across different scales, the utility of and practical barriers to the use of different analytical approaches (e.g. BIM query, network analysis and multi-modelling) in relation to practical problems and use cases faced in delivery.

The research involves Imperial College London, the University of Sheffield and Newcastle University working collaboratively to evaluate the potential to connect the digital twin with analytic approaches such as multi-modelling and network analysis. The team have collaborated over the past two years, and are working with the Lloyd’s Register Foundation/Alan Turing Institute Data-Centric Engineering Programme and the UK Collaboratorium for Research in Infrastructure & Cities (UKCRIC).

The major industry partner is Tideway, the company responsible for building the Thames Tideway Tunnel, a major new sewer through the centre of London, with the ambition of protecting the River Thames from pollution. In this project there has been intensive use of digital modelling, with operational models driving design decisions, as well as an emergent set of increasingly sophisticated modelling practices developing over time (for a background on the project see, for example, Crawford et al., 2017).

1.3: Methods

This research is multi-method, with substantial interdisciplinary work and industry collaboration being undertaken in order to examine the potential to combine a set of relevant analyses. These methods include:

1. Field research – collecting and prioritising data to evaluate critical interdependencies through field research with Tideway. This involved workshops with all members of the research team, interviews with key stakeholders on the project, access to digital models and systems, and in-depth work to trace interdependencies and their consequences on project delivery.

2. Evaluation of systems analyses techniques – for different issues arising as a result of interdependencies, we seek to evaluate how a digital twin can be used in systems analyses. These analyses combined techniques from:

   a) Systems engineering (e.g. using model-based systems engineering approaches to identify infrastructure interdependencies and emergent properties, and to visualise and (where possible) quantify uncertainties) – presenting a matrix view of directional interdependencies;

   b) BIM query to support identification of system interdependencies from a digital model, using model-checking to automate classification and analyses of element/subsystem relationships;

   c) Network analyses (extending classical dynamical systems theory) – analysing nodes and connections in infrastructure systems and systems-of-systems; and

   d) Collaborative model-based engineering of cyber-physical interactions to deliver well-founded computational methods and tools that enable semantically heterogeneous models for smart infrastructure interventions to be constructed.

3. Synthesis, validation and presentation of findings – synthesising findings in relation to the techniques studied, to articulate, across different scales, and seek to validate findings with regard to the utility of and barriers to the use of different analytical approaches (network analysis, multi-modelling). This synthesis considers utility and barriers in relation to practical problems faced in delivery and available combined data (and uncertainties therein). We worked together as a team to present and evaluate the findings and present them back to the industry partners and a reference group on 2 July 2019.

2. Findings

The research findings suggest insights with which to further develop leading practice in order to achieve the ambition of using the digital twin for analysing systems interdependencies so as to understand projects as interventions in infrastructure systems-of-systems. Challenges include the sharing of data across organisational boundaries, the development of common ontologies and the identification of modelling methods for different use cases through the delivery and operations of infrastructure. There are also challenges in making the advanced modelling methods used in research available to infrastructure decision makers.

2.1: Identifying infrastructure interdependencies and emergent properties

2.1.1: The Tideway case

The Thames Tideway Tunnel (Tideway) case was chosen as an orienting case for this work as it involves a major ongoing intervention in London, developing infrastructure that crosses or touches a range of existing communication services, water mains, bridges, river walls, buildings, and other services and tunnels (https://www.tideway.london/). Figure 1 shows the Tideway scheme.

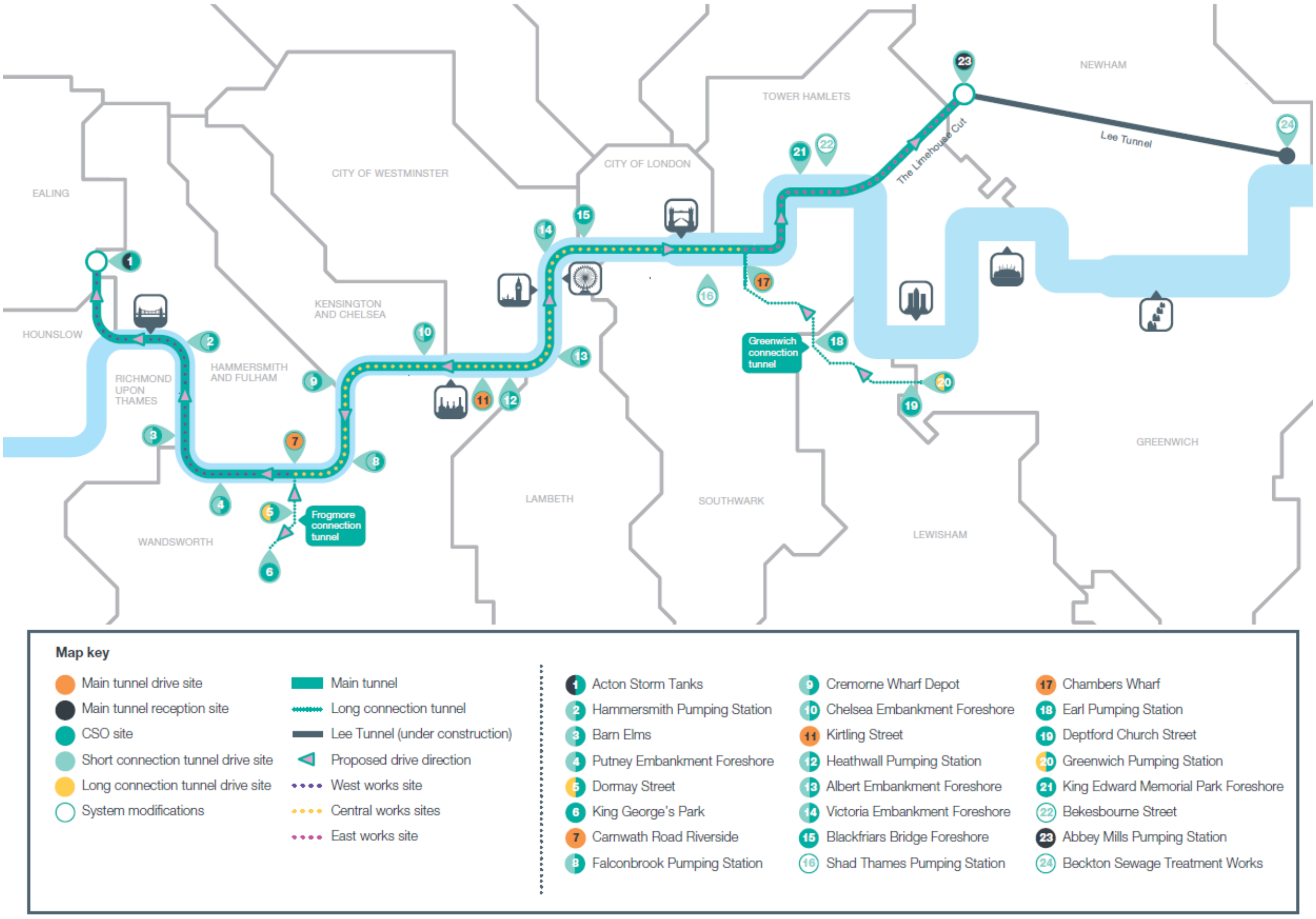


Figure 1. The route of the Tideway scheme (used with permission)

Delivery of the Thames Tideway Tunnel involves many organizations in planning, design and construction. Bazalgette Tunnel Limited (BTL) was set up to finance, build, maintain and operate the tunnel. There are three consortia responsible for the physical construction: BAM Nuttall Ltd, Morgan Sindall Plc, and Balfour Beatty Group (BMB), who are responsible for Tideway West; Ferrovial Agroman UK Ltd and Laing O'Rourke Construction (FLO), who are responsible for Tideway Central; and Costain Ltd, Vinci Construction Grands Projets, and Bachy Soletanche (CVB), who are responsible for Tideway East. AMEY is responsible for systems integration and process control, with project management being carried out by Jacobs.

The operation of the tunnel is scheduled to begin in 2024, with Thames Water and Tideway retaining responsibility for operations and maintenance. There will be relatively few sensors within the operating tunnel, as this is a challenging environment for data and electrical networks. Tideway connects with London's existing sewer network at a number of points. At these connection locations, there are components such as valves or penstocks and monitors that enable operation of the connected system. The modelling for this research (undertaken between October 2018 and July 2019) was partially shaped by the point in the project delivery at which the research team were active. It does not include historic modelling approaches that informed the initial development of the scheme requirements and operating philosophies.

To have a fully developed digital twin for use in systems analyses, as well as geometric information on the dimensions and location of the tunnel and the topology of the wider sewer network, practitioners would need temporal datasets of inflows or outflows, information on interacting systems (SCADA, sensor networks, etc.), and models of Combined Sewer Overflow (CSO) intercepts. Substantial modelling work has been done within Tideway. Detailed use of models for design decision-making includes work in areas such as simulation of historic and design storms (Crawford et al., 2017), air management (Geogaki et al., 2017), and fluid dynamics in the shafts (Plant et al., 2017). As such models become more inter-connected, steps are made towards a ‘virtual operating system’, however through this practice there is an emergent, rather than designed, approach to connecting these different types of models.

Our analysis of this orienting case built on a dataset developed from approximately 50 open documents taken from the Thames Tideway Tunnel website and includes CAD drawings and planning documents. In addition to the open data, Tideway were able to share some non-public datasets with the research team, but not all pre-existing model data.

2.1.2: Initial overview of interdependencies/evaluation

Following feedback from discussions with members of the Tideway team, we have modelled the Tideway project using both causal loop models and model-based systems engineering (in the Systems Modelling Language (SysML)). We divided infrastructure interdependencies into two broad areas: 1) interdependencies that arise during delivery and 2) interdependencies that arise during operation. In each case we consider interdependencies with existing infrastructure as well as within the assets that compose the project.

A first step was to articulate the various subsystems present within Tideway and determine their topology and potential interdependencies between the operation of the Thames Tideway Tunnel and interdependencies with London’s wider infrastructure system. After completing this high-level topology, the relationship between the systems and subsystems was articulated in order to determine the functional relationships between systems components. In doing so, it was possible to identify, at a high level, the dependencies that exist between Tideway’s components and subsystems at the asset, project and systems-of-systems level.

Figure 2 shows how, for any infrastructure project, interdependencies can be considered at the level of the: 1) infrastructure system-of-systems, 2) project as an intervention in that system-of-systems, and 3) assets delivered and how they are coordinated to deliver the outcomes and experiences planned through the project. As shown, the literature suggests that there will be geospatial, cyber, physical and logical interdependencies (Rinaldl et al., 2001) within and between these levels, with a range of emergent properties across levels.

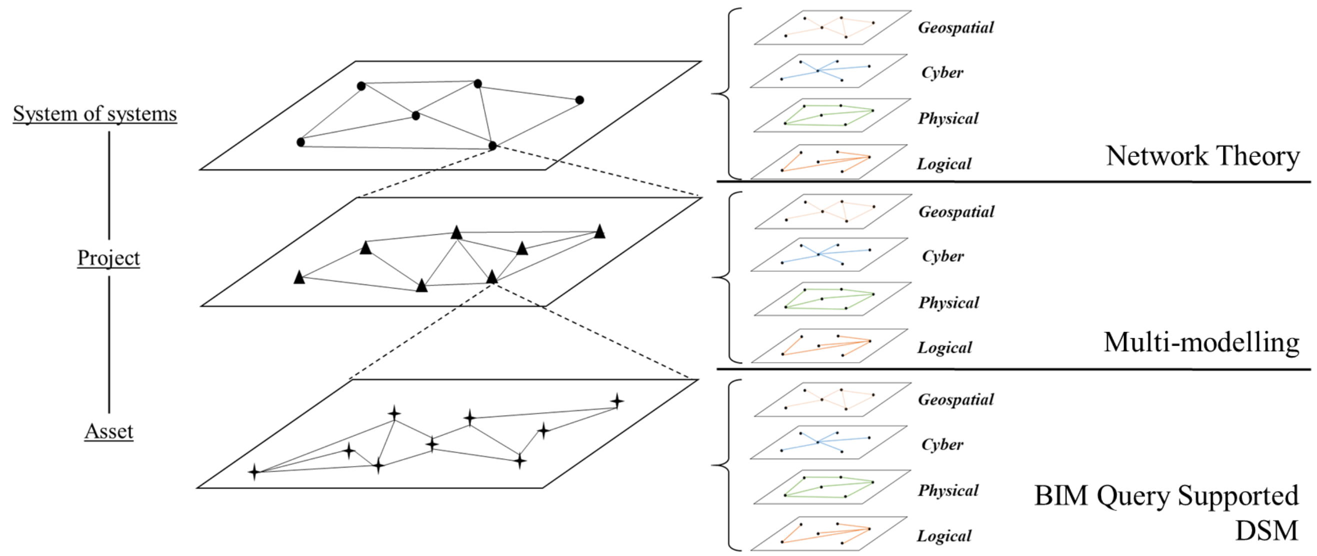


Figure 2. Interdependencies across levels (from the assets to the project and the system-of-systems), showing the level at which different modelling techniques were applied in this research

2.1.3: Prioritising systemic issues

We sought to understand how practitioners in the project prioritise and rank such known and emergent interdependencies. Feedback from professionals was on:

- *Operations: Interdependencies during operation / with existing infrastructure*
  - We were informed that the value of the digital twin lies in the maintenance and operations phase, using data delivered through the project; and
  - Yet, on Tideway there are relatively limited interdependencies with existing infrastructure (e.g. compared with Crossrail) due to the nature of operations and the depth of the tunnel, with relatively limited operational interdependencies beyond valves/penstocks etc.;

- *Design/delivery: Interdependencies that arise during delivery / with existing infrastructure*
  - Here, interdependencies with known existing infrastructure are not a major problem (but some interdependencies are not known);
  - To use existing digital models to their full ability, questions arise as to how data is created and disseminated through organisations involved in delivery; and
  - Using a digital twin would be a large step change in delivery processes.

Given the focus on emergent interdependencies that are not known and planned for but are identified during the process of delivery, the research team did not seek to quantify interdependencies but rather to gain a qualitative understanding and evaluate their impact.

2.1.4: Identifying intervention points and scenarios

From this work to understand key interdependencies and systemic issues, we identified a set of motivating questions or ‘use cases’, to inform our understanding of situations in which practitioners might want to use the digital twin to address an issue through analysing systems interdependencies, as well as what might be the associated intervention points and scenarios. This work revealed different understandings of infrastructure (or ontologies) that underpin different questions. Firstly, we consider the questions on operations, as the understanding of operations is fundamental to the understanding of projects *as interventions* into infrastructure systems.

- *Operations:* Such questions include:

1. How to manage the facilities according to updated operation records?
2. What is the effect on system behaviour of changing network topology or node properties?
3. How will the system behave given a set of environmental conditions?
4. Is the topology of my network as intended in the design stage? What are the interdependencies between my network and other infrastructure systems?
5. What is the current health status of the system? Does it require any interventions?

Figure 3 represents, for operations, these motivating questions and use cases, and how these arise across scales from the consideration of infrastructure as an asset or network.

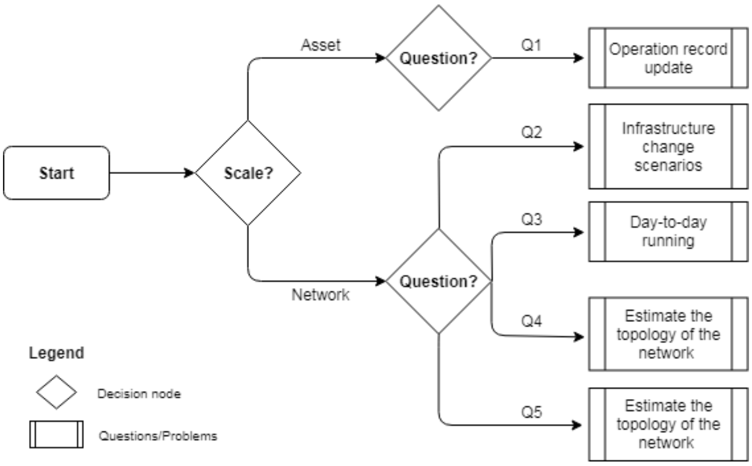


Figure 3. Example motivating questions on operations of infrastructure systems

- *Design/delivery:* Such questions include:

6. What are the system consequences of a late design change?
7. What are the operational consequences of asset interdependencies?
8. What are the interdependencies between the assets developed in the project and the wider infrastructure network? Or between assets developed in one part of the project and assets developed in another part of the project?
9. Which are the critical parameters in the design?
10. What are the most influential nodes in the network? How much of the network can they control?
11. How robust is my network design to node failures?

Figure 4 represent, for design/delivery, these motivating questions and use cases, and how these arise across scales from the consideration of infrastructure as an asset or network.

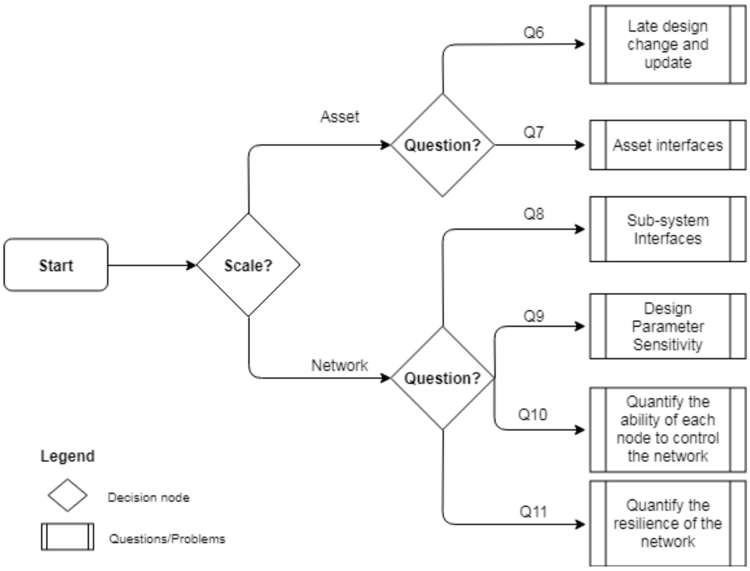


Figure 4. Example motivating questions on design/delivery of infrastructure systems

These questions have informed our modelling work. We have focused on identifying measures to identify interdependencies in operations and delivery in order to use a digital twin to reduce the risk of systems failures (such as a cascading failure). To articulate under which circumstances the various approaches had value, the teams have worked on a decision tree, as will be discussed later in this report, after we have discussed modelling methods and data in detail. The following sections of findings discuss using BIM techniques (in section 2.2), network analysis (in section 2.3), and multi-modelling (co-modelling) (in section 2.4). As noted above, techniques from both systems dynamics (causal loop diagrams) and model-based systems engineering (SysML) were used to obtain an overview of the systems from qualitative data.

2.2: Systems engineering and the digital twin

This work focuses on identifying system interdependencies from a digital model (BIM), using data obtained from Tideway and converted into the Industry Foundation Classes (IFC) file format. Section 2.2.1 focuses on identifying infrastructure elements, and their associated interdependencies across levels. Section 2.2.2 aims at developing the approaches and strategies for model checking through the use of BIM query methods at the asset level. Section 2.2.3 aims to demonstrate the potential of the these methods in aiding late-stage design through using the Thames Tideway Tunnel as an initial case study and providing strategies for model development to identify and analyse systems interdependencies.

2.2.1: Identify interdependencies from the digital twin

The Design Structure Matrix (DSM) provides a simple, compact, visual matrix representation of complex systems, which is effective in addressing decomposition and integration problems and managing iterative tasks (e.g. design process). There is a substantial body of research on the use of the DSM (e.g. Browning, 2001). However, manual effort is often required to clarify (and/or quantify) these interdependencies in the matrix, and researchers that use the DSM often proceed by reviewing architectural diagrams, system schematics and other relevant documents or interviewing the relevant engineers and experts. This research draws on a new trajectory of work that is beginning to seek to automate this process (e.g. Gopsill et al., 2016), and has focused upon developing a prototype method with which to generate a DSM from a digital twin using predefined element/subsystem relationships. As shown in Figure 5, this DSM provides a matrix view of directional interdependencies between elements (e.g. A has an impact on B).

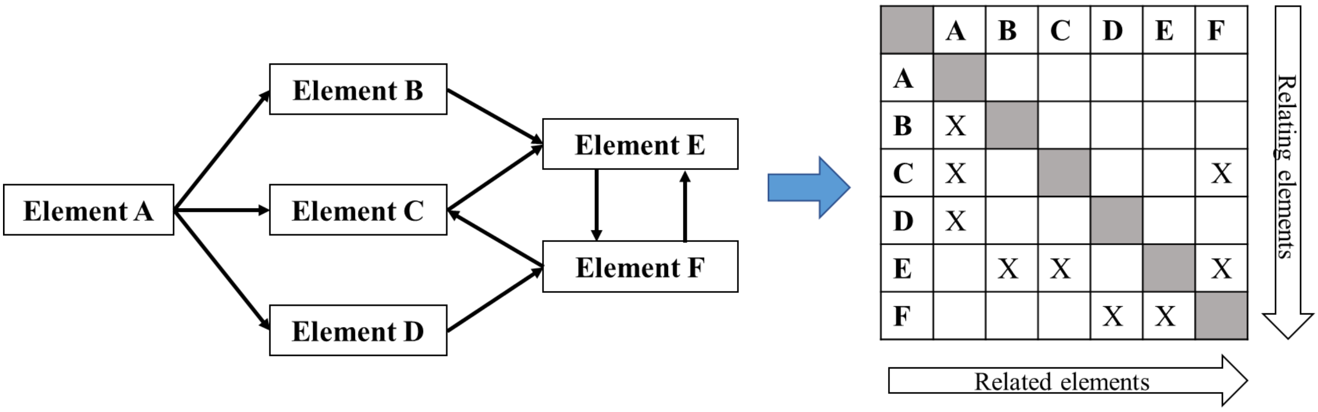


Figure 5. Related elements and the representation of their directional interdependencies in the DSM

We developed a Digital-DSM through the use of a BIM model provided by Tideway. Based on extant literature (BuildingSMART, 2013; Delany, 2019; Borrmann et al., 2015; Daum & Borrmann, 2014; Mazairac & Beetz, 2013; Nepal et al., 2012, etc.), two essential building blocks of the Digital-DSM (i.e. elements and interdependencies) were classified and identified. We further defined the elements for the Digital-DSM, at different scales drawing on the Uniclass 2015 classification system and prior work (Senthilkumar & Varghese, 2009; Pimmler & Eppinger, 1994; Saoud et al., 2017), as shown in Figure 6, and articulated in Chen & Whyte (2020).

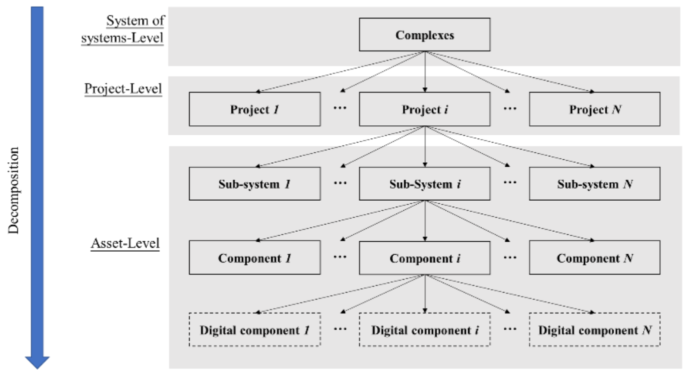


Figure 6. Structure of elements (complexes, project, sub-system, components, digital components) and interdependencies in the Digital-DSM

In this work we consider the *geospatial* (Daum & Borrmann, 2014; Mazairac & Beetz, 2013; Nepal et al., 2012; Borrmann & Rank, 2009; Egenhofer & Franzosa, 1991; Clementini & Felice, 1995), *physical* (Borrmann et al., 2015; BuildingSMART, 2013) and *logical* (BuildingSMART, 2013) interdependencies (not cyber, or cyber-physical, interdependencies) across different scales, as shown in Table 1.
This work provides the theoretical basis for developing a practical method of generating a Digital-DSM, based on geometric and semantic information within IFC data.

Table 1. Definition of interdependencies at different levels

	System-of-systems	**Project**	**Asset**
Geospatial	The regional or wider environmental event can create state changes in all complexes.	A local environmental event can create state changes in all projects involved in complex(es).	The physical adjacency or topological relationship between each asset.
Physical	The overall state of each complex is dependent on the material output(s) of the other.	The state of each project is dependent on the material output(s) of the other.	An objectified relationship between a material definition and elements or element types to which this material definition applies.
Logical	The state of each complex depends on the state of the other via a mechanism that is neither a physical, cyber nor geospatial connection.	The state of each project depends on the state of the other via a mechanism that is neither a physical, cyber nor geospatial connection.	The state of each asset depends on the state of the other via a mechanism that is functional, or that is neither a physical, cyber nor geospatial connection.

2.2.2: Strategy for model checking

Two strategies have been developed for automatically generating the Digital-DSM from IFC data (Table 2): (1) the IFC mining method, which directly adopts the IFC data generated in order to search for corresponding entities, geometric information, and IFC relationships in the lines of an IFC file through text mining or a program developed specifically; for example, the 'Foundations Piles' can be exactly identified by searching for this global identifier (GlobalId) in the lines of IFC (Figure 7); and (2) the BIM query method, which adopts BIM query languages for extracting entities, geometric information, and IFC relationships, e.g. BIM Query Language (BIMQL) (Mazairac & Beetz, 2013) and 3D Spatial Query Language (Borrmann & Rank, 2009).

```
#101356=IFCBUILDINGELEMENTPROXY('1Jzfx_e9jHTY_A$E26KkLK',#42,'Civil-Concrete--C-G215-M_Foundations
Piles','1,PW_WORKDIR:dms06592\\100-DW-CVL-PWR1X-363760.dgn,Default:133980','Civil-Concrete:C-G215-M_Foundations
Piles',#102161,#1946,'133980.245406*100-DW-CVL-PWR1X-363760!Default',$);
```

Figure 7. IFC mining method for identifying 'Foundations Piles'

- In this research, for identifying elements, both the IFC mining method and the BIM query method can be useful and reliable.
- For identifying geospatial interdependencies, this work used the BIM query method along with the 9-intersection model (9-IM) developed by Egenhofer and Franzosa (1991) and Clementini and Felice (1995), where the geospatial interdependencies at the asset level were defined according to the topological relations between interior, boundary and exterior elements (Egenhofer & Franzosa, 1991; Borrmann & Rank, 2009).
- For identifying physical and logical interdependencies, both the IFC mining method and the BIM query method are recommended to search for specifications in an IFC file (Table 2).

Table 2. Strategy for identifying interdependencies at the asset level using IFC data (for geospatial, physical and logical interdependencies)

Interdependencies	Specifications	Strategies
Geospatial	*Disjoint; Contain; Within; Touch; Overlap*	BIM query method (e.g. BIMQL, 3D Spatial Query Language)
Physical	*IfcRelAssociatesMaterial*	IFC mining method/BIM query method
Logical	*IfcRelAssignsToGroup*	IFC mining method/BIM query method
	IfcRelAssignsToResource	IFC mining method/BIM query method

2.2.3: Demonstrating the approach

The potential of the Digital-DSM and generating approaches in analysing systems interdependencies and predicting the outcome of late-stage design changes is demonstrated through the use of an Open BIM approach based on IFC. This research used the Tideway project as a case study to demonstrate how our approach/strategy can be applied in analysing systems interdependencies in practice, which further helps to clarify the complexity in complex systems.

Two strategies have been adopted in the case study. The IFC mining method was adopted to address the IFC data (Figure 8a) generated from the DGN format file in AECOsim Building Designer. A program using Matlab (Figure 8b) has been developed to identify elements and interdependencies in an IFC file, which succeeded in identifying all elements at the asset level (Figure 8c). The BIM query method has also been adopted using the BIMserver and BIMQL engine (Figures 9a and 9b) for querying elements (Figures 9c and 9d).

We have also tried to identify the interdependencies in the provided data from Tideway, however, only limited interdependencies have been identified (Figure 9b) because the raw data is in DGN format, which was converted into an IFC format using AECOsim Building Designer, but much information can be lost during this process. With well-defined IFC data or BIM models provided, our approaches are able to provide satisfactory results for identifying elements and interdependencies with which to construct the Digital-DSM.

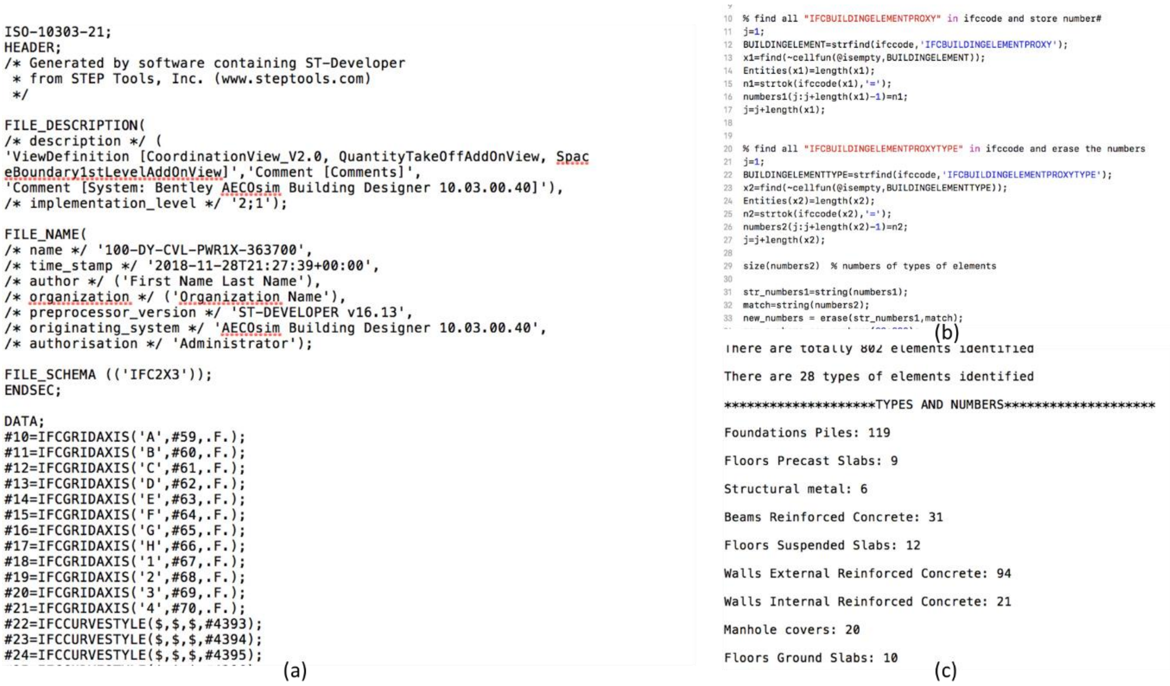


Figure 8. IFC mining method for identifying Digital-DSM elements and interdependencies

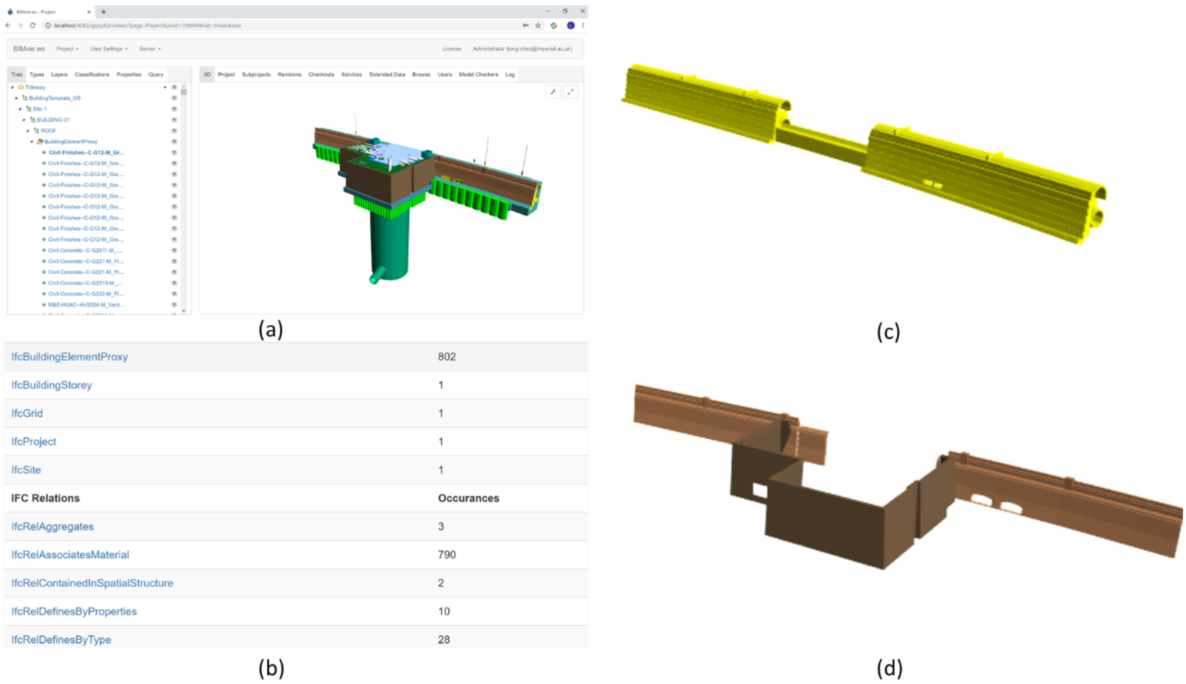


Figure 9. BIM Query Method (BIMQL) for identifying Digital-DSM elements and interdependencies

2.3: Network analyses and the digital twin

This work explores the use of network analysis in identifying interdependencies within a digital twin framework. Infrastructure systems are complex networks that consist of different interconnected components. Because of the complex nature of these systems, it is difficult to determine all relevant interdependencies between different types of infrastructure. While interdependencies that are created by physical interconnections are easy to identify and map, there are subtle interconnections within and between infrastructure systems which cannot be determined outright and which require analytical tools to be identified. Network analysis and network identification in particular are important tools that can be used to detect hidden interdependencies in complex networks. In contrast to the DSM approach presented above, the network identification method used in this section produces an estimate through the use of time-series data collected from the infrastructure system.

Network analysis and identification methods are essential to the development of digital twins for national infrastructure in both design and operational phases. In the design phase, network analysis tools, such as network controllability, resilience metrics, etc., can inform the optimal deployment of sensors, the design of the control systems, and can form the basis for robust design optimisation. In the operational stage, network identification and estimation methods are required to continuously update the structure and parameters of the digital twin in order to identify faults, changes in topology or unusual behaviour.

2.3.1: Network analyses and the Tideway case

As an example, the Thames Tideway Tunnel is a major intervention into the existing London infrastructure which creates new interdependencies with and between the existing infrastructure through physical connections with the existing sewer network, power grid, and control and communication network. Network analysis tools can be used to quantify the impact of the intervention on the robustness and resilience of the infrastructure systems that are affected by the project.
While the majority of interdependencies created are known, it is very likely that the project will create additional interdependencies that are difficult to identify and quantify at the design stage, and which may only manifest under certain conditions in the operational stage. In this context, by using network identification tools to analyse real-time sensor data, it is possible to identify and characterise existing hidden interdependencies or interdependencies that develop at a later stage during the operational lifetime of the sewer. In addition, the information provided by the network identification tools can be used in conjunction with the digital twin to implement condition-based monitoring and predictive maintenance, so as to detect leaks or blocked conduits and determine maintenance needs.

Interdependencies between infrastructure systems are modelled as a system of coupled networks in which nodes from one or more subnetworks depend on nodes in other subnetworks. Analysing the interdependencies is important because the failure of nodes in one network can lead to the failure of dependent nodes in other networks, which results in a cascade of failures (Kenett et al., 2014). Consequently, a failure in a small fraction of nodes in one network can have a significant impact on the other interdependent networks. A real-world example is that of the 2003 Italy blackout, which was caused by a cascade of failures between the power stations and the communication infrastructure (Buldyrev, et al., 2010). To prevent such events, it is important to understand potential vulnerabilities in infrastructure systems (Dunn, et al., 2013). To this end, percolation theory is introduced (Cohen, et al., 2000) in order to study the stability of interconnected networks. In this framework, the impact of removing a node or an edge from one network is assessed so as to define the network reliability and analyse the network failure process (Li, et al., 2015). Furthermore, by analysing complex systems as coupled networks, it is possible to deliver additional insight that may alter the basic assumptions upon which network theory relies for the single networks case (Buldyrev et al., 2010; Parshani, et al., 2010).

One possible application of the network identification tools is to estimate where buried infrastructure that is not currently known might exist. Finding unknown buried infrastructure can cause delays in the construction stage and can generate disruptions in other urban infrastructure systems. Estimating their position aims to mitigate such delays. From a network perspective, this is a partial network identification problem in which the nodes and part of the edges are known, and the complete set of edges need to be estimated. When real data is available at each node of the network, the identification tools can be used to infer where additional edges exist between the nodes of the complex network. However, this approach does not provide an estimate of the exact location of the buried infrastructure. Instead, it provides an estimate of the start and end points (origin and destination nodes) of the subterranean infrastructure that was not known to exist.

2.3.2: Resilience and robustness through network analyses

In addition to the network identification tools, once the topology of the complex network is known, different resilience measures are available in the network analysis literature for assessing the robustness of the network design. A review of different resilience measures that can be used for transportation infrastructure is presented by Sun, Bocchini and Davison (2018). These include travel time, congestion index, resilience index, and capacity and can help to identify vulnerable components in the transportation infrastructure.

To demonstrate the approach, in the absence of real data from the sewage system, we introduce a synthetic model of two interdependent infrastructure networks, as depicted in Figure 10, one which represents a major project intervention, such as the Thames Tideway Tunnel, and the other representing the existing infrastructure system, which could be the existing sewer network, the power grid or a communication system (Kenett et al., 2014).

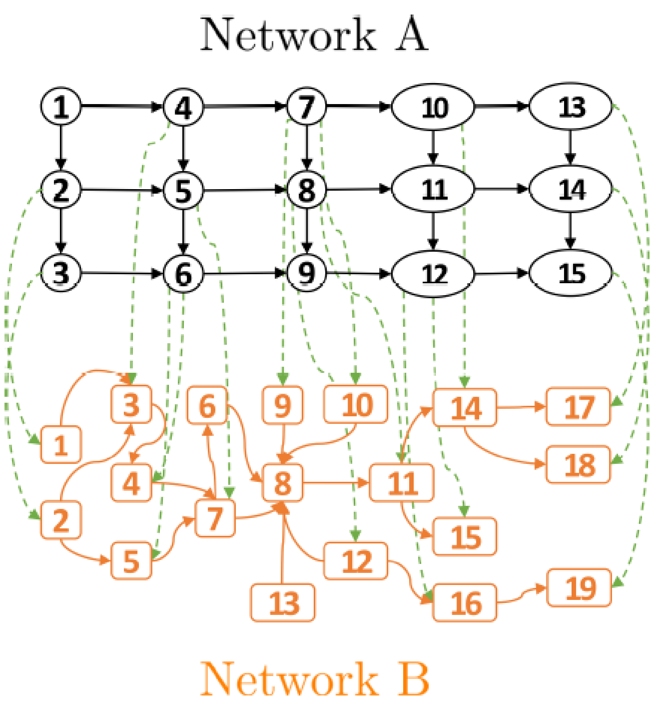


Figure 10. Complex network that contains two interdependent infrastructure systems

Our synthetic model assumes a large number of interdependencies between the nodes of the two networks. In order to capture these interdependencies in the identification process, the two networks are modelled as a single, complex network which contains the nodes from both networks.

We apply a network identification method based on Conditional Granger Causality (CGC) (Ding, et al., 2006) in order to determine the network topology in a worst-case scenario, while assuming no prior knowledge of the topology of either network. Arguably, the assumptions made in this numerical study are more challenging, as we do not assume any a priori partial knowledge of the topology of the networks involved, which is available for the Thames Tideway Tunnel with the existing sewage network.

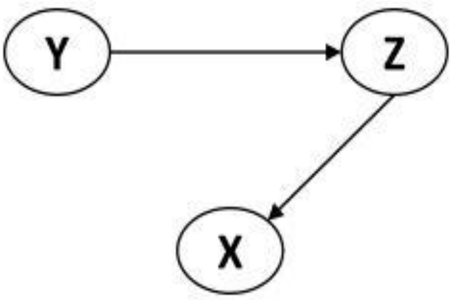
The network topology is inferred by evaluating the statistical interdependence between spatiotemporal data streams generated by sensors located in nodes of the network.

Conditional Granger causality is used to estimate the complete network topology for each infrastructure system, plus the interdependencies between the two networks. The rationale for this is to illustrate the capabilities of this approach in a more challenging scenario.

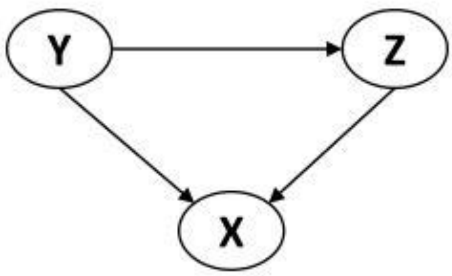
Each node in the complex network in Figure 10 is assumed to be equipped with a sensing unit that measures a relevant physical quantity (e.g. water flow, electrical current, etc.) at given time instants. Thus, time series data is generated in each node of the complex network. Based on this data, CGC can be used to infer the topology of the network. Granger causality relies on the idea that node X depends on node Y when the time series generated in node Y can be used to improve the prediction of the data for node X (Wiener, 1956). Granger formalised this idea in the context of linear regression models (Granger, 1969). An extension to the nonlinear case using kernels is presented in Marinazzo, Pellicoro and Stramaglia (2008). An alternative approach to Granger-causality-based identification is to use transfer entropy to measure the time-directed information transfer between jointly dependent processes. Interestingly, the Granger causality and the transfer entropy are equivalent for Gaussian variables (Barnett, et al., 2009).

2.3.3: Approach to Granger causality

The main limitation to using only Granger causality is that it cannot distinguish between a direct and an indirect connection between two nodes. For example, an analysis using only the Granger causality between any pair of nodes cannot distinguish between the network depicted in Figure 11a and the one shown in Figure 11b. In both cases there is an interdependence between Y and X, but for the case in Figure 11b there is also a direct relationship between Y and X. To overcome this limitation, CGC is used to determine whether or not the interdependence between Y and X is intermediated by Z. Thus, CGC is used in the following to identify the complete network topology from data collected at each node.



a) Example of an indirect connection between Y and X



b) Example of a direct and indirect connection between Y and X

Figure 11. Two cases of interdependence between Y and X

To evaluate numerically the performance of this approach, the complex network topology in Figure 10 is used to generate data at each node. For the first set of numerical results, it is assumed that the data at each node depends only on values from other nodes at the previous timestamp. Moreover, no partial information on known edges is incorporated into the estimation process. In the numerical experiments, all of the edges are estimated based on the data available from each node.

Using CGC-based analysis, the topologies of network A and network B are successfully identified. The numerical experiments are conducted using the Matlab Multivariate Granger Causality (MVGC) toolbox (Barnett & Seth, 2014), and the simulation time on an i7 7700HQ processor with 16 GB of RAM is around 1.5 seconds. This work can be used to identify interdependencies between network A and network B, and vice versa. Specifically, all of the interdependencies assumed in the model presented in Figure 10 are successfully identified using CGC-based analysis.

However, these simulation results assume that data at each node depends only on values from other nodes at the previous timestamp. This is equivalent to a one-time-period propagation delay between any two nodes of the complex network. To relax this assumption, the following numerical results consider a larger value for the maximum propagation delay. In the numerical experiments, the value of the propagation delay is estimated using the Akaike Information Criterion (McQuarrie & Tsai, 1998).

In practice, some connections/interdependencies might produce an effect after more than one time period. For example, in a sewage system, different pipe lengths generate different propagation delays between the nodes of the network. To analyse this scenario, the following numerical results assume a propagation delay of up to three time periods. This means that a change of value in one node can take up to three time periods to produce an effect on the other interdependent nodes. In practice, the value of the propagation delay depends on the types of infrastructure systems being modelled. Because different infrastructure systems operate with different timescales, the propagation delay parameter needs to be tuned based on the “slowest” system incorporated into the model.

The numerical results obtained for the case in which the maximum propagation delay is equal to three time periods are similar to those of the previous case. In particular, all of the edges between nodes from network A and all of the edges between nodes in network B, as well as all of the interdependencies between the two networks, are successfully identified. The main difference between the two cases with different propagation delays is that the computational time for the second case is around 30 seconds on the same i7 7700HQ processor with 16 GB of RAM. This shows that this identification method is scalable and can deal with large networks in different propagation delay scenarios.

2.3.4: Conclusions

CGC has been successfully used to demonstrate that interdependencies between two infrastructure networks can be inferred from time series data recorded at the nodes of the networks. Using data generated for each node of the directed network model depicted in Figure 10, all of the connections between the nodes are identified using CGC and the MVGC toolbox. The numerical simulation study carried out demonstrates that the proposed approach can correctly identify all interdependencies even in the presence of variable propagation delays along the edges of the network. The low computational costs of the method make it suitable for large infrastructure systems.

For sewage systems, CGC-based network identification could be used for detecting blockages in pipes. From a network perspective, a blocked pipe is equivalent to a missing edge between two nodes. In this context, online network identification carried out using real-time data can be used to detect both blockages and leaks by comparing the topology estimated in real time with the topology from the design stage.

The main limitation of CGC-based network identification is that it does not provide an estimate of the strength of the interdependence and does not capture aspects such as congestion or partial capacity loss (Goldbeck, et al., 2019). More advanced network identification algorithms can be used to identify not only the connectivity but also the dynamics of the flows along the edges of the network, providing important insight into the state of the system in the operational stage. Dynamic network identification requires a longer computational time but provides an estimate of the strength and dynamics of the interdependence, allowing the operator to observe how the properties of the network change over time and to make predictions. A general framework for dynamic network identification for complex systems is presented by Van Den Hof, Dankers, Heuberger and Bombois (2013). For urban infrastructure systems, a dynamic network identification model that combines network and asset representations is proposed by Goldbeck et al. (2019).

Another aspect that makes this modelling approach difficult to implement in practical settings is that it requires data to be shared between different organisations. Precisely, the complex network model assumes the availability of data streams from all of the integrated infrastructure systems for the identification process. In practice, different organisations might be reluctant to share their data.

The network identification based on Granger causality and the dynamic network identification tools can be used sequentially to provide a better estimate of the system's health status and predict maintenance needs. Precisely, static network identification which has a low computational cost can be performed for the whole network, while dynamic network identification is used only in parts of the network where there are deviations with respect to the expected network topology. Following the same rationale, the low computational cost of CGC-based network identification can be combined with a more computationally demanding approach such as multi-modelling. Network identification can be used to identify regions of the network in which there are deviations with regard to the desired functionality and a more in-depth analysis of that region can be conducted using the multi-modelling approach presented below.

2.4: Multi-modelling and the digital twin

Multi-modelling is an approach to the modelling of a system where instead of attempting to model the whole system behaviour as a single model or a single paradigm, we allow the system to be decomposed into multiple component models. Each of these component models may then be produced using the most appropriate modelling tools and techniques, before being exported as a Functional Mockup Unit (FMU). For example, a software system is typically described as a Discrete Event (DE) model, while a physical process such as fluid flows is better described using differential equations in a Continuous-Time (CT) environment. A multi-model configuration then describes how these FMUs should be connected in order to form a holistic model of the complete system.

The INTO-CPS toolchain[2], upon which the multi-modelling will be based, includes a Design Space Exploration (DSE) capability (Fitzgerald, et al., 2017). The DSE functionality is generally employed to find the optimal designs of a system by varying the values of model parameters and then observing the performance of the simulated system using one or more objective measures. It is also possible,

[2] http://http://into-cps.org/

however, to use the DSE functionality to vary parameters that, rather than changing the design of the system, change the environment in which the system is simulated.

2.4.1: Identify abstraction levels and modelling needs
This section aims to show how multi-modelling may be used in infrastructure projects, such as the Tideway project, in which work not only crosses organisational boundaries but also includes both cyber and physical elements.

The multi-modelling process begins with the decomposition of the system that is to be modelled into a set of models. This decomposition is generally performed using a system modelling language such as SysML. At this stage we also identify, at an abstract level, the nature of the connections between those models. In the case of a simplification of the Thames Tideway Tunnel example, we identify four subsystems that will form the basis of the multi-model, i.e. the existing sewer network, the Thames Tideway Tunnel, control of the penstock gates, and the treatment works fed by the sewers and the Thames Tideway Tunnel, as depicted in Figure 12.

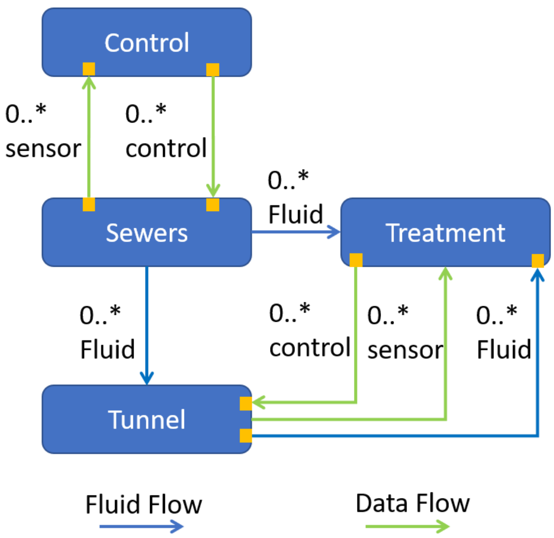


Figure 12. Concept SysML

We may then explore the many options within this conceptual structure, considering different numbers of connections between the models, where these connections represent the data or physical transfers between the modelled systems. Instances of the multi-model structure in which the cardinalities of all connections are specified may then be used as the skeleton for building detailed models of the four identified components. These detailed models contain further details regarding the cardinality and connectivity of the subcomponents that they contain; for example, Figure 13 shows a more detailed model of the control and sewer components, including the penstocks, sensors, and other subcomponents. It should be noted that this approach extends in principle to rainwater and hydrology.

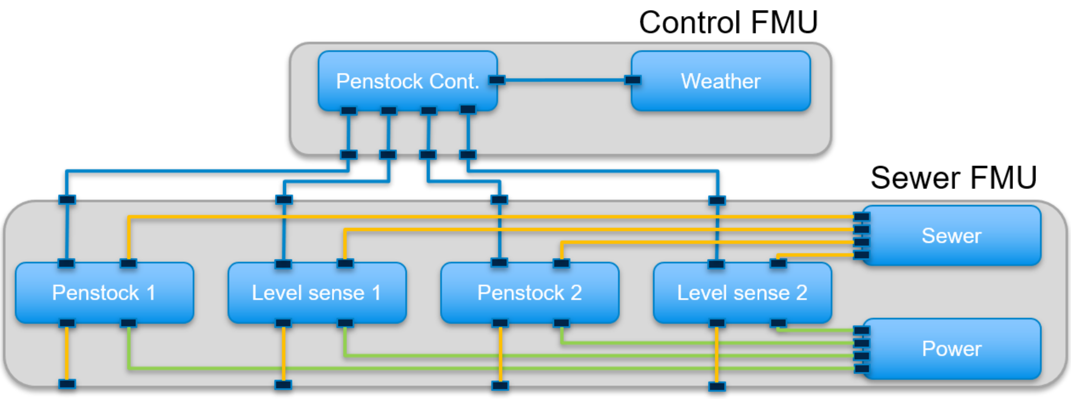


Figure 13. Further details added to the SysML model of the control and sewer components

The detailed SysML models can then be used to help generate discrete-event or continuous-time models, which may then be exported as FMUs. The resulting FMUs are then employed in multi-models and simulated in order to learn how that particular system configuration will behave. When a suitable system configuration has been obtained by exploring a range of potential configurations, the multi-model configuration and parameters of the generated FMUs may then be used as part of the specification for the next stage in the design process, such as the creation of detailed 3D geometry in BIM models. The multi-model itself may then be retained and maintained, forming the basis of the operational digital twin.

2.4.2: Example model

A representative portion of the London sewer network and the Thames Tideway Tunnel has been determined that includes the significant elements that we have identified that would exist in the full sewer network/Thames Tideway Tunnel (Figure 14). This system contains:

- *Street-level sewers* that capture rainfall at multiple points and transport this water to the main sewer. The street-level sewers are a simplification of the sewer network that spreads throughout London.
- *The main sewer*, which receives water from the street-level sewers and transports it to a treatment plant. The main sewer is connected to CSOs and penstocks, and possesses water-level sensors at each of these points.
- *CSOs*, which are the points at which water can flow from the main sewer into the Thames. Each of these can be parameterised by the height of the lip, indicating the maximum water level at that point before discharge begins.
- *Penstocks*, which are the points at which water may flow into the Thames Tideway Tunnel itself. Flow into the tunnel is regulated by penstock doors that may be opened and closed by the controller.
- *A treatment plant*, which represents a water treatment plant. The process by which water is treated is not modelled; instead, this component has a fluid storage capacity and a rate at which such water is processed.
- *Tideway*, which represents the Thames Tideway Tunnel itself. It receives water from the penstocks (when open) and transports it towards the pumps.
- *Pumps*, which lift the water from the end of the Thames Tideway Tunnel and feed it into the treatment plant.

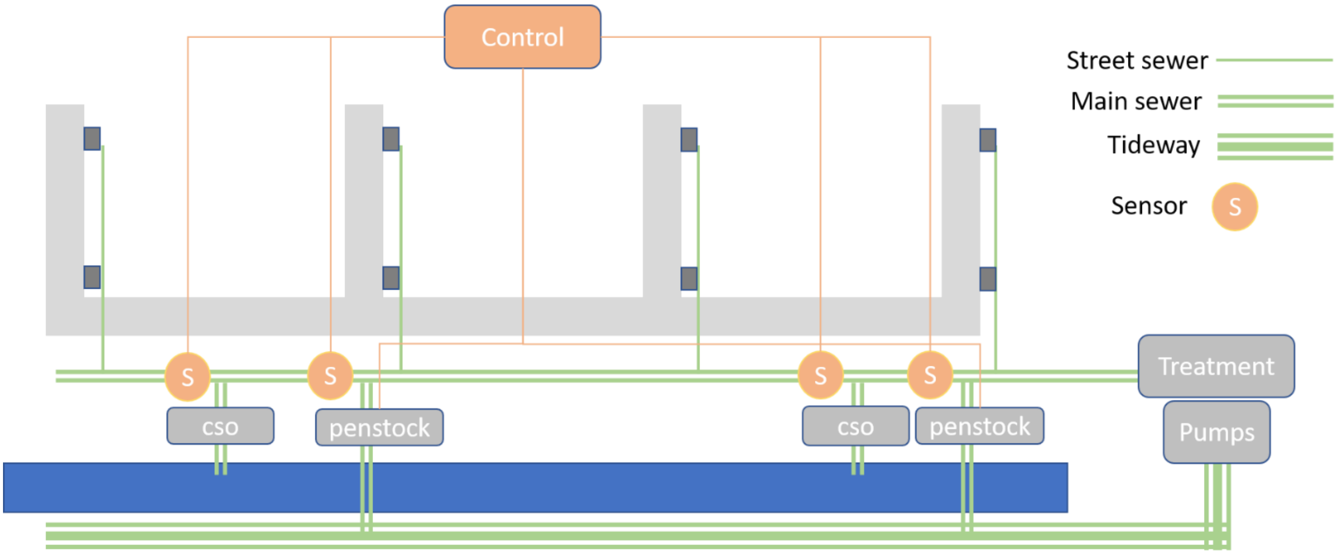


Figure 14. Example sewer system for demonstration

The system is decomposed into four models that are inspired by the organisational boundaries that exist in the Tideway project, as depicted in Figure 15.

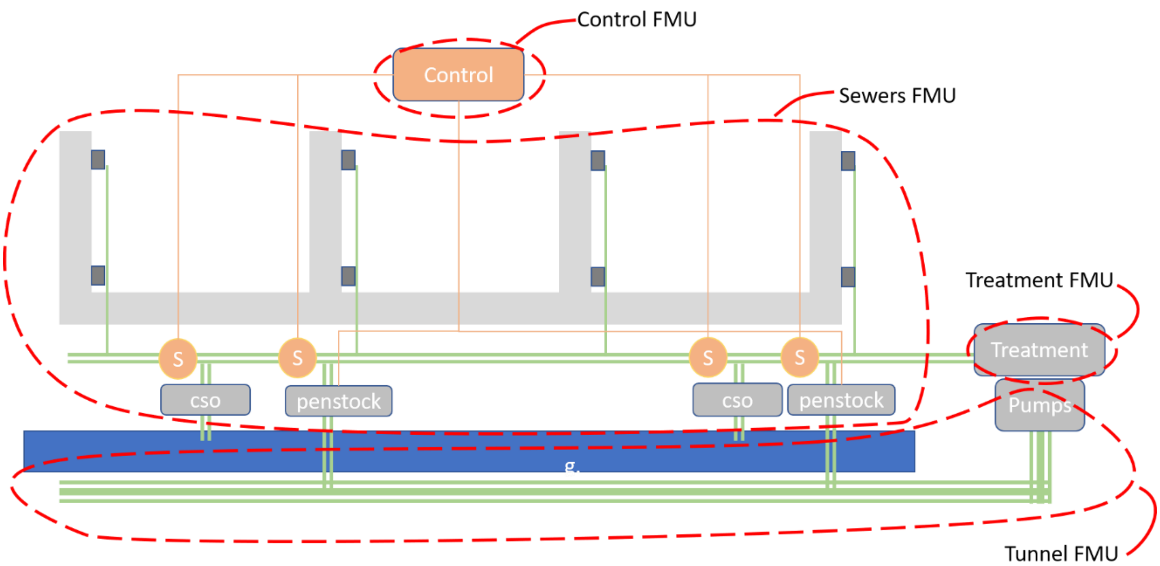


Figure 15. Decomposition of system into FMUs

The “Sewers” model represents an existing sewer infrastructure. In reality this model should span the whole sewer and rainwater catchment in the urban area; here it contains the street sewers, main sewers, CSOs, penstocks and sensors of the example system. Since this model is very physical in nature, it is constructed using differential equations in a CT environment. While the main elements are physical in nature, the model also contains the fluid-level sensors that provide the control system with the data it requires; and these sensors form part of the boundary between the physical and cyber domains. In this case, the sensors will simply output a signal that is proportional to the distance of the sensor from the fluid surface and, therefore, may easily be represented in the CT environment.

The “Control” model represents control logic that opens and closes the penstock gates according to the sensor levels. It will be responsible for computing the fluid levels based upon the signals received from the sensors in the Sewers model and using that information to control the opening status of the penstock gates. Since this model is concerned with control, it is constructed using the VDM-RT DE environment.

The “Tunnel” model represents the Thames Tideway Tunnel itself and is very much a model of the physical behaviour of the fluid that flows along the tunnel, along with the pumps that it contains. It houses no sensors other than those around pumps that remove fluid from the tunnel in order to feed the treatment plant.

The final element in this multi-model is the “Treatment” plant. The model of the treatment plant considers the process to be a single tank that has a maximum volume and processing rate, which fits well with the CT modelling paradigm. The treatment model has access to the sensor data around the pumps in the tunnel and uses this to control the operation of the pumps.

In terms of exchanges between the models:

- Control reads the level sensors in the sewer
- Control sets the opening state of the penstocks in the sewer
- Sewer computes the rate of fluid entering the tunnel via the penstocks
- Sewer computes the flow rate of fluid entering the treatment plant
- Treatment plant reads the fluid-level sensors at the pumps in the tunnel
- Treatment plant sets the pump speed
- Tunnel computes the flow rate of fluid entering treatment via the pumps

The multi-model resulting from the creation of these models is shown in Figure 16.

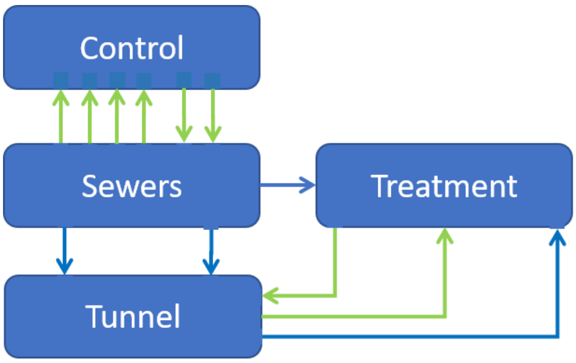


Figure 16. Structure of example multi-model

2.4.3: Systematic analysis of scenarios/interventions

A challenge in any large-scale project, such as the Thames Tideway Tunnel, lies in the time required to design, build and commission both the physical and the cyber assets of which it is composed. In order to optimise the time taken to commission the system and to minimise the time between completion of construction and achieving project sign-off, Tideway are exploring the development of a predictive digital twin. The intention of this is to support the training of controllers and operators of the system and maximise the understanding of system operations performance through each rainfall event available as compared to the system design using a combination of available models, e.g. rainfall data, sewer models and a control system representation. Initially, the digital twin should be verified as being predictive by comparing simulation results with actual sensed data. Then, by exercising the twin through a wide range of rainfall scenarios using the DSE facility and creating objective functions that report on the quantity of sewage discharged into the Thames during each scenario, it is possible to describe the range of scenarios in which the outflow quantity meets the requirements.

2.4.4: Analysis supported by the multi-modelling approach

The multi-model may now be used for analysis of the system. The first analysis was a by-product of the process of decomposing the system into FMUs. As outlined previously (section 2.4.1), by identifying the interfaces of each of the elements which are to be modelled, connecting them to the interfaces of the FMUs and then connecting the interfaces of the FMUs, we are making explicit the operational dependencies between these modelled elements. Figure 13 shows some of the internal elements within the control and sewer FMUs along the connections between these elements. This analysis addresses questions 7 and 8 of the decision tree, as discussed in section 2.1.4 (Figure 4) and further discussed below.

The second form of analysis arises from conducting a simulation of the model. Here we may observe the behaviour that the model was created to represent, e.g. the variable traces as shown in Figure 17. When conducting a simulation the engineer may choose to make use of the ability to replace one FMU representation of the model with another, depending on the purpose of the analysis. For example, if the engineer wishes to understand the effects of making significant changes to the system, such as multiple changes to the connectivity in a sewer network, then the engineer may make use of a high-fidelity FMU, based upon expensive numerical methods such as finite element fluid mechanics. Here the BIM data described in section 2.2 would be used to construct the models. On the other hand, for the day-to-day running of a system, the operations manager may utilise a machine learned model generated from recorded data. Here the training of such a model may be expensive, but the use of the model may be relatively less expensive, facilitating faster simulation of the system behaviour, so long as it remains within some bounds of the original training data. Here we might use a system identified model, as described in section 2.3.

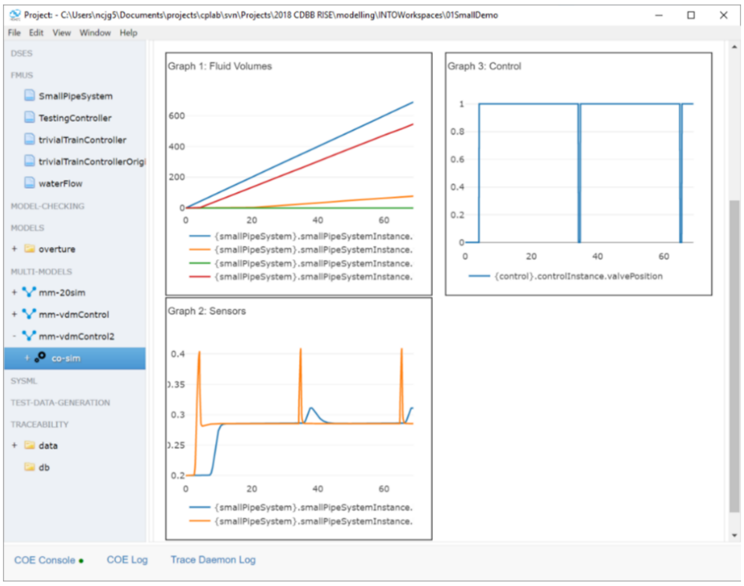


Figure 17. Representative portion of the example system produced

The final analysis that will be described here is referred to as Design Space Exploration (DSE). The primary role of DSE is to optimise a system towards some goals by varying the values of one or more parameters of the modelled design until an optimal set of parameters are obtained. In such an analysis the engineer defines the set of parameters that are to be explored, along with a range of values for each parameter, thereby defining the search space. They also define a set of objective functions which act upon the results of the simulation to produce measures that characterise the behaviour of the simulated system in a way that is important, such as the energy used by a vehicle or, in the case of the Tideway example, the quantity of fluid released via the CSO. Such an approach is suited to analysing system-level properties such as flexibility or robustness.

It is possible to use the DSE facility to determine the relationship between the parameters and the objective measures. By considering the amount of variation of each objective measure for each value of each parameter, we may determine which parameters are most influential with respect to the objectives, and from that determine which parameters must be controlled in the design. This addresses question 9 of the decision tree as discussed in section 2.1.4 (Figure 4) and further discussed below.

3. Discussion

To summarise the modelling approaches and sources of data discussed above, and how they relate to the ‘use case’ questions on *operation* raised in section 2.1.3, Figure 18 shows a decision tree for operations. This expands upon Figure 3, to show the questions and decisions (from left to middle) connected with data sources and modelling techniques (from right to middle) that may be associated with systems analyses using the digital twin.

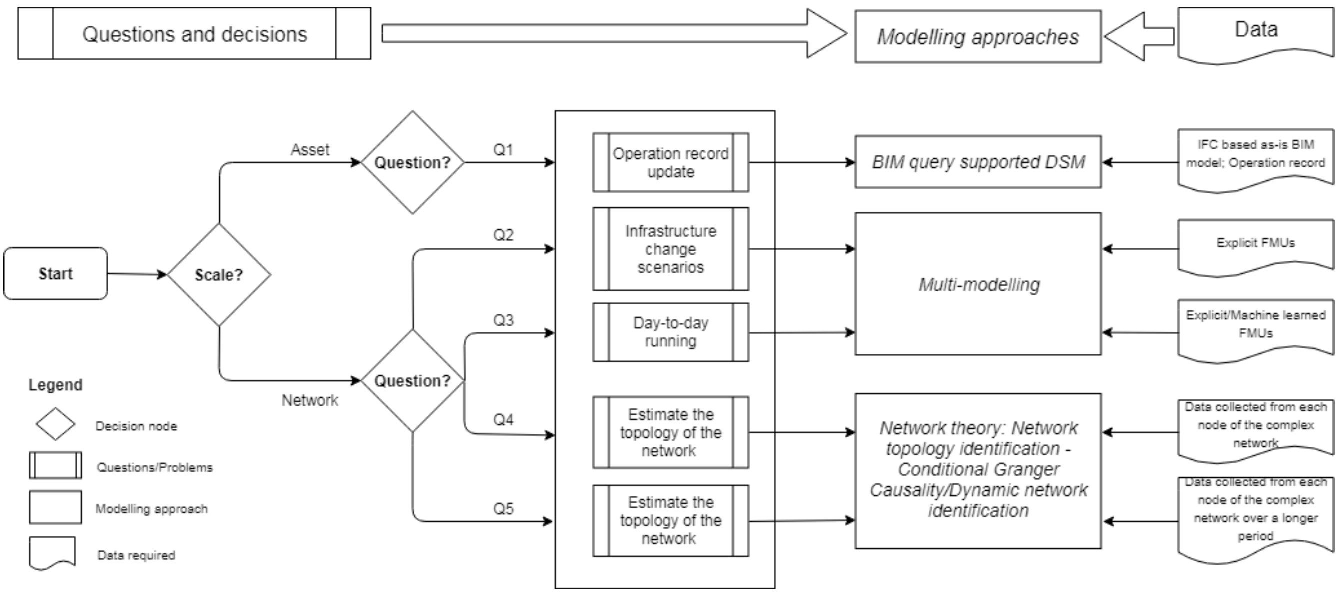


Figure 18. A decision tree for operations which shows the questions and decisions (from left to middle, as shown in Figure 3) and also data and modelling (from right to middle) associated with these systems analyses using the digital twin (an earlier version is discussed in Babovic et al., 2019)

To summarise the modelling approaches and sources of data discussed above, and how they relate to the ‘use case’ questions on *design and delivery* raised in section 2.1.3, Figure 19 shows a decision tree for operations. This expands upon Figure 4, to show the questions and decisions (from left to middle) connected with data sources and modelling techniques (from right to middle) that may be associated with systems analyses using the digital twin.

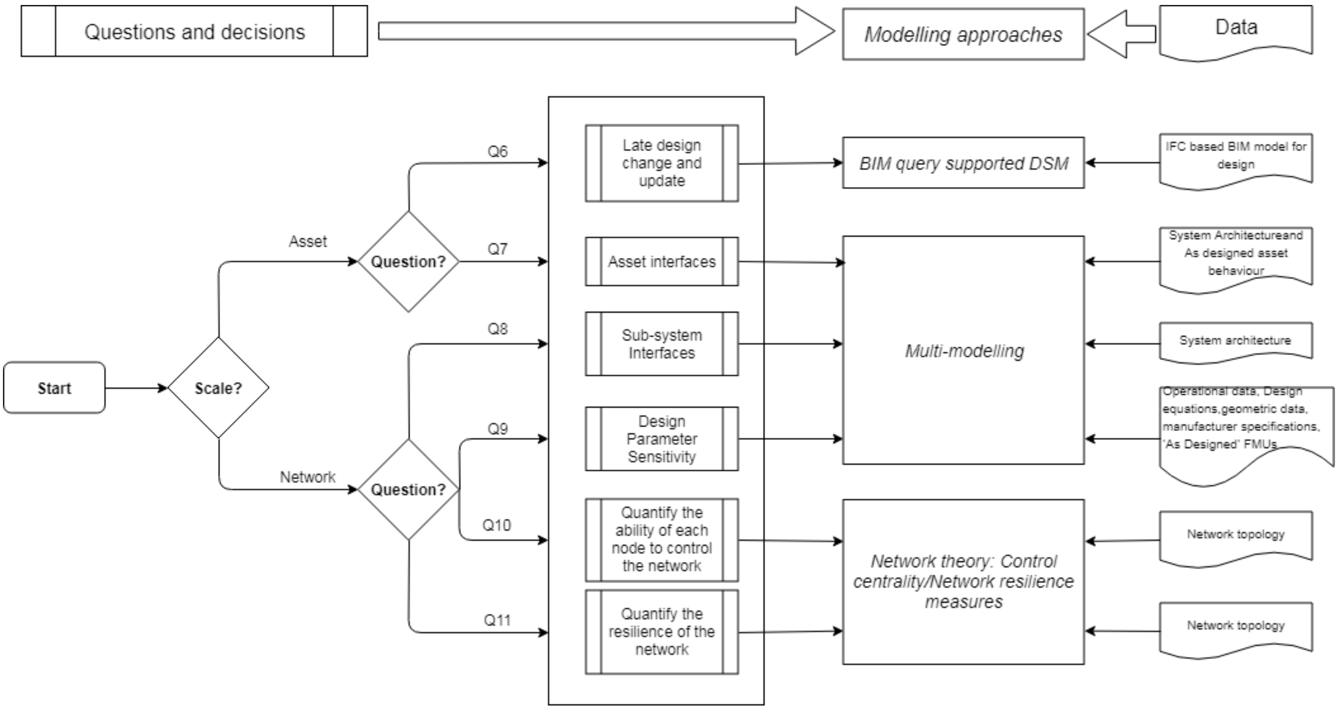


Figure 19. A decision tree for design/delivery which shows the questions and decisions (from left to middle, as shown in Figure 4) and also data and modelling (from right to middle) associated with these systems analyses using the digital twin (an earlier version is discussed in Babovic et al., 2019)

These decision trees for operation and design/delivery recognise the questions and decisions that arise for practitioners, and how, in relation to the modelling approaches considered in this research, these may be affected by the timeframe (delivery or operations), the level of analysis and ontology (scale – network/asset) and the shareability of data across organisations and data types that affect the modelling methods used with the digital twin. They set out a series of questions or use cases and then consider the data and modelling approaches that may be used to address them. The decision tree is thus intended as a representation of the choices, the questions and decisions and the modelling approaches and data that might be used to address these in the conditions that we considered in the research project. It is not intended to represent a generic framework.

4. Conclusions, Implications and Further Work

4.1: Overview

This research sought to articulate the extent to which a digital twin can be used to generate new insight into systems relationships and interdependencies. In addressing the objectives, the research develops knowledge on:

1. *Critical interdependencies:* Addressing objective 1, the work to identify critical interdependencies emerging in Tideway, both in the infrastructure system, and in the enabling production system is summarised in section 2.1 of this report. While we intended to quantify and rank these, we found that practitioners feel relatively confident in managing known interdependencies in both delivery

and operations. Our focus became gaining a qualitative understanding of interdependence in this context. To enhance the resilience of infrastructure, there is the potential to use the digital twin to identify interdependencies from near-real-time data on operating systems, as well as to understand projects as interventions into infrastructure systems. Our findings suggest that unknown interdependencies and emergence should be the focus of future research.

2. *Application of modelling approaches:* Addressing objective 2 was a significant focus of this work, through the substantial research focused on applying and extending modelling approaches, as summarised in sections 2.2, 2.3 and 2.4 of this report. Contributions of this work are to the emerging trajectory of research on the digital twin, and include advances in the Digital-DSM, network analysis and multi-modelling, and their application to infrastructure projects. Through this work we sought to identify critical interdependencies in time for decision makers on the project to make decisions by linking digital data. One issue that arises in this work concerns the types of data required for modelling, with near-real-time data being needed for some types of modelling, such as network analysis and multi-modelling.

3. *Use cases for different analysis approaches:* Addressing objective 3, the work to articulate, across different scales, the utility of and practical barriers to the use of different analysis approaches (e.g. network analysis, multi-modelling) in relation to practical problems and use cases faced in delivery (objective 3) is summarised in the discussion section (section 3). Here we expand on the motivating questions and use cases set out in Figures 3 and 4, to provide decision-trees in Figures 18 and 19 to represent the choices, questions and decisions and the modelling approaches and data in the conditions that we considered in the research project.

This research identifies challenges and opportunities for extending leading practice in order to analyse systems interdependencies through the use of a digital twin. The sharing of data across organisational boundaries, the development of common ontologies and the identification of modelling methods for different use cases through the delivery and operation of infrastructure are all ongoing challenges. On leading infrastructure projects, such as Tideway, there has been significant emergence of digital practices, as modelling and data use have become more sophisticated over the last 20 years. The research starts theoretically, highlighting the potential of a consistent basis for extending the modelling and enabling analytics through a disciplined approach.

4.2: Policy implications

This delivers a range of policy insights. There is a particular opportunity for the findings to inform work on the digital twin, in response to the ‘Data for the Public Good’ report (NIC, 2017), both at the Alan Turing Institute and through the Digital Framework Task Group (DFTG) Digital Twin hub, set up following the work on the ‘Gemini Principles’ (Bolton et al., 2018).

The findings support the CDBB mission to support the digitally enabled transformation of the full lifecycle of the built environment in order to increase productivity and improve economic and social outcomes in the UK and, where appropriate, internationally. Through the academic bridgehead, it ensures that the Digital Built Britain programme is cognisant of new and emerging research and technological developments that will impact upon the built environment in the years and decades to come. The main research topic is CDBB: “10. Exploring ways to leverage data and information to deliver a digital built Britain”. Additionally, the research addresses: “4. Exploring the implications of a digital built Britain for economic infrastructure” (the work with Tideway brings a specific empirical focus on water and sanitation, but the results will be more broadly applicable to built infrastructure). The main area of interest is that of “Complex integrated systems” in which we share CDBB’s interest in the insight, understanding, modelling and management of complex integrated systems so as to enable and manage interaction and cascade events, and in the implications of organisational interactions around these systems. Other areas of interest include “Digital, Construction, Manufacturing, and Data and Information”. These are well targeted towards CDBB’s mission and governmental policies such as the Industrial Sector deal. They explore ways in which to leverage data and information in order to deliver a digital built Britain and complex integrated systems. Many of the issues that need to be addressed in “Transforming Construction” are systemic, and this research deepens our ability to address infrastructure systems as being complex, and to see their production systems holistically.

4.3: Practical implications

For practitioners working in the operation or delivery of infrastructure, and interested in linking together assorted datasets and historical modelling activities in order to develop a consistent basis for a ‘digital twin’, the work has practical implications:

- It sets out scenarios in which network analyses, multi-modelling and BIM query may be useful in addressing particular questions in operations and in the design and delivery of infrastructure projects as interventions in infrastructure systems. It provides decision trees with which to simply understand these questions and ‘use cases’. Thus, for practitioners interested in the exploitation of data and analytics to enhance the natural and built environment, the findings suggest some questions and approaches with respect to using the digital twin to analyse systems interdependencies.
- By taking a theoretical approach, drawing on research disciplines, this work also suggests and provides steps towards a framework through which digital approaches can be implemented and potentially integrated.
- The work provides insight into the understanding, prioritisation and management of interdependencies at the asset, project and systems-of-systems level, where these may be geospatial, physical, cyber or logical. It also suggests the need for a particular emphasis on the importance of unknown interdependencies and emergence in order to understand the project as an intervention in infrastructure systems-of-systems.
- The work also has implications for how practitioners and researchers might work together in this emerging trajectory of work on the digital twin. Our experience suggests that this requires a high bandwidth of communication, accompanied by a high intensity of interactions and the co-creation of knowledge, across partners, where the industry is a co-developer of insights, rather than being a source of data, and the research team engage in order to understand the mechanisms of what is delivered and what may be available. We recommend an early data stocktake that considers ownership, levels of abstraction, potential users of models, as well as a regular pattern of interaction, particularly in the early stages of work.

The Tideway team are intending to develop a virtual operating system for the tunnel system. The intention of the system is to support providing training for operators and controllers of the system, aid scenario planning and decision making for system commissioning and enable assurance of system performance against expected design conditions. Using a combination of existing available models, their challenge is to understand what level of integration is available and appropriate in order to meet the intent of this digital twin. Further exploration of the multi-model work may be particularly applicable in the Tideway case.

4.3: Future research

As a relatively small empirical study (conducted between October 2018 and July 2019), this work is an early step in the emerging trajectory of research on digital twin engineering. It has some limitations and the work both suggests some areas for further research and provides a basis for such future work.

4.3.1: Limitations of the existing study

The work has examined three generic modelling techniques in detail — BIM query, network analysis, and multi-modelling — with each being considered primarily at one level of analysis, as shown in Figure 2. There is more work to be done to consider the potential of these approaches across scales, as well as their combination with other modelling approaches, and to address different use cases and scenarios.

4.3.2: Areas for further research

This work has demonstrated how a digital twin can be used to generate new insights into systems relationships and interdependencies on infrastructure projects. In order to deliver and operate infrastructure holistically, it is necessary to integrate knowledge from various information sources. However, even in leading practice, this remains a major challenge due to limited interoperability and interaction between islands of information. Future work should aim to address this through the application of a standardised data model based on linked data principles and knowledge engineering techniques. This would serve as a foundation for further research towards enhanced interoperability and integration between the diverse information sources and model formats on infrastructure projects.

There are hence a number of directions for further research in this area. Two examples are the extension of the set of modelling to include linked data, and the consideration of digital twins across scales.

Linked data: Here, in addition to the developed approaches, future work should consider the potential of linked data-enabled modelling approaches to enrich analyses for systems interdependencies. Emerging research questions for further research include:

- How can linked data and knowledge modelling techniques be used to infer relational interdependencies on infrastructure projects?
- How can a linked data-enabled system be used to identify the constituent ‘parts’ or ‘systems’ for multi-modelling the operation of an infrastructure system?
- How can linked data-enabled systems be used to test the impact of alternative ‘parts’ and ‘systems’ on the overall performance of an infrastructure system?
- How can linked data be used to automatically incorporate past weather data and future forecasts for simulating the operation of an infrastructure system?

The city-scale digital twin: The pursuit of a city-scale digital twin is necessarily heterogeneous in its use of file formats, driving the development of a platform that is capable of coping with significant variations in system protocols. The city-scale digital twin should support further exploration of uncertainty. Many infrastructure projects are not built as designed and many are not operated as intended. The implication of this may be as simple as from which electrical panel a pump is fed or a more subtle variation in how demand is managed at peak times. The impact of these potential uncertainties in structure and dynamics can be better explored when the system under investigation is embedded within a wider spatially and logically defined set of systems. The Tideway project provided the opportunity to explore how a large consortium of actors collaborate on a single, large project. The challenge for cities is that although such single, large interventions are relatively rare, it is not uncommon for the same scale of activity to be occurring across several projects. A city-scale digital twin offers the opportunity to better understand the relationships and interdependencies across such multiple projects.

While this project has involved a consideration of sensitivity analyses and visualisation, in relation to the use of systems engineering approaches, BIM query, network analyses and multi-modelling, there is also an opportunity to significantly extend research in these areas in future work.

This work has implications for the work of researchers associated with the Data and Analytics Facility for National Infrastructure (DAFNI) — an £8million investment in new data, computation and visualisation facilities, being delivered by the Science and Technologies Facility Council (STFC) at the Rutherford Appleton Laboratory (RAL) in Harwell. Whereas this is a relatively small-scale project, researchers can use this technology platform to develop tools and services for data capture and assimilation, model simulation of national infrastructure systems, machine learning, and visual analytics that will be continuously refined and updated. We envisage that the new data analytics, modelling and simulation tools that we are exploring in this study will be made available in DAFNI in the future, allowing stakeholders to explore scenarios and simulate, evaluate, optimise and visualise design in order to inform major investment commitments in real time.

It may also inform the Urban Observatory at Newcastle University (www.urbanobservatory.ac.uk/) and the Urban Flows Observatory at the University of Sheffield, which aims to create a better understanding of energy, water and resource usage, supporting its air, energy and materials research areas. To support this research at a city level requires a set of tools with which to manage the information collected and create ways in which to display, analyse and interact with the data so as to develop insights into and further understanding of the interrelationships and interdependencies.

4.3.3: Contribution to further work

The contribution to further work of this nine month research project on analysing systems interdependencies using a digital twin is to help researchers, practitioners and policy makers to both (1) see practical steps to using and connecting heterogeneous data to answer questions about the operation and design/delivery of infrastructure systems, and (2) understand projects as interventions into infrastructure systems-of-systems.

Acknowledgements

We are grateful for the input of Tideway, particularly Dr Sian Thomas, Dr Patrick Owen and Peter Vale, who commented on this report. We also acknowledge the input of Dr Giuliano Punzo from the University of Sheffield for input in the delivery of the project, and of Mark Enzer and Professor Mark Girolami for reviewing near-final drafts. The authors retain responsibility for any errors and omissions.

References

Allen, F. & Babus, A. (2009) Networks in finance. In Kleindorfer, P.R., Wind, Y. & Gunther R.E. (eds) The network challenge: strategy, profit, and risk in an interlinked world, Prentice Hall. 367379.

Babovic, F., Chen, L., Gamble, C., Genes, C., Fitzgerald, J., Pierce, K., Punzo, G. & Whyte, J. (2019) A decision-tree for modelling approaches to use a digital twin in analysing systems interdependencies in infrastructure projects; Annual Systems Engineering Conference, November 19–20, Leeds, UK.

Barnett, L., Barrett, A.B. & Seth, A. K. (2014) Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables, Physical Review Letters, 103, 238701-1-4.

Barnett, L. & Seth, A. K. (2014) The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. Journal of Neuroscience Methods, 223, 50–68.

Berkowitz, B. & Ewing, R. P. (1998, Jan 01) Percolation theory and network modelling applications in soil physics. Surveys in Geophysics, 19(1), 23–72.
Bolton, A., Butler, L., Dabson, I., Enzer, M., Evans, M., Fenemore, T., Harradence, F., Keaney, E., Kemp, A., Luck, A., Pawsey, N., Saville, S., Schooling, J., Sharp, M., Smith, T., Tennison, J., Whyte, J., Wilson, A., & Makri, C. (2018) Gemini Principles. Report (CDBB_REP_006) https://doi.org/10.17863/CAM.32260 [accessed 9 July 2019]
Borrmann, A., König, M., Koch, C. & Beetz, J. (2015) Building Information Modeling: Technology Foundations and Industry Practice. Springer, Cham.
Borrmann, A. & Rank, E. (2009) Topological analysis of 3D building models using a spatial query language. Advanced Engineering Informatics, 23(4), 370–385.
Borrmann, A., Van Treeck, C. & Rank, E. (2006) Towards a 3D spatial query language for building information models. In: Proceedings of the 11th International Conference on Computing in Civil and Building Engineering (ICCCBE-XI), Montreal, Canada.
Browning, T. R. (2001) Applying the design structure matrix to system decomposition and integration problems: a review and new directions. *IEEE Transactions on Engineering Management*, *48*(3), 292–306.
BuildingSMART (2013) Industry Foundation Classes: IFC4 Official Release. Available Online: http://www.buildingsmart-tech.org/ifc/IFC4/final/html/link/ifcrelationship.htm [accessed 9 July 2019]
Buldyrev, S.V., Parshani, R., Paul, G., Stanley, H.E. & Havlin, S. (2010) Catastrophic cascade of failures in interdependent networks. Nature, 464(7291), 1025.
Carrington, P. J., Scott, J. & Wasserman, S. (2005) Models and methods in social network analysis (Vol. 28). Cambridge University Press, UK.
Chen, L. & Whyte, J. (2020) Analysing Interdependencies of Complex Engineering Systems Using a Digital Twin-Driven Design Structure Matrix, ASCE Construction Research Congress 2020, 8–10 March, Tempe, Arizona, USA.
Clementini, E. & Felice, P. D. (1995) A model for representing topological relationships between complex geometric features in spatial databases. Information Sciences, 90(1–4), 121–136.
Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. (2000) Resilience of the internet to random breakdowns. Physical Review Letters, 85(21), 4626.
Colbourn, C. J. (1987) The combinatorics of network reliability, international series of monographs on computer science (Vol. 198) (No. 7) Oxford University Press, New York, USA.
Crawford, D., Hon, A. & Thomas, S. (2017) Tideway: The way to plan, procure, design and construct a deep CSO control tunnel under London UK, Proceedings of the Water Environment Federation Technical Exhibition and Conference (WEFTEC), 5349–5355.
Daum, S. & Borrmann, A. (2014) Processing of topological BIM queries using boundary representation based methods. Advanced Engineering Informatics, 28(4), 272–286.
Delany, S. (2019) UNICLASS Classification, Available Online: https://toolkit.thenbs.com/articles/classification. [accessed 9 July 2019].
Ding, M., Chen, Y. & Bressler, S. L. (2006) Granger causality: basic theory and application to neuroscience. Handbook of Time Series Analysis: Recent Theoretical Developments and Applications, 437–460.
Dunn, S., Fu, G., Wilkinson, S. & Dawson, R. (2013) Network theory for infrastructure systems modelling. Proceedings of the Institution of Civil Engineers-Engineering Sustainability, 166(5), 281–292.
Egenhofer, M. J. & Franzosa, R. D. (1991) Point-set topological spatial relations. International Journal of Geographical Information Systems, 5(2), 161–174.
Fitzgerald, J., Gamble, C., Payne, R. & Lam, B. (2017) Exploring the Cyber-Physical Design Space. INCOSE International Symposium, 27: 371–385.
Geogaki, S., Baur, R., Desing, B. & Crawford, D. (2017) London's CSO control improvement programme – Air management challenges, Proceedings of the Water Environment Federation Technical Exhibition and Conference (WEFTEC), 3215–3236.
Goldbeck, N., Angeloudis, P. & Ochieng, W. Y. (2019) Resilience assessment for interdependent urban infrastructure systems using dynamic network flow models. Reliability Engineering and System Safety, 188, 62–79.
Gopsill, J. A., Snider, C., McMahon, C. & Hicks, B. (2016) Automatic generation of design structure matrices through the evolution of product models. Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AI EDAM), 30(4), 424–445.
Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society, 424–438.
Hon, A., Treacher, M., Dolan, D. & Crawford, D. (2017) If – then – else. A balance between protecting the assets through limited inflow and maximising CSO capture, Proceedings of the Water Environment Federation Technical Exhibition and Conference (WEFTEC), 2679–2708.

Kenett, D. Y., Gao, J., Huang, X., Shao, S., Vodenska, I., Buldyrev, S. V., Paul, G., Stanley, H. E. & Havlin, S. (2014) Network of interdependent networks: overview of theory and applications. In Networks of Networks: The Last Frontier of Complexity (pp. 3–36). Springer.
Li, D., Zhang, Q., Zio, E., Havlin, S. & Kang, R. (2015) Network reliability analysis based on percolation theory. Reliability Engineering & System Safety, 142, 556–562.
Marinazzo, D., Pellicoro, M. & Stramaglia, S. (2008) Kernel-Granger causality and the analysis of dynamical networks. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 77(5), 1–9.
Mazairac, W. & Beetz, J. (2013) BIMQL – An open query language for building information models. Advanced Engineering Informatics, 27(4), 444–456.
McQuarrie, A. D. R. & Tsai, C. L. (1998) Regression and Time Series Model Selection. World Scientific Publishing, Singapore.
NIC (2017) Data for the public good, National Infrastructure Commission, London, UK.
Nepal, M. P., Staub-French, S., Pottinger, R. & Webster, A. (2012) Querying a building information model for construction-specific spatial information. Advanced Engineering Informatics, 26(4), 904–923.
Otte, E. & Rousseau, R. (2002) Social network analysis: a powerful strategy, also for the information sciences. Journal of Information Science, 28(6), 441–453.
Parrott, A. & Warshaw, L. (2017) Industry 4.0 and the digital twin: Manufacturing meets its match, Report. https://www2.deloitte.com/insights/us/en/focus/industry-4-0/digital-twin-technology-smart-factory.html [accessed 2 July 2019]
Parshani, R., Buldyrev, S. V. & Havlin, S. (2010) Interdependent networks: Reducing the coupling strength leads to a change from a first to second order percolation transition. Physical Review Letters, 105(4), 48701.
Pimmler, T. U. & Eppinger, S. D. (1994) Integration analysis of product decompositions. Proc. ASME 6th Int. Conf. on Design Theory and Methodology, Minneapolis, USA.
Plant, J., Crawford, D. & Nielson, K. (2017) Design standards for in-shaft vertical deaeration and energy dissipation for deep storage-conveyance tunnels, Proceedings of the Water Environment Federation Technical Exhibition and Conference (WEFTEC), 1453–1481.
Rinaldi, S. M., Peerenboom, J. P. & Kelly, T. K. (2001) Identifying, understanding, and analyzing critical infrastructure interdependencies. IEEE Control Systems Magazine, 21(6), 11–25.
Saleh, M., Esa, Y., Onuorah, N. & Mohamed, A. A. (2017) Optimal microgrids placement in electric distribution systems using complex network framework. In 6th International Conference on Renewable Energy Research and Applications (pp. 1036–1040).
Saoud, L. A., Omran, J., Hassan, B., Vilutienė, T. & Kiaulakis, A. (2017) A method to predict change propagation within building information model. Journal of Civil Engineering and Management, 23(6), 836–846.
Senthilkumar, V. & Varghese, K. (2009) Structured methodology to formulate drawing dependency structure matrix for construction design. Architectural Engineering and Design Management, 5(4), 225–248.
Smith, D. A. & White, D. R. (1992) Structure and dynamics of the global economy: network analysis of international trade 1965-1980. Social Forces, 70(4), 857–893.
Sun, W., Bocchini, P. & Davison, B. D. (2018) Resilience metrics and measurement methods for transportation infrastructure: the state of the art. Sustainable and Resilient Infrastructure, 1–32.
Van Den Hof, P. M. J., Dankers, A., Heuberger, P.S.C. & Bombois, X. (2013) Identification of dynamic models in complex networks with prediction error methods - Basic methods for consistent module estimates, Automatica. 49(10), 2994–3006.
Whyte, J., Fitzgerald, J., Mayfield, M., Coca, D., Pierce, K. & Shah, N. (2019) Projects as interventions in infrastructure systems-of-systems, International Committee on Systems Engineering (INCOSE) International Symposium 2019, July 20–25, Orlando, USA.
Wiener, N. (1956) The theory of prediction. Modern Mathematics for Engineers. New York: McGraw-Hill.