Designing DAFNI : a national facility for modelling infrastructure

Brian Matthews, Sam Chorlton, Peter Oliver, Ron Fowler, and Erica Yang

STFC Scientific Computing Department

Brian.Matthews@stfc.ac.uk

Abstract. DAFNI is a major UK national facility to advance infrastructure system research. DAFNI will host national infrastructure datasets and provide a complex hybrid-cloud platform for modelling, simulation and visualisation. We discuss some of the challenges and aims of the DAFNI system, outline its architecture and technical challenges, and summarise its status.

1 Introduction

The infrastructure systems of a country or region, including energy supplies, water systems, transport routes, digital networks, land use, and the built environment, are subject to environmental (e.g. climate, geology, hydrology), social and economic pressures. Researchers in a variety of disciplines, including environmental sciences, geography, civil engineering, urban planning and economics use computer modelling and analysis of infrastructure to explain and predict the effects of changes, whilst policy makers use the outputs of such models to make planning decisions. Infrastructure systems are becoming ever more complex, and models are becoming more detailed, combining data from different infrastructures and disciplines, and at different scales, from a country or a region down to an individual locality or building. Thus, there is a need for advanced high-performance and high-throughput computing, large-scale data infrastructure to manage and combine data, together with cloud systems for on-demand remote access.

The Data Analytics Facility for National Infrastructure (DAFNI)[1] is a major national facility under development in the UK to provide world-leading capability to advance infrastructure system research. It involves a consortium of 14 universities led by the University of Oxford, and is being developed and hosted by the Science and Technology Facilities Council's (STFC) Scientific Computing Department, as part of the UK Collaboratorium for Research on Infrastructure and Cities (UK-CRIC). It will provide a scalable hybrid cloud platform supporting storage and querying of heterogeneous national infrastructure datasets in a multi-modal architecture, and will support the execution, creation and visualisation of complex modelling applications to enable significant new advances in infrastructure research, and to improve the readiness of research tools and methods for real-world challenges at scale, nationally and internationally. This platform will improve the quality and opportunities for National Infrastructure Systems research whilst reducing the complexity of using data and models for end users.

2 Motivations and objectives

The Organisation for Economic Co-operation and Development estimates that globally US$53 trillion of infrastructure investment will be needed by 2030 (OECD 2012). In line with this, the UK's National Infrastructure Plan has set aside over £460 billion of investment for the next decade. However, the impact this investment is hard to predict as projections are underpinned by the quality of the analytics used to inform decision making. Advanced by big data analytics, simulation, modelling

[1] https://www.dafni.ac.uk/

and visualisation are now providing the possibility to improve this situation, but there are a number of challenges to be overcome, including:

Distributed teams: analytics are currently undertaken as an isolated activity at disparate institutions with minimal instances of coalescing and collaboration of outputs. However, infrastructure networks and their interactions are inherently complex and heterogeneous, with interactions with people and the environment. Handling this complexity is beyond the capacity of any one team.

Data Heterogeneity: The variety and variability in data has become a limiter for the modelling community and presents a constant hurdle with respect to model collaboration, interoperability and accessibility with extensive subject matter expertise required to exploit each model. Existing data arrangements such as EDINA[2] provide shared academic access to a number of the data sources used in modelling activities, but these represent a subset of the total data landscape.

Maintaining traceability: there is a need to ensure results are reliable and repeatable. It will therefore be essential to store versioned copies of the datasets to support this. This will cause an exponential growth in scale as the range of supported datasets increases and the platform ages.

Data security: licensing of data and models in this field is complex. The difficultly in ensuring that security is maintained presents a barrier to data sharing, and safeguarding the integrity of the data for researchers and data providers represents a key challenge the development of a common platform.

Model granularity: the increase in data availability and resolution has enabled new modelling applications with increasing granularity of modelling applications, with a corresponding increased demand for computational resources. The resource availability limits the ability for modelling activities to understand impacts of simulations at a national scale whilst maintaining fine grain resolution.

Consequently, the shared DAFNI platform is being developed to provide a dedicated compute resource specifically for the National Infrastructure modelling community. It will improve the quality and opportunities for research; and reduce the complexity of all aspects related to conducting the research including data access and processing, model execution, security and visualisation. It will provide specific optimisations to support tasks such as sensitivity analysis and parameter optimisation required by the community. The combination of these facets with a functional platform that addresses the data, licencing and scalability challenges delivers a platform enabling research in areas such as:

- modelling of energy, transport, digital and water networks, for system planning and optimisation;
- real-time data assimilation from sensors and the Internet of Things (IoT) to enable more efficient and reliable operation of infrastructure networks;
- modelling of changing patterns of demand for infrastructure services to enable investment planning;
- modelling of extreme events and their impact upon infrastructure networks to target vulnerabilities and enhance network resilience.

3 DAFNI Architecture and Capabilities

Although the problem space that DAFNI is addressing is broad and complex, an initial requirement gathering exercise identified core capabilities, as illustrated in Figure 1, and briefly described below.

National Infrastructure Database (NID): a centrally managed access point to national infrastructure and other datasets required to support infrastructure research. This includes: a centrally managed datastore; an Extract Transform and Load (ETL) framework to maintain data currency and interoperability; a data catalogue; and a data access and publication service.

[2] The University of Edinburgh Centre for Digital Expertise (https://edina.ac.uk/)

National Infrastructure Modelling Service (NIMS): support for users to improve performance of existing models, reduce the complexity of creating models and facilitate the creation of complex system-of-systems models. This will include: workflow framework and creator; a workflow engine; a model catalogue and a data transformation library.

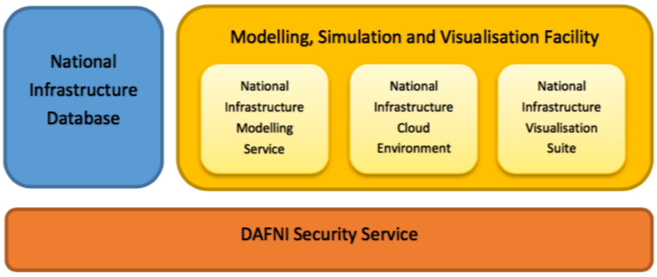

Figure. 1. DAFNI Core Capabilities. Cloud, modelling and visualisation services are grouped into a Modelling, Simulation and Visualisation facility

National Infrastructure Cloud Environment (NICE): a high performance, flexible and scalable hybrid cloud environment with: a cloud services user interface and command line interface; a number of PaaS offerings (e.g. data science notebooks and databases); and a centralised resource pool in support of PaaS and workflow manager services.

National Infrastructure Visualisation Suite (NIVS): a set of high-resolution visualisation tools to facilitate understanding of data, models, outputs and translation of findings to decision makers. This will include: traditional visualisation as a service (e.g. graph and tabular representations); and advanced visualisation as a service (e.g. virtual/augmented reality)

DAFNI Security Service (DSS): tools and processes to manage security aspects of the platform, which allows users to seamlessly access and use services they have rights to without being hindered by the platform, while at the same time maintaining security and integrity of data. Services will include authentication, authorisation, monitoring, and accounts management.

These components will be implemented in a micro-services architecture. This allows the capabilities within DAFNI to be independent with an extensible and flexible delivery of the platform in line with the evolving nature of the National Infrastructure modelling landscape. Following a structured design approach, a hierarchical overview of the platform has been derived leveraging the capabilities and functions outlined as part of the core capabilities analysis. We highlight three key features within the design that present a research challenge for the platform. They are the Central Datastore (CD), Query Manager (QM) and the Model Workflow Engine (MWE).

Central Datastore: The CD is the focal point for the storage of National Infrastructure datasets. The architecture aims to use the multi-modal approach utilising relational, graph and non-relational database technologies to facilitate the data heterogeneity. This is complemented by an object store for access to datasets and a MetadataDB to maintain a record and audit trail of the diverse datasets. This combination of database technologies allows the data to be stored in the manner most suitable and often therefore most performant. This reduces the complexity of ingesting the datasets, and allows models interacting with existing databases to remain predominately the same. Although there are a number of general purpose database technologies[3], these often sacrifice performance or detail in favour or interoperability.

Query Manager: The QM provides a means through which to extract data from the CD and to generate new insights. The QM is attempting to solve the extremely complex task of enabling querying of data across databases and across database technologies. DAFNI aims to create a new cross database query

[3] An example of this is MongoDB which although a document database is designed as a general database with support for documents, graphs and in part can enable use of relational data (https://www.mongodb.com/what-is-mongodb).

manager specifically to address the distinct requirements of its modelling community. This will likely represent one of the biggest implementation challenge for the project.

Model Workflow Engine: System-of-systems modelling is hard because of the required model interoperability. Each model has a unique set of dependencies making coupling extremely time consuming and often an impossible task. To address this, DAFNI is developing the MWE that will utilise containerisation to encapsulate functionality and dependencies whilst maintaining the workflow structure. Each workflow will consist of a series of chained containers characterising each operation with a centralised job manager to handle data collection and data exchange between the containers. This flexibility can allow for more dynamic allocation of resources within DAFNI and allows for use of any operation that can be containerised to be used within the workflows (e.g. data transformation and visualisation).

4 Current Status

The DAFNI construction programme (2017–2021) is currently in progress, with an early detailed requirements and design phase, and detailed architecture. At the time of writing, it is undergoing its initial implementation and deployment.

As the DAFNI platform is evolving, a series of pilots is validating the functionality and refine the platform requirements. The first pilot focused on the NISMOD-1 *System of Systems* modelling application (Ives, Pant & Robson 2018) developed as part of ITRC-MISTRAL project (Mendes 2016) and hosted at Newcastle University. NISMOD-1 is a collection of codes that currently run on a single machine supporting five models of UK infrastructure: Energy Supply; Water Supply; Solid Waste; Transport; and Waste Water. The models explore the needs of these infrastructure components based on estimates of trends in areas such as population growth, economic growth, and climate change. A key need for NISMOD-I is sensitivity analysis: determining whether the uncertainty of a given input parameter change the “preferred” solution to an infrastructure problem. Without proper understanding of sensitivity, predictions are of limited use. With a large number of input parameters to each of the NISMOD models, a full sensitivity analysis requires running very many simulations while varying each input in turn. This is highly compute intensive and is impractical to run through the existing NISMOD GUI. The first pilot ported the NISMOD-1 system onto the DAFNI cluster and provided a batch processing system to submit multiple sensitivity analysis jobs using HTCondor. As a result, the NISMOD-1 team have successfully run a number of sensitivity analyses on the Water Supply models and achieved a speed up of 10 times over that of the original service.

The NISMOD-1 sensitivity analysis is an example of the benefits that can be derived by moving existing, proven infrastructure models onto a high throughput cluster. The analysis can easily be extended using private and public cloud systems. Moving the data as well as the software to the DAFNI system is key to obtaining scalable performance.

As DAFNI is developed and deployed, it will address the increasing the demand of research capabilities, in order for the UK’s national infrastructure research effort to remain at the cutting edge. Further, as DAFNI begins to develop into maturity it could act also as a focus for government and industry, and it is working towards ensuring the expected availability and platform robustness to achieve this goal.

References

Organisation for Economic Co-operation and Development. 2012. Strategic Transport Infrastructure Needs to 2030, OECD Publishing,

Ives, M., Pant, R., Robson, C. 2018 NISMOD: http://www.itrc.org.uk/nismod/#.WrT2ypPFJ24.

Mendes, M. 2016. The UK Infrastructure Transitions Research Consortium: Mistral Programme: 2016 - 2020, http://www.itrc.org.uk/wp-content/PDFs/ITRC-Mistral-intro-2016.pdf.