Dafni: a computational platform to support infrastructure systems research

Brian Matthews, Jim Hall, Michael Batty, Simon Blainey, Nigel Cassidy, Ruchi Choudhary, Daniel Coca, Stephen Hallett, Julien J. Harou, Phil James, Nik Lomax, Peter Oliver, Aruna Sivakumar, Theodoros Tryfonas, Liz Varga

Published version information

Citation: B Matthews et al. Dafni: a computational platform to support infrastructure systems research. Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction (2023).

DOI: 10.1680/jsmic.22.00007

Submitted: 31 January 2022

Published online in 'accepted manuscript' format: 20 March 2023

Manuscript title: Dafni: a computational platform to support infrastructure systems research

Authors: Brian Matthews[1], Jim Hall[2], Michael Batty[3], Simon Blainey[4], Nigel Cassidy[5], Ruchi Choudhary[6], Daniel Coca[7], Stephen Hallett[8], Julien J. Harou[9], Phil James[10], Nik Lomax[11], Peter Oliver[1], Aruna Sivakumar[12], Theodoros Tryfonas[13], Liz Varga[14]

Affiliations: [1]Scientific Computing Department, Science and Technology Facilities Council, Didcot, UK. [2]School of Geography and the Environment, University of Oxford, Oxford, UK. [3]Centre for Advanced Spatial Analysis, University College London, London, UK. [4]Engineering and Physical Sciences, University of Southampton, Southampton, UK. [5]Department of Civil Engineering, University of Birmingham, Birmingham, UK. [6]Department of Engineering, University of Cambridge, Cambridge, UK. [7]Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK. [8]Centre for Environmental and Agricultural Informatics, Cranfield University, Cranfield, UK. [9]Department of Mechanical, Aerospace & Civil Engineering, University of Manchester, Manchester, UK. [10]School of Engineering, University of Newcastle, Newcastle, UK. [11]School of Geography, University of Leeds, Leeds, UK. [12]Department of Civil and Environmental Engineering, Imperial College, London, UK. [13]Department of Civil Engineering, University of Bristol, UK. [14]Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK.

Corresponding author: Brian Matthews, Scientific Computing Department, UKRI - Science and Technology Facilities Council, Rutherford Appleton Laboratory, Chilton, Didcot, OX11 0QX, UK. Tel.: +44 (0)1235 446648

E-mail: brian.matthews@stfc.ac.uk

Abstract

Research into the engineering of infrastructure systems is increasingly data-intensive. Researchers build computational models to explore scenarios such as investigating the merits of infrastructure plans, analysing historical data to inform system operations, or assessing the impacts of infrastructure on the environment. Models are more complex, at higher resolution and with larger coverage. Researchers also require a ‘multi-systems’ approach to explore interactions between systems, such as energy and water with urban development, and across scales, from buildings and streets to regions or nations. Consequently, researchers need enhanced computational resources to support cross-institutional collaboration and sharing at scale. The Data and Analytics Facility for National Infrastructure (Dafni) is an emerging computational platform for infrastructure systems research. It provides high-throughput compute resources so larger data sets can be used, with a data repository to upload data and share it with collaborators. Users’ models can also be uploaded and executed using modern containerisation techniques, giving platform independence, scaling and sharing. Further, models can be combined into workflows, supporting multi-systems modelling, and generating visualisations to present results. Dafni forms a central resource accessible to all infrastructure systems researchers in the UK, supporting collaboration and providing a legacy, keeping data and models available beyond a project’s lifetime.

1. Introduction

The infrastructure systems of a country or region, including energy supplies, water systems, transport networks, digital communications, land use, and the built environment are key investments for economic, social and environmental well-being (Thacker et al. 2019), and one estimate suggests that US$94 trillion of investments will be required by 2040 for new and replacement infrastructure (Global Infrastructure Hub, 2017). However, the impact of this investment is hard to predict as infrastructure is subject to environmental, social and economic pressures. Researchers in across disciplines, including environmental sciences, geography, civil engineering, urban planning and economics use computational modelling and analysis to explain and predict the effects of change on infrastructure systems, whilst policy makers use the outputs of such models to inform 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 a locality or building (Hall 2019). Thus, there is a need for advanced large-scale computing and data infrastructure to manage and analyse data, together with cloud systems for on-demand remote access.

The Data and Analytics Facility for National Infrastructure (Dafni) (see www.dafni.ac.uk) is a major national facility under development in the UK to provide world-leading capability to advance infrastructure systems research. It provides a scalable platform supporting storage and querying of heterogeneous national infrastructure data-sets, and the execution, creation and visualisation of complex modelling applications. This platform improves the quality and opportunities for National Infrastructure Systems research whilst reducing the complexity of using data and models for end users. Thus Dafni enables new advances in infrastructure research, and improve the readiness of research tools and methods for real-world challenges at scale, nationally and internationally.

This paper presents Dafni, discussing the motivations, aims and approach behind its development. It goes on to discuss its architecture, and give more details on its approach to handling data and supporting user models in multi-systems workflows, Some pilot studies are discussed further, demonstrating how Dafni is being used to support research, including support for systems-of-systems modelling. Finally, the paper discusses emerging themes for new developments. In particular, there is a need for a richer information framework for data integration and exchange using common standards and semantics, while Digital Twins present additional challenges, with the combination of sensor networks and real-time data analysis adding an additional layer of complexity, so the role of Dafni to support an ecosystem of Digital Twins is considered.

2. Motivations and objectives

2.1 Challenges computational modelling of national infrastructure

Research undertaken in universities, exploring new models and algorithms, provide the leading edge of innovation in infrastructure systems analysis. Examples include QUANT (Batty and Milton, 2021), SPENSER (Lomax and Smith 2020), UDM (Ford et al., 2019), and NISMOD (Hall et al. 2016, Hall et al. 2017). This research can be leveraged to exploit modern computing capacity and cloud computing technology (e.g. Microsoft Azure, Amazon Web Services, Kubernetes) coupled with advances in big data analytics, simulation, modelling and visualisation to scale up and integrate such models. This approach provides more detailed, high-quality projections of the impact of infrastructure development decisions on the natural, economic and social environment, so that more effective choices can be made in the provision of new infrastructure, and thus that investment can best support human flourishing (Schooling et al., 2021). However, a number of challenges need to be overcome in order to take advantage of these advances in computing.

Using large-scale computing. The increase in data availability and resolution has enabled new modelling applications with increasing resolution and spatial and temporal coverage, with a corresponding increased demand for computational resources. However, maintaining large-scale resources, such as peta-scale data repositories or compute clusters, is costly and requires specialist skills, and high-performance computing (HPC) systems are technically challenging to access. Thus the compute resources available to individual research groups may be limited, making iterative development and optimisation processes time consuming and slow to complete. This restricts the ability of modellers to understand impacts of simulations at a national scale whilst maintaining fine-grain resolution.

Data sharing and security. Data can be difficult to find and access, while licensing of data and models can be complex, with varied commercial and security conditions presenting a barrier to data sharing between organisations. A common approach to data security is needed, backed by specialised skills and processes so that data can be shared and accessed with trusted partners.

Maintaining traceability. The need to ensure results are reliable and repeatable makes it essential to store versioned copies of the underlying datasets, with auditable provenance of results.

Distributed teams. Analyses 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 with each other, people and the environment are inherently complex and heterogeneous, and handling this complexity can become beyond the capacity of single teams.

Multi-systems modelling and data integration. For models to reflect more accurately real-world situations, there is a need for them to capture the interactions between systems in multi-systems models. These multi-systems models can be along two axes. Firstly, the components within a system can be aggregated into systems-of-systems at a higher-scale. Thus equipment items can be aggregated into models of plants, which themselves can be aggregated with other features into models of organisations, or of geographic localities, which in turn can be aggregated into cities, regions, or nations. Secondly, the interactions between different infrastructure systems, such as water, transport, energy, waste, communications and the built environment can be integrated into a common infrastructure model, with interactions with the natural, social and economic environments taken into account. This later case is becoming increasingly important, in for example the effects on the power distribution network of the change of transport to electric vehicles (Chaudry et al, 2022), or the effects on water supply of economic activity resulting from new transport links (ITRC Mistral 2020). The variety and variability of these models presents a significant challenge as extensive domain expertise is required to exploit each model. Further, the models themselves need to be interoperable, via programmatic interfaces and common libraries. Data needs to be shared and exchanged across the models and domains, and across different scales and semantic representations. Thus a common data integration framework is needed for a flexible multi-systems modelling system.

2.2 Objectives of Dafni

In response to these challenges, Dafni has been developed as a shared platform to provide a dedicated compute resource for the National Infrastructure modelling community. Dafni has been supported by the UK Collaboratorium in Research on Infrastructure and Cities (UKCRIC, see https://www.ukcric.com/) in a 4 year development phase (2017-21) involving a consortium of 12 UK Universities, led by the University of Oxford. The Scientific Computing Department of the UK’s Science and Technology Facilities Council (STFC) was commissioned by the consortium as development partner and host. STFC’s role is the support of national scientific research infrastructure and was seen as being well-suited to the delivery of the platform.

The objectives of Dafni are to provide a common platform to support scalable, collaborative research into infrastructure systems, as follows.

A common platform for sharing and combining data and models. The Dafni platform provides a common computing hub for the infrastructure systems research community to store data and models and make them available to trusted collaborators.

A shared space to support collaborations and build multi-systems models. The shared platform can enable collaborations to build and execute more complex multi-system models at scale, accessing common data and combining shared models into workflows.

A legacy environment. Access to models, data and results in the repository can be made available and usable for the long-term, providing a legacy environment persisting beyond the lifetime of individual research projects, and traceability of the provenance of results.

Dafni is intended to improve the opportunities for and quality of research; and reduce the complexity of all aspects related to conducting the research in a high performance computing environment, including data access and processing, model execution, security, and visualisation. It enables the combination of these features into a functional platform that addresses the data, licencing, and scalability challenges identified above.

2.3 Analogous facilities

Within the research infrastructure landscape there are other facilities that have a similar role to Dafni within their respective domains, including the following.

The Australian Urban Research Infrastructure Network (AURIN 2022) provides compute infrastructure and expert support for urban, regional and social science researchers across Australia. It develops advanced data and analytic capability for the adoption of high-impact research within government and industry, holding reference data sets for long-term availability, and providing simulation and visualisation capability for decision support. It does not provide a user environment with capability for users to supply their own models and data resources and construct their own workflows.

The Biodiversity and Climate Change Virtual Laboratory (BCCVL) (Hallgren et al. 2016) is an Australian government funded initiative aiming to reduce the barrier to entry into high-resolution climate change and biodiversity impact modelling, utilising high-end HPC infrastructure for non-technical literate researchers. Via the 'virtual data laboratory' users can access over 4000 climate datasets and 300 environmental descriptors collocated onto a common geospatial and temporal grid. Further, users can execute pre-validated, managed models and either download results for custom offline post-processing or utilise one of several pre-defined techniques to analyse their results.

The Urban Centre for Computation and Data (UrbanCCD) (UrbanCCD 2022) is a joint initiative at the University of Chicago and Argonne National Laboratory to support the study of urban science. The UrbanCCD does not provide a dedicated computing facility, but researchers may make use of the Argonne Leadership Computing Facility for batch computing.

JASMIN (Lawrence et al. 2013) is a globally unique data intensive supercomputer for environmental science and currently supports over 1500 users on over 200 projects. JASMIN users research topics ranging from earthquake detection and oceanography to air pollution and climate science. JASMIN provides the UK and European climate and earth-system science communities with the ability to access very large sets of environmental data, which are typically too big to download and process using their own computers. This reduces the time it takes to test new ideas and get results from months or weeks to days or hours.

3. Dafni architecture and capabilities

3.1 Architecture overview

Dafni is designed around a number of core components, as illustrated in Figure 1, and briefly described below.

Dafni is hosted on a dedicated hardware cluster currently providing some 792 cores and 10 GPU nodes, with 2 PB of storage with a combination of fast and long-term storage available, which can be configured for different performance characteristics. Long term storage uses the MinIO object-store system (see https://min.io/), while the compute cluster is configured using Kubernetes. Kubernetes is an open-source container orchestration system for automating software deployment, scaling, and management (see https://kubernetes.io/); this allows the flexible deployment of user applications. Dafni has developed a number of components on this foundation to support user-applications.

- **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 data-store; a data catalogue; and a data access and publication service.
- **National Infrastructure Modelling Service (NIMS).** The NIMS provides support to improve performance of existing models, reduce the complexity of creating models and facilitate the creation of multi-systems models. It includes a model catalogue and a workflow creation and execution framework based on ARGO (see https://argoproj.github.io/).
- **National Infrastructure Cloud Environment (NICE).** The NICE provides a scalable cloud environment with a number of Platform as a Service (PaaS) offerings to users, including Jupyter notebooks (see https://jupyter.org/). Currently the NICE is used to within the internal architecture of Dafni, to deploy services within the cluster.

- **National Infrastructure Visualisation Suite (NIVS):** The NIVS supports visualisation tools to facilitate understanding of data, models, outputs and translation of findings to decision makers. This includes traditional visualisation as a service (e.g. graph and tabular representations) and user developed analyses using Jupyter Notebooks.
- **Dafni Security Service (DSS):** The DSS manages the security of the platform, which allows users to seamlessly access and use those services they have rights to, while at the same time maintaining security and integrity of data. Services include authentication, authorisation, monitoring, and group management.

These components have been implemented in a microservice architecture (Jamshidi et. al. 2018). This allows the capabilities within Dafni to be developed independently with an extensible and flexible delivery of the platform in line with the evolving nature of the National Infrastructure modelling landscape. Two central components, the NID, and the NIMS, are discussed in more detail.

3.2 The National Infrastructure Database

The National Infrastructure Database (NID) is the foundation of Dafni, a core service that allows researchers to upload, access and share datasets which are necessary to their research. It then manages the provision of data to models, workflows and visualisations, with outputs from model executions published back to the NID, allowing the research community access to the latest model outputs.

The NID uses a MinIO object storage instance with a capacity of up to 900 terabytes. The adoption of object storage allows Dafni to be flexible and store any data in any format required. MinIO provides a Cloud Native solution which integrates seamlessly into Dafni's underlying Kubernetes environment. This is supported by databases which store and manage the metadata records for each dataset, providing Data Search and Data Versioning capabilities around the data-store itself.

Dafni researchers interact with the data store via the Dafni Data Repository, illustrated in Figure 2, a tailor-made repository service that allows researchers to upload data to the NID and manage the access to that data, allowing others on the platform to access it either globally, individually or through groups. In addition, researchers can update their datasets and create new versions, while all registered users on the Dafni platform can access and download the open access datasets.

Metadata. Dafni has adopted a rich metadata schema, based on DCAT V2 (World-Wide Web Consortium, 2020), a World-Wide Web Consortium recommendation for interoperability between data catalogues, augmented with additional features supporting geospatial data, such as INSPIRE categories (INSPIRE 2022) and Geonames (see http://www.geonames.org/) for spatial coverage. This provides a Search and Discovery service on the Dafni platform and positions the platform for interoperability with other data stores. The approach is to encourage users to provide a rich metadata record of data from the start, thus supporting the access and reuse of data according to the FAIR data principles (Wilkinson et al. 2016).

The metadata combines top-level contextual and licencing information with more detailed dataset attributes, which drill down to the file level. This is combined with a description of the dataset's ownership and publication history in order to provide traceability and link each dataset on Dafni to its infrastructure research community. The metadata is indexed by the Data Search and Discovery service, built using ElasticSearch (see https://www.elastic.co), a powerful full-text search and analytics engine. Users can find datasets of interest to their research via a text search or by spatio-temporal filtering. Filters by data source, theme and file format are also supported.

3.3 The National Infrastructure Modelling Service

The National Infrastructure Modelling Service (NIMS) encompasses both the model catalogue and model workflow systems on Dafni. The purpose of the NIMS is to allow Dafni users to run user supplied models through the use of workflows without specialised knowledge of HPC systems or programming.

Containerisation. The execution of user generated models and their combination into multi-systems models is challenging because of the compatibilities required between models. Each model, developed by independent groups of researchers and software engineers, has a set of dependencies on programming language, packages and libraries. These dependencies make porting models onto a common platform a complex and time consuming process, a significant barrier to the use of high-performance computing. Further, coupling models together requires the sharing of data in interoperable formats and access to APIs for models to communicate.

To simplify these challenges, the Dafni NIMS utilises containerisation using the Docker packaging system (see https://www.docker.com) to encapsulate functionality and dependencies. Docker builds self-contained packages encapsulating the model executable together with its execution environment, and also bundling configuration and library files. A model definition file in the YAML ("YAML Ain't Markup Language" see https://yaml.org/) format is also provided to accompany the "dockerised" model, specifying the interfaces, input parameters and data sets, and outputs to the model, together with metadata that will be displayed about the model catalogue. Dockerised models can then be uploaded onto the platform and can be deployed and executed via the Kubernetes system. Thus Dafni can execute user code independent of their dependencies.

Model Catalogue. Models are uploaded into a Model Catalogue, illustrated in Figure 3a, a repository of models, based on the Harbor, an open-source system providing a registry of containers (see https://goharbor.io). User metadata describing the model is supplied by the user on upload, providing a searchable catalogue, subject to the user and group access permissions set within the DSS.

Workflows. Workflows allow users to create multi-systems models and to output the results of these workflows to share with other users. Each workflow consists of a series of chained containers characterising each operation with a centralised job manager to handle data collection and data exchange between the containers. On execution, the Kubernetes orchestration engine allocates resources and deploys the workflow into "pods" across a number of nodes in the cluster where each can be executed on their own resources. This flexibility can allow for more dynamic allocation of resources within Dafni and allows any operation that can be containerised to be used within the workflows (e.g. data transformation and visualisation).To build a workflow, users construct a series of interconnected steps, as shown in

Figure 3b. The step types are described below.

Model. The model step facilitates the execution of a model. Users can choose the model from the model catalogue and set any input parameters for the model, with data selected from the NID. Models can also be chained together in the workflow, passing output data from model into the inputs of the next to allow for multi-systems modelling. For example, a model which simulates population growth can be chained to a model which relied on population numbers to predict house prices, thus allowing the exploration of the effect on house prices of different demographic scenarios.

Iterator. Iterators allow the same step in the workflow to be repeated multiple times whilst changing parameters within a given range either randomly or with a pre-defined increment. This allows multiple executions of the same model to be completed in parallel to one another where possible, so many runs of the same model can be completed across a range of values, or across random values in Monte-Carlo simulations where the same model can be run multiple times with different parameters.

Publisher. The publisher step takes outputs from a model and ingests them into the NID. The user supplies metadata about the resultant dataset which will be displayed in the data catalogue.

Visualisation. The visualisation step takes the outputs of a model and creates a visualisation builder containing those outputs using the NIVS. This allows the user to go directly from the results of a finished workflow into generating graphs or charts from those results in a visualisation builder, or via a user programmable Jupyter notebook.

4. Using Dafni

The initial phase of the Dafni construction programme (2017–2021) was a requirements and design study that developed a detailed architecture. As the Dafni platform evolved, a series of pilots validated the functionality and refined requirements while demonstrating the benefits of its additional computing power. These pilots included railway station planning (Young et al., 2019) and demand prediction (Young et al., 2019), 5G cell tower placement, house demand and pricing, and urban and economic development. Further a programme of Dafni "Champions" was introduced, looking at case studies in transport, including using the MATSim multi-agent transport simulation framework (Horni et al., 2016) and exploring how Dafni might support a digital twin of road traffic in conjunction with the Sheffield Urban Observatory.

A significant pilot involved working closely with the UK Infrastructure Transitions Research Consortium's (ITRC, see https://www.itrc.org.uk/) NISMOD system, a key example of a collaborative environment within infrastructure systems research. Before implementation on Dafni, NISMOD access was only available to members of the immediate research group and the model had not been optimised for more general research challenges. The first Dafni pilot focused on the NISMOD-1 System of Systems modelling application developed as part of the ITRC project and hosted at Newcastle University. NISMOD-1 ran on a single machine supporting five models of UK infrastructure: Energy Supply (Chaudry et al., 2022); Water Supply (Dobson et al., 2020); Solid Waste; Transport (Blainey & Preston, 2019); 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-1 is sensitivity analysis: determining whether the uncertainty of a given input parameter changes the “preferred” solution to an infrastructure problem (Pianosi et al., 2016); without proper understanding of this 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, a highly compute intensive process. The first pilot ported the NISMOD-1 system onto the Dafni cluster and provided a batch processing system to submit multiple sensitivity analyses. As a result, the NISMOD-1 team have successfully run a number of sensitivity analyses on the Water Supply models and achieved a speed of up to 10 times faster than the original service. This demonstrates the benefits that can be derived by moving existing, proven infrastructure models onto a high throughput cluster; moving the data as well as the software to the Dafni system is key to obtaining scalable performance. The work on the NISMOD pilot has continued through the development of NISMOD-2 and its implementation on Dafni as the platform has evolved. Workflows supporting the NISMOD scenarios are now available on Dafni, which provides NISMOD users with a long-term execution environment.

Further projects are now using the Dafni platform. The Open Climate Impact project (OpenClim see https://gtr.ukri.org/projects?ref=NE/T013931/1) is developing a modelling framework to explore the impact of future climate change scenarios on infrastructure, exploring such factors as flood events in urban environments, the effect of extreme heat events on the population, and the effect on agriculture. The project has particular emphasis on adapting the environment to climate change, and the mitigating effects those adaptations might have. Dafni is being used in the project as a common modelling framework to connect the different models and to provide a legacy space so that the workflow can be accessed in the long term. The Centre for Greening Finance and Investment (see https://www.cgfi.ac.uk/) is also planning to use the Dafni platform similarly to host and develop a shared data and modelling framework to explore how environmental change will impact the risks on investment, insurance and other activities within the finance industry.

5. Future developments

5.1 Enhanced data framework

Data remains the central driver for the future research and exploitation of computational models of infrastructure systems, and richer handling of data would enhance the power and range of Dafnit for researchers. The following extensions and enhancements to the NID are being explored.

- **FAIR Data publication and Data Curation.** Dafni supports a metadata description for data and models, and thus partially satisfies the FAIR principles. In order for Dafni to support reusable reference data within the research community, this needs to be enhanced to support a data publication pipeline, underpinned by data curation processes to update and maintain data for the long term.
- **Scaling and querying large data.** The handling of large data sets within workflows can be inefficient as data copying and transfer is a high-latency exercise, and frequently queries are applied early in the workflow to extract the relevant data-slice suitable for processing. Large-data immutable datasets can be treated as static objects

which can be accessed in a common manner across different processes, with data-slicing taking a "data-cube" approach.

- **Interoperability Framework.** The data and modelling framework in Dafni has the advantage of being generic and thus can accept data in any format. However, in linking models into workflows, there still remains the need to undertake data manipulation tasks, such as queries, format transformation, projections between scales and other data transformations. By providing enhanced support for particular data formats and providing a suite of "data adaptors" or "data transforms" in an interoperability framework, the process of data manipulation can be simplified.
- **Semantic framework.** A further extension to the interoperability framework would be to introduce the use of ontologies. By supporting a selected suite of ontologies, rich data enhanced mappings can be supported within workflows as well as enhancing the search and discovery service. An exploration of use and availability of suitable ontologies with recommendation for future development was undertaken within the Dafni Champions programme (Varga et al. 2021, Varga et al. 2022).

5.2 Dafni and digital twins

The concept of Digital Twins (Batty, 2018; Callcut, 2021) has emerged over the last decade as a key technology for the future planning delivery and operation of infrastructure systems. There has been a high level of interest from government and industry in investing in Digital Twins as a tool to predict, optimise and control the outcomes of infrastructure investment. Initiatives such as the Digital Twin Hub (see https://digitaltwinhub.co.uk/) have been developing frameworks for combining digital twin models into a ‘National Digital Twin’ (NDT), “a digital model of our national infrastructure which will be able both to monitor our infrastructure in real-time, and to simulate the impacts of possible events“ (National Infrastructure Commission, 2017), via the sharing of data and computational resources into a common digital twin ecosystem. Dafni can play a role in the development and deployment of Digital Twins for infrastructure systems by providing support for features that a NDT would require to be effective.

A NDT would require an ecosystem of models from a wide range of sources, which can be combined into large-scale, multi-system digital twins. Running multiple twins at scale would require high-performance computing environments which allows models to be executed rapidly, scaled up in resolution, and in geographical and temporal range. The platform independent approach of Dafni offers the basis of such an environment. Further, a NDT would need to support the combination of models into new workflows to support connected digital twins, and provide visualisations of results for human decision support, again supported within Dafni.

Digital Twins also bring significant challenges for data management and integration, and a NDT would require a wide range of different data sources to be brought together into a shared trusted data space, in a common information management framework. For the sustainability of the NDT, this would need to be maintained and curated for the long-term. Again Dafni already offers the NID, which could form the basis of such a data space.

Thus, Dafni can provide a hub for a digital twin infrastructure, supporting the research and development required to explore the opportunities of deploying digital twins within infrastructure systems development and operation.

Dafni is working with initiatives such as the UKCRIC Urban Observatory programme (see https://urbanobservatory.ac.uk) and interacting with key stakeholders, including the Connected Places Catapult (https://cp.catapult.org.uk/). It has developed pilots DTs, including on traffic management with Sheffield University. Further development on Dafni is exploring how to provide additional functionality to support Digital Twins, including extending the information management framework in Dafni, as discussed above, interacting with real-time input and streaming data systems, and working with Machine Learning to form adaptive models for decision making from historic data.

6. Conclusions

Dafni is an infrastructure platform to support the development of sharing of multi-systems models of national infrastructure. The data and model sharing allows access to models across infrastructure systems, and across collaborations. Dafni thus offers the infrastructure systems engineering community a space to leverage their research into wider and deeper applications.

Further, Dafni also supports collaborations between researchers, government and industry. Data and computation is central to the infrastructure engineering practise; the UK National Infrastructure Commission observes that “Data is part of infrastructure and needs maintenance in the same way that physical infrastructure needs maintenance” (National Infrastructure Commission, 2017). Dafni provides the basis for a trusted, common, vendor-neutral hub for data sharing and exchange with support for maintaining the value of these assets for the longer term. Further, it is recognised that while large-scale computing is valuable to solve new business and research challenges, it remains hard to access and use for non-specialists (see for example (Government Office for Science, 2021)). Dafni provides a user environment that seeks to overcome some of these technical barriers.

Dafni has transitioned from a development project to a service platform. This enables Dafni's operational growth to increase usage and capability to support research in EPSRC's Engineering Programme and related fields, so that the UK's national infrastructure research can remain at the cutting edge. Further, it allows Dafni to continue with its aim to support changing and sustainable infrastructure through working with government and industry.

Acknowledgements

Dafni has been supported by the EPSRC Grants EP/R012202/1 *UKCRIC National Infrastructure Database, Modelling, Simulation and Visualisation Facilities* (2017-21) and EP/V054082/1 *Dafni-Research Only Strategic Equipment (Dafni-ROSE) (2021-23)*. The authors would like to thank the members of the Dafni support and development team in STFC and the collaborators and champions involved specifying requirements and develop pilot cases within the Dafni environment.

References

AURIN (2022). What we do. See https://aurin.org.au/about-aurin/what-we-do/ (accessed 24 January 2022)

Batty M. (2018) Digital twins. *Environment and Planning B: Urban Analytics and City Science*. 45(5):817-820. https://doi.org/10.1177/2399808318796416

Batty M. and Milton R (2021). A New Framework for Very Large-Scale Urban Modelling, Urban Studies, 58(15), 3071–3094 https://doi.org/10.1177/0042098020982252

Blainey SP & Preston JM (2019) 'Predict or Prophesy? Issues and Trade-Offs in Modelling Long-Term Transport Infrastructure Demand and Capacity', Transport Policy 74(2):165-173. https://doi.org/10.1016/j.tranpol.2018.12.001

Callcut M, Agliozzo JC, Varga L, McMillan L (2021) Digital twins in civil infrastructure systems. Sustainability 13(20), 11549 https://doi.org/10.3390/su132011549

Chaudry M, Jayasuriya L, Blainey S, Lovric M, Hall JW, Russell T, Jenkins N & Wu J (2022), 'The implications of ambitious decarbonisation of heat and road transport for Britain's net zero carbon energy systems', *Applied Energy*, 305:117905 https://doi.org/10.1016/j.apenergy.2021.117905

Dobson B, Coxon G, Freer J, Gavin H, Mortazavi-Naeini M and Hall JW (2020). The spatial dynamics of droughts and water scarcity in England and Wales, Water Resources Research, 56 (9): e2020WR027187. https://doi.org/10.1029/2020WR027187 .

Ford A, Barr S, Dawson R, Virgo J, Batty M, Hall J. (2019) A multi-scale urban integrated assessment framework for climate change studies: a flooding application. Computers, Environment, and Urban Systems, 75, 229-243. https://doi.org/10.1016/j.compenvurbsys.2019.02.005

Global Infrastructure Hub, (2017). Global Infrastructure Outlook: 2017. https://www.oxfordeconomics.com/recent-releases/99f4fa86-a314-4762-97c6-fac8bdcbe40a (accessed 24 January 2022)

Government Office for Science (2021). Large-scale computing: the case for greater UK coordination. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1018875/UK_Computing_report_-_Final_20.09.21.pdf (accessed 24 January 2022)

Hall JW, Tran M, Hickford AJ and Nicholls RJ. (eds.) (2016). The Future of National Infrastructure: A System of Systems Approach, Cambridge University Press. ISBN: 978-1-10758-874-5

Hall JW. Thacker S, Ives MC, Cao,Y, Chaudry M, Blainey SP and Oughton E.J (2017). Strategic analysis of the future of national infrastructure, Proceedings of the Institution of Civil Engineers: Civil Engineering, 170(1): 39-47. https://doi.org/10.1680/jcien.16.00018

Hall, J.W. (2019). A Simulation Tool to Guide Infrastructure Decisions: System-of-Systems Modeling Aids Prioritization and Uncertainty Planning. IEEE Systems, Man and Cybernetics Magazine, 5(3): 10-20. https://doi.org/10.1109/MSMC.2019.2913565.

Hallgren W, Beaumont L, Bowness A, Chambers L, Graham E, Holewa H, Laffen S, Mackey B, Nix H, Price J, Vanderwal J, Warren R and Weis G. (2016). The Biodiversity and

Climate Change Virtual Laboratory: Where ecology meets big data. Environmental Modelling & Software, vol. 76, pp. 182-186, https://doi.org/10.1016/j.envsoft.2015.10.0251364

Horni, A., Nagel, K. and Axhausen, K.W. (eds.) (2016) *The Multi-Agent Transport Simulation MATSim.* London: Ubiquity Press. https://doi.org/10.5334/baw.

INSPIRE (2022). Topic categories in accordance with EN ISO 19115. https://inspire.ec.europa.eu/metadata-codelist/TopicCategory (accessed 24 January 2022)

ITRC Mistral (2020). A sustainable Oxford-Cambridge corridor? Spatial analysis of options and futures for the Arc https://www.itrc.org.uk/wp-content/uploads/2020/01/arc-main-report.pdf (accessed 24 January 2022)

Jamshidi P, Pahl C, Mendonça N C, Lewis J and Tilkov S. (2018). Microservices: The Journey So Far and Challenges Ahead. *IEEE Software*, vol. 35, no. 3, pp. 24-35, May/June 2018, https://doi.org/10.1109/MS.2018.2141039

Lawrence B N, Bennett V L, Churchill J, Juckes M., Kershaw P, Pascoe S, Pepler S,. Pritchard M. and Stephens, A. (2013) Storing and manipulating environmental big data with JASMIN, *2013 IEEE International Conference on Big Data*, 2013, pp. 68-75, https://doi.org/10.1109/BigData.2013.6691556.

Lomax N, and Smith AP, (2020). Dafni Pilot 4: SPENSER - Synthetic Population Estimation and Scenario Projection model. https://dafni.ac.uk/wp-content/uploads/2020/05/dafni-pilot-4-dafni-hosts-population-forecast-model.pdf (accessed 24 January 2022)

National Infrastructure Commission (2017). Data for the Public Good. NIC, London, UK https://nic.org.uk/app/uploads/Data-for-the-Public-Good-NIC-Report.pdf (accessed 24 January 2022)

Pianosi F, Rougier J, Freer J, Hall JW, Stephenson DB, Beven K, Wagener T (2016). Sensitivity analysis of environmental models: a systematic review with practical work. Environmental Modelling and Software, 79:214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Schooling J, Enzer M and Broo DG (2021). Flourishing systems: re-envisioning infrastructure as a platform for human flourishing. Proceedings of the Institution of Civil Engineers – Smart Infrastructure and Construction, https://doi.org/10.1680/jsmic.20.00023

Thacker S, Adshead D, Fay M, Hallegatte S, Harvey M, Meller H, O'Regan N, Rozenberg J and Hall JW (2019). Infrastructure for Sustainable Development, Nature Sustainability, 2 324–331. https://doi.org/10.1038/s41893-019-0256-8.

Urban Center for Computations and Data (2022) About UrbanCCD. see http://www.urbanccd.org/about/#about-urbanccd (accessed 24 January 2022)

Varga L, McMillan L, Hallett S, Russell T, Smith T, Truckell I, Postnikov A, Rodger S, Vizcaino N, Perkins B, Matthews B, Lomax N (2021). Infrastructure Research Ontologies. http://dafni.ac.uk/wp-content/uploads/2021/05/IRO-final-report-31-03-2021.pdf (accessed 24 January 2022)

Varga L, McMillan L, Hallett S, Russell T, Smith T, Truckell I, Postnikov A, Rodger S, Vizcaino N, Perkins B, Matthews B, Lomax N. (2022). Infrastructure and city ontologies. Proceedings of the Institution of Civil Engineers – Smart Infrastructure and Construction, https://doi.org/10.1680/jsmic.22.00005

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 https://doi.org/10.1038/sdata.2016.18

World-Wide Web Consortium (2020). Data Catalog Vocabulary (DCAT) - Version 2.

W3C Recommendation. https://www.w3.org/TR/vocab-dcat-2/ (accessed 24 January 2022)

Young MA & Blainey SP (2018) Developing railway station choice models to improve rail industry demand models, Transportation Planning and Technology 41(1):80-103. https://doi.org/10.1080/03081060.2018.1403745

Young MA, Blainey SP, Gowland T & Nagella S (2019). An Automated Online Tool to Forecast Demand for New Railway Stations and Analyse Potential Abstraction Effects. Transport Practitioners’ Meeting, 10-11 July 2019, Oxford. https://eprints.soton.ac.uk/432493/ (accessed 24 January 2022)

Figure 1. Dafni core components

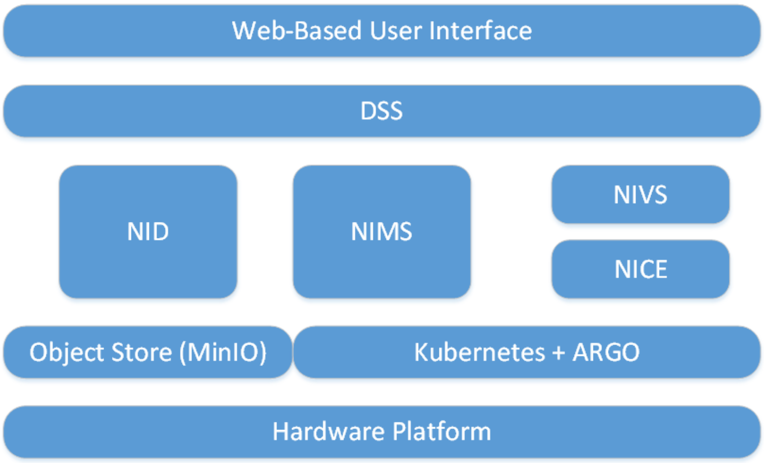
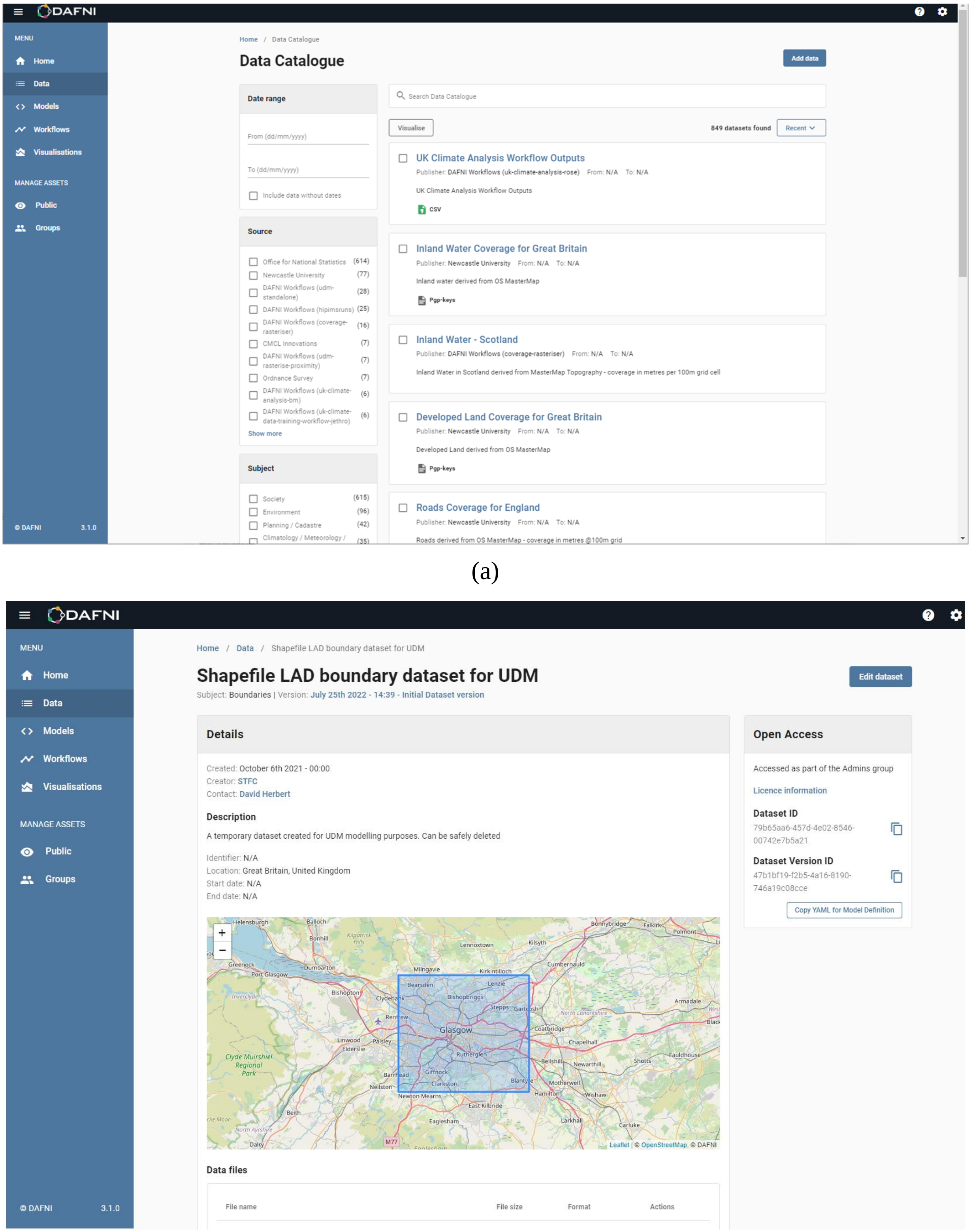


Figure 2. The Dafni NID a) the data catalogue, b) an entry for an example data set



DAFNI

MENU
- Home
- Data
- Models
- Workflows
- Visualisations

MANAGE ASSETS
- Public
- Groups

© DAFNI 3.1.0

Home / Data Catalogue

Data Catalogue

Add data

Date range

From (dd/mm/yyyy)

To (dd/mm/yyyy)

□ Include data without dates

Source

□ Office for National Statistics	(614)
□ Newcastle University	(77)
□ DAFNI Workflows (udm-standalone)	(28)
□ DAFNI Workflows (hipimsruns)	(25)
□ DAFNI Workflows (coverage-rasteriser)	(16)
□ CMCL Innovations	(7)
□ DAFNI Workflows (udm-rasterise-proximity)	(7)
□ Ordnance Survey	(7)
□ DAFNI Workflows (uk-climate-analysis-bm)	(6)
□ DAFNI Workflows (uk-climate-data-training-workflow-jethro)	(6)

Show more

Subject

□ Society	(615)
□ Environment	(96)
□ Planning / Cadastre	(42)
□ Climatology / Meteorology /	(35)

Search Data Catalogue

Visualise   849 datasets found   Recent

□ **UK Climate Analysis Workflow Outputs**
Publisher: DAFNI Workflows (uk-climate-analysis-rose)   From: N/A   To: N/A
UK Climate Analysis Workflow Outputs
csv

□ **Inland Water Coverage for Great Britain**
Publisher: Newcastle University   From: N/A   To: N/A
Inland water derived from OS MasterMap
Pgp-keys

□ **Inland Water - Scotland**
Publisher: DAFNI Workflows (coverage-rasteriser)   From: N/A   To: N/A
Inland Water in Scotland derived from MasterMap Topography - coverage in metres per 100m grid cell

□ **Developed Land Coverage for Great Britain**
Publisher: Newcastle University   From: N/A   To: N/A
Developed Land derived from OS MasterMap
Pgp-keys

□ **Roads Coverage for England**
Publisher: Newcastle University   From: N/A   To: N/A
Roads derived from OS MasterMap - coverage in metres @100m grid

(a)

DAFNI

MENU
- Home
- Data
- Models
- Workflows
- Visualisations

MANAGE ASSETS
- Public
- Groups

© DAFNI 3.1.0

Home / Data / Shapefile LAD boundary dataset for UDM

Shapefile LAD boundary dataset for UDM

Edit dataset

Subject: Boundaries | Version: July 25th 2022 - 14:39 - Initial Dataset version

Details

Created: October 6th 2021 - 00:00
Creator: STFC
Contact: David Herbert

Description

A temporary dataset created for UDM modelling purposes. Can be safely deleted

Identifier: N/A
Location: Great Britain, United Kingdom
Start date: N/A
End date: N/A

Data files

File name	File size	Format	Actions

Open Access

Accessed as part of the Admins group

Licence information

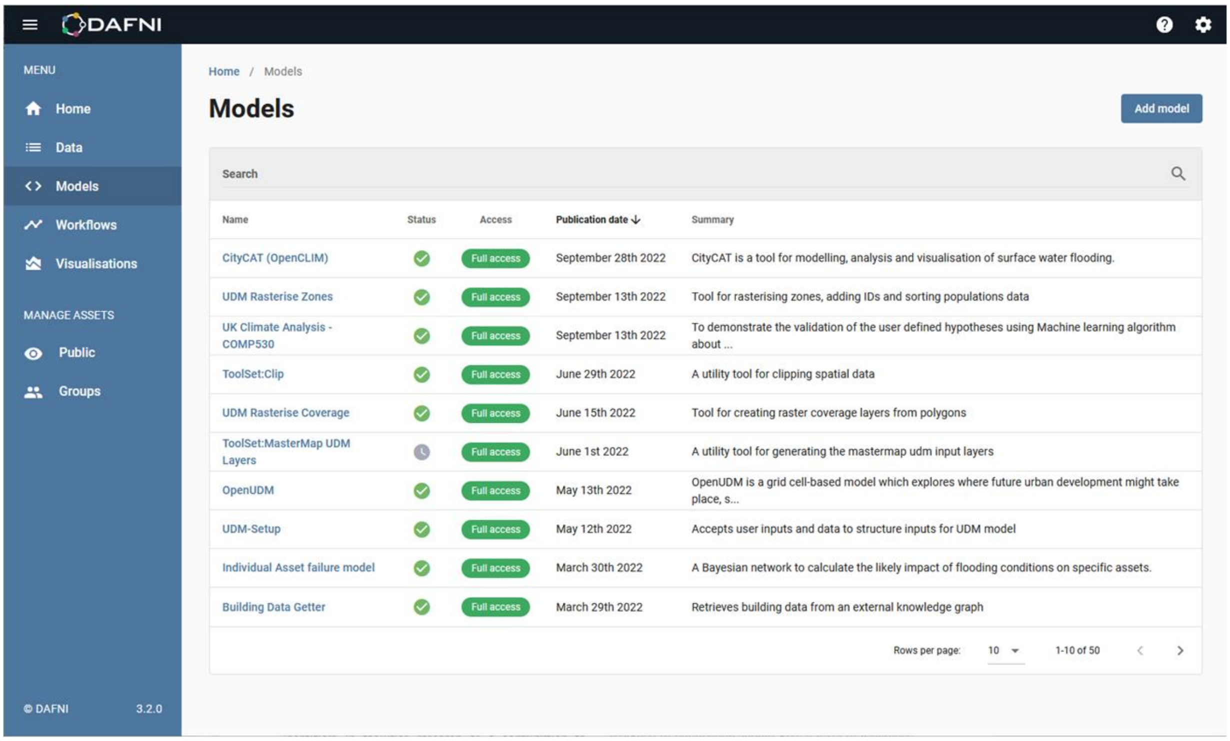
Dataset ID
79b65aa6-457d-4e02-8546-00742e7b5a21

Dataset Version ID
47b1bf19-f2b5-4a16-8190-746a19c08cce

Copy YAML for Model Definition

(b)

Figure 3. The Dafni NIMS a) the model catalogue, b) building a workflow

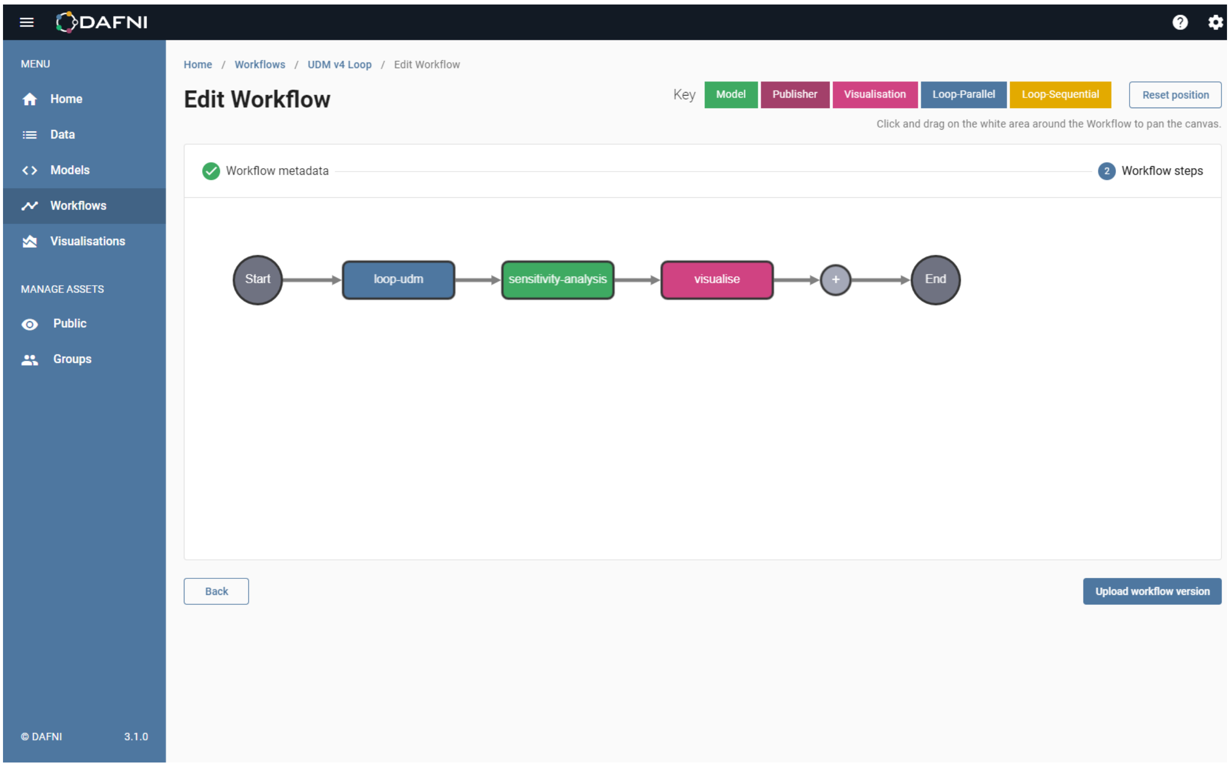


Models

Name	Status	Access	Publication date ↓	Summary
CityCAT (OpenCLIM)	✓	Full access	September 28th 2022	CityCAT is a tool for modelling, analysis and visualisation of surface water flooding.
UDM Rasterise Zones	✓	Full access	September 13th 2022	Tool for rasterising zones, adding IDs and sorting populations data
UK Climate Analysis - COMP530	✓	Full access	September 13th 2022	To demonstrate the validation of the user defined hypotheses using Machine learning algorithm about ...
ToolSet:Clip	✓	Full access	June 29th 2022	A utility tool for clipping spatial data
UDM Rasterise Coverage	✓	Full access	June 15th 2022	Tool for creating raster coverage layers from polygons
ToolSet:MasterMap UDM Layers	🕓	Full access	June 1st 2022	A utility tool for generating the mastermap udm input layers
OpenUDM	✓	Full access	May 13th 2022	OpenUDM is a grid cell-based model which explores where future urban development might take place, s...
UDM-Setup	✓	Full access	May 12th 2022	Accepts user inputs and data to structure inputs for UDM model
Individual Asset failure model	✓	Full access	March 30th 2022	A Bayesian network to calculate the likely impact of flooding conditions on specific assets.
Building Data Getter	✓	Full access	March 29th 2022	Retrieves building data from an external knowledge graph

Rows per page: 10 1-10 of 50

(a)



Edit Workflow

Click and drag on the white area around the Workflow to pan the canvas.

(b)