A spatiotemporal database framework for infrastructure systems analytics and modelling

Robson C.[*1], Barr S.[†1], Pregnolato M.[‡1], Ji Q. [§1]
[1]School of Engineering, Newcastle University, Newcastle (UK)

January 2018

Summary
The size, scale, variety and complexity of spatial datasets required for the analysis of critical infrastructure is ever increasing. Decision makers require knowledge at fine temporal and spatial scales, such as at the household scale, resulting in demand for significantly larger datasets then previously used. Data storage and management mechanisms for such datasets require novel solutions. A heterogeneous federated database solution is presented where buildings are considered as the primary spatial entity, and where multiple optimal SQL/NoSQL databases are employed to store the wide range of datasets required in the analysis of infrastructure systems. A pilot study is presented for the city of Newcastle-Upon-Tyne where local building level electricity distribution networks are coupled with buildings for which household characteristics are associated.

KEYWORDS: spatiotemporal database, infrastructure systems, heterogeneous data, multiscale analytics

1. Introduction and background

The improved availability and quality of fine spatial scale infrastructure asset data (Barr S.L. *et al.*, 2016), along with the developments in data analytics and modelling (Ouyang, 2014), offer the greater ability to understand and develop new insights of infrastructure systems and networks at an unprecedented spatial and temporal resolution. In particular, there is an increasing prevalence of spatial data for critical infrastructure systems, such as data on demand modelling (Evans *et al.*, 2017), which is coupled with a growing interest in ensuring the security of the services such infrastructure systems can provide (HM Treasury, 2010; HM Treasury, 2016), such as transport, communications and energy. The potential opportunities within the infrastructure field given spatial-temporal data at very fine scales are significant, and to realise this, data management solutions which extend upon file based systems are required to handle the unique array of varied data types and scales which are indispensable to future work in this area.

The spatio-temporal framework, NISMOD-DB++, presented in this paper extends an existing framework (NISMOD-DB) which supported spatiotemporal infrastructure modelling at the regional to national scale (Barr S.L. *et al.*, 2016). NISMOD-DB++ has been designed to allow analysis and modelling at the intra-urban level of infrastructure assets and buildings through employing a federated database architecture (a set of cooperating databases) allowing multiple disparate databases to be seamlessly and efficiently linked together (Sheth and Larson, 1990; Dharmasiri and Goonetillake, 2013). This work is part of the Infrastructure Transitions Research Consortium MISTRAL project (Multi-Scale Infrastructure Systems Analytics) and supports the development of fine scale spatial

[*]craig.robson1@ncl.ac.uk
[†] stuart.barr@ncl.ac.uk
[‡] maria.pregnolato@ncl.ac.uk
[§] q.ji2@ncl.ac.uk

infrastructure sector models for transport, energy, digital communications and waste sectors, which in combination and employed together form an interdependent long-term infrastructure planning model platform.

2. Database framework

NSIMOD-DB++ (Figure 1) uses a federated database approach where a single master PostgreSQL database is used as a connector to other databases, be they SQL or NoSQL solutions (Dharmasiri and Goonetillake, 2013). Such a federated design means that the most appropriate/optimal database storage can be utilised for different data sets, allowing optimal ingestion, storage and retrieval (querying) of them via a single common meta-database. Moreover, by building meta-relationships between data-sets within different databases, it becomes possible to perform queries over multiple disparate databases to retrieve data and information. This implementation allows for SQL (relational) solutions, such as for structured/tabular data to be used where most appropriate, while unstructured data, such as spatially fine-scale networks can be stored within a graph database, a native and very efficient storage solution for data where the relationships within it are important (Angles and Gutierrez, 2008).

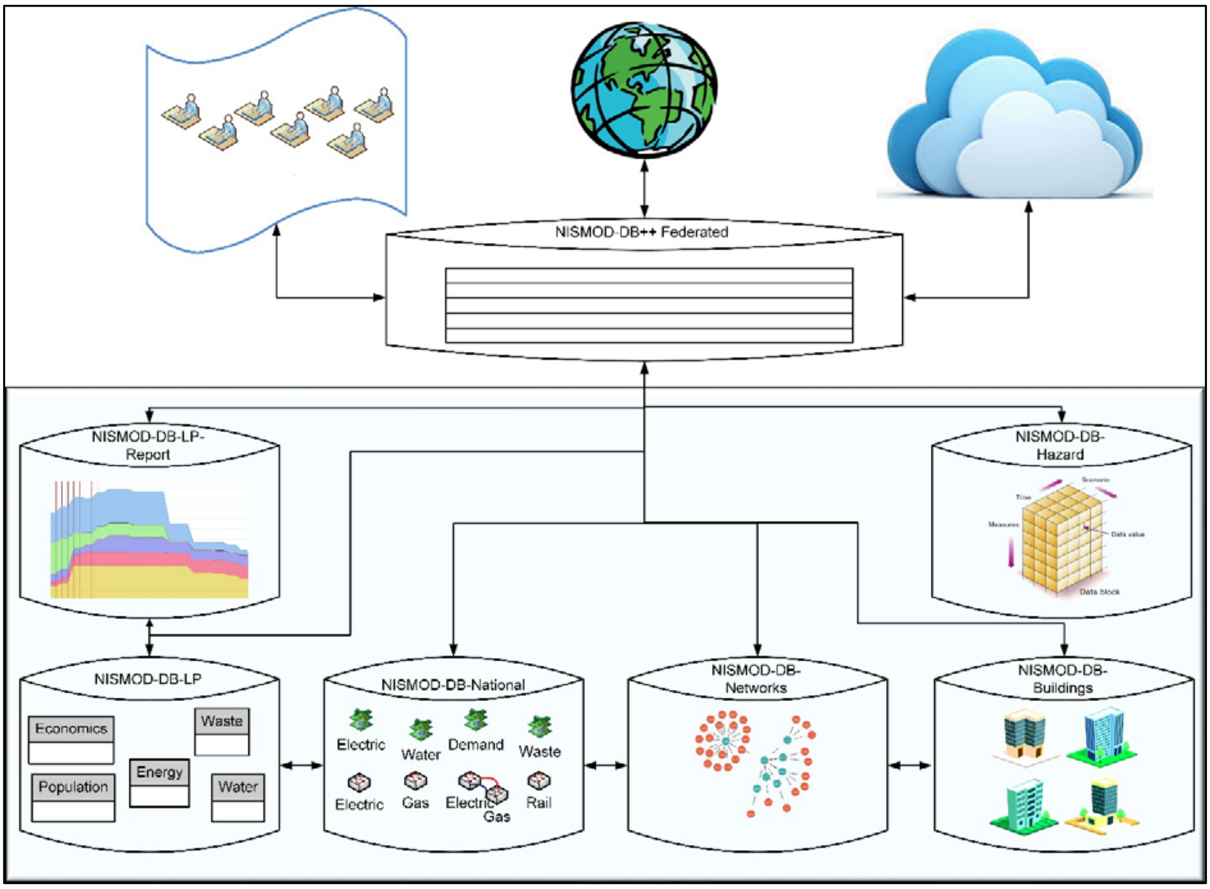

Figure 1: Database framework for NISMOD-DB++.

3. Pilot study

The application of the database framework for storing, handling and serving spatio-temporal datasets has been demonstrated through a case study using the city of Newcastle-Upon-Tyne. A range of datasets have been used to show the effectiveness of the framework from a number of perspectives, including the ability to handle structured spatial data such as buildings and household structure, as well as unstructured spatial data such as infrastructure networks. Ordnance Survey MasterMap data has been used as a primary dataset providing household level mapping data, including building footprints (Figure 2). Using Ordnance Survey AddressBase Premium, the activity of each building has been identified (residential, mixed-use, non-residential), along with the building form of residential properties (terraced, flat, semi-detached, detached) (Figure 2). Using outputs from a household microsimulation model, household characteristics are then assigned to each building, providing a detailed profile of each residential address.

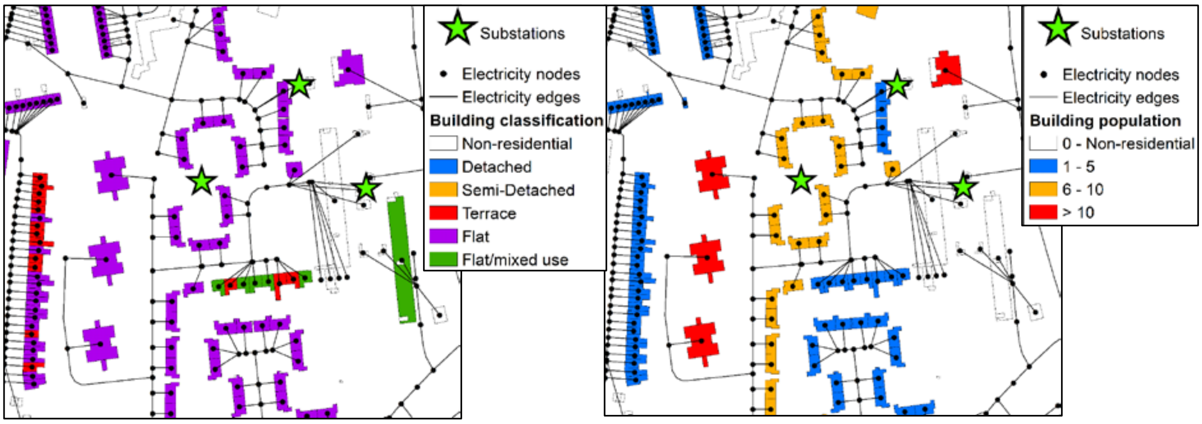

Figure 2: Building classification results (left) and resident population per building (right) with the electricity distribution network.

To enable a full evaluation of the ability of the federated database approach, an electricity distribution network is also used (Figure 3), generated through a heuristic algorithm using the MasterMap buildings, the road network and the existing known electricity substations (as identified using the Ordnance Survey Points of Interest dataset). Mapping the supply from 11kv substations to each building, when combined with the known National Grid transmission network, gives a complex network for the whole of Newcastle-Upon-Tyne of 191,595 nodes and 190,989 edges, with 640 sub-networks. This is then stored within a Neo4j graph database designed for the storage of spatial infrastructure networks (NISMOD-DB-Networks) within the NISMOD-DB++ federated database architecture.

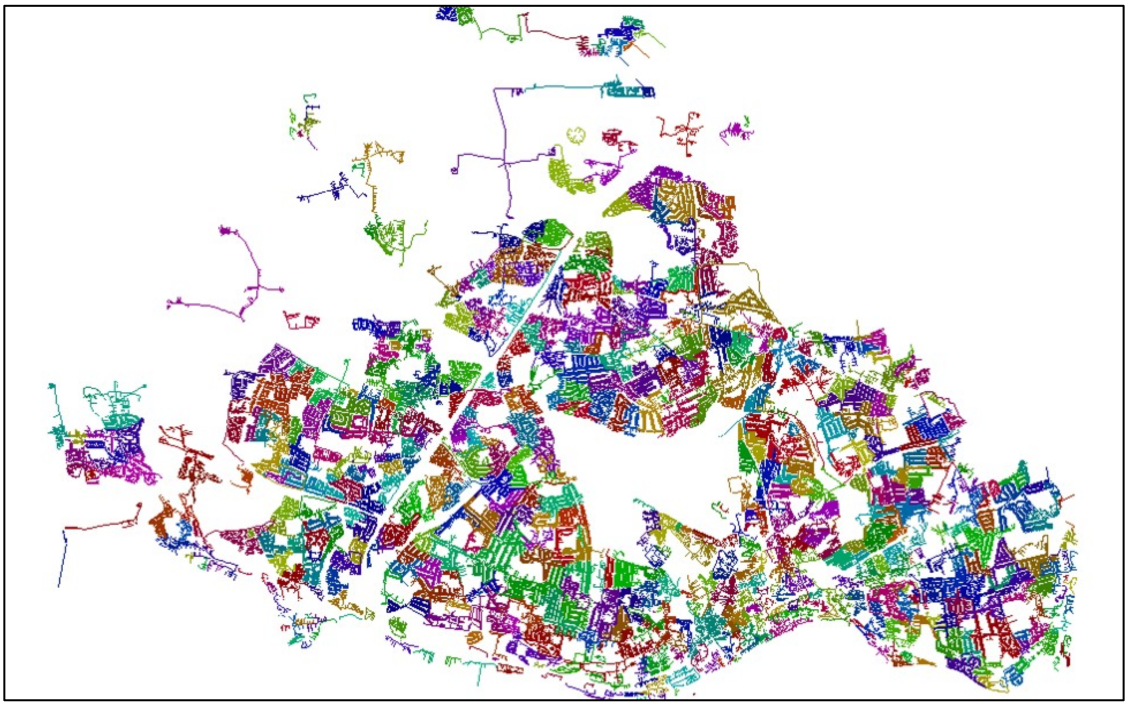

Figure 3: Low-voltage electricity distribution network for Newcastle-Upon-Tyne (Contains Ordnance Survey data, Crown copyright 2017).

By employing a federated database approach and foreign data wrappers in PostgreSQL, data across multiple databases can be queried as if part of the same database. This allows datasets of different types to be seamlessly integrated and employed for complex queries, increasing the potential for new insights to be learned and improving performance. This is demonstrated (Figure 4) through combining the household level data stored within a PostgreSQL database with the electricity network data stored within a graph database, though connected through a foreign data wrapper. Using the function 'cypher', this allows a query (red text, Figure 4) to be run over the graph database that is optimised for the analysis of networks from within the relational database. By then joining the output of this query, in this case a list of buildings and their supplying substation, with household details from the relational database, insights into the residential properties reliant on different sub-networks within he electricity network can be quickly ascertained (Figure 5).

```
SELECT net_id, sum("Dwellers") as "No of people", count("Dwellers") as "No of households"
FROM nismod_household.households as households
JOIN (SELECT toid, net_id FROM cypher('MATCH (n:ElectricityDistNode) WHERE n.net_id IN [272,206,102,29]
RETURN n.toid as toid, n.net_id as net_id'), json_to_record(cypher) as x(toid varchar, net_id varchar)) as
dwellings
ON households."DwellingID" = dwellings."toid"
GROUP BY net_id;
```

Figure 4: Example query using data from a PostgreSQL database and a graph (Neo4j) database.

net_id	No of people	No of households
29	346	194
206	372	137
272	153	93

Figure 5: Outputs from the query (Figure 4) showing the number of people and related households dependent on particular electricity distribution sub-networks in Newcastle-Upon-Tyne.

4. Conclusion

The presented spatio-temporal database framework enables the storage, management and optimised querying of varied spatial datasets more easily and efficiently through the ability to store data in a database type/structure which is more optimised to the data type. The designed solution employs a spatially enabled relational database through PostgreSQL (and PostGIS), coupled with a graph database (Neo4j), allowing the developed foreign database wrappers within PostgreSQL to couple such databases seamlessly. The effectiveness of the solution is demonstrated with a pilot study whereby the ability to combine unique and varied datasets is shown. Future work will see the framework employed for a wider set of datasets highlighting the ubiquitous nature of the solution as well as on data for the whole UK, massively increasing the scale of the data being handled.

5. Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council (ESPRC), for the MISTRAL (Multi-scale Infrastructure Systems Analytics) programme grant awarded to the Infrastructure Transitions Research Consortium (ITRC) (Grant EP/N017064/1).

6. Biography

Craig Robson is an ITRC-MISTRAL Research Associate in the School of Engineering at Newcastle

University. His research interest include the robustness and resilience of critical spatial infrastructure networks to hazards, and data management and visualisation systems for large and complex datasets.
Stuart Barr is a Professor of Geospatial Systems Engineering in the School of Engineering at Newcastle University.
Maria Pregnolato is an ITRC-MISTRAL Research Associate and early career EPSRC Research Fellow in the School of Engineering at Newcastle University. Her research interests include future housing and infrastructure development, network modelling and urban resilience.
Qingyuan Ji is a PhD student in the School of Engineering at Newcastle University, associated to the ITRC-MISTRAL project, who is working on heuristic algorithms for the generation of fine-scale networks.

References

Angles, R. and Gutierrez, C. (2008) 'Survey of graph database models', *ACM Computing Surveys*, 40, pp. 1-39.

Barr S.L., Alderson D., Ives M.C. and Robson C (2016) 'Database, simulation modelling and visualisation for national infrastructure', in Hall, J.W., Tran, M., Hickford, A.J. and Nicholls, R.J. (eds.) *The Future of National Infrastructure: A System-of-Systems Approach*. UK: Cambridge University Press.

Dharmasiri, H.M.L. and Goonetillake, M.D.J.S. (2013) *2013 International Conference on Advances in ICT for Emerging Regions (ICTer)*. 11-15 Dec. 2013.

Evans, S., Liddiard, R. and Steadman, P. (2017) '3DStock: A new kind of three-dimensional model of the building stock of England and Wales, for use in energy analysis', *Environment and Planning B: Urban Analytics and City Science*, 44(2), pp. 227-255.

HM Treasury (2010) *Strategy for National Infrastructure*. UK: HM Treasury,.

HM Treasury (2016) *National Infrastructure Delivery Plan 2016-2021*.

Ouyang, M. (2014) 'Review on modeling and simulation of interdependent critical infrastructure systems', *Reliability Engineering & System Safety*, 121(0), pp. 43-60.

Sheth, A.P. and Larson, J.A. (1990) 'Federated database systems for managing distributed, heterogeneous, and autonomous databases', *ACM Comput. Surv.*, 22(3), pp. 183-236.