# DAFNI: a computational platform to support infrastructure systems research

24 January 2022

4931 words

## Author 1

- Dr. Brian Matthews
- Scientific Computing Department, Science and Technology Facilities Council, Didcot, U.K.
- **0000-0002-3342-3160**

## Author 2

- Prof Jim Hall
- School of Geography and the Environment, University of Oxford, Oxford, U.K
- 0000-0002-2024-9191

## Author 3

- Prof Michael Batty
- Centre for Advanced Spatial Analysis, University College London, London, U.K
- 0000-0002-9931-1305

## Author 4

- Dr. Simon Blainey
- Engineering and Physical Sciences, University of Southampton, Southampton, U.K
- 0000-0003-4249-8110

## Author 5

- Prof. Nigel Cassidy
- Department of Civil Engineering, University of Birmingham, Birmingham, U.K
- 0000-0002-1492-5413

## Author 6

- Dr. Ruchi Choudhary
- Department of Engineering, University of Cambridge, Cambridge, U.K
- 0000-0001-6915-3433

## Author 7

- Prof. Daniel Coca,
- Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, U.K
- 0000-0003-2878-2422

## Author 8

- Prof. Stephen Hallett,
- Centre for Environmental and Agricultural Informatics, Cranfield University, Cranfield, U.K
- 0000-0002-8776-7049

## Author 9

- Prof. Julien J. Harou,
- Department of Mechanical, Aerospace & Civil Engineering, University of Manchester, Manchester, U.K

- 0000-0003-1827-6155

**Author 10**

- Prof. Phil James
- School of Engineering, University of Newcastle, Newcastle, U.K
- 0000-0001-9248-0280

Author 11
- Dr Nik Lomax,
- School of Geography, University of Leeds, Leeds, U.K
- 0000-0001-9504-7570

**Author 12**

- Dr Peter Oliver
- Scientific Computing Department, Science and Technology Facilities Council, Didcot, U.K.
- 0000-0002-0445-8313

**Author 13**

- Dr Aruna Sivakumar
- Department of Civil and Environmental Engineering, Imperial College, London
- 0000-0003-2721-8299

**Author 14**

- Dr Theodoros Tryfonas
- Department of Civil Engineering, University of Bristol, U.K.
- 0000-0003-4024-8003

**Author 15**

- Prof. Liz. Varga
- Department of Civil, Environmental and Geomatic Engineering, University College London, London, U.K
- 0000-0001-6955-478X

**Full contact details of corresponding author.**

**Dr. Brian Matthews**

**Scientific Computing Department,**

**UKRI - Science and Technology Facilities Council,**

**Rutherford Appleton Laboratory, Chilton, Didcot, OX11 0QX, UK**

**brian.matthews@stfc.ac.uk**

**+44 (0)1235 446648**

**Abstract (150 – 200 words)**

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 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.  User models can also be uploaded and executed using modern containerisation techniques, giving platform independence, scaling and shared.  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.

**Keywords chosen from ICE Publishing list**. Information technology; Data; Numerical modelling; Infrastructure Planning; Digital Twin

# 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).   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 a variety of 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 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 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 datasets, and the execution, creation and visualisation of complex modelling applications.  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. Thus DAFNI will enable 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 it's architecture, and give a 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 development. 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; 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 coupled with advances in big data analytics, simulation, modelling and visualisation to scale up and integrate such models. This 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 of models, 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 varying 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.
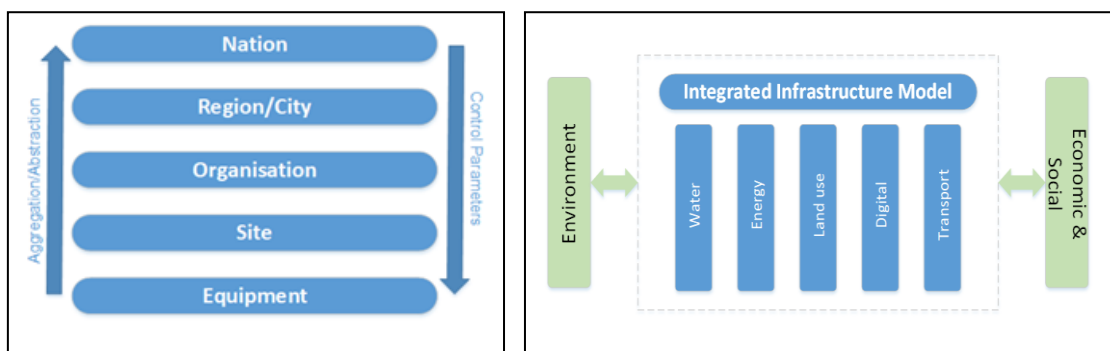


**Figure 1:  Integrating multi-systems models a). across scales; b). across sectors**

81  **Multi-systems modelling and data integration**.  As models reflect more and more real-world

82  situations, there is a need for them to reflect the interactions between systems in multi-systems

83  models.   These multi-systems models can be seen along two axes as illustrated in Figure 1.

84  Firstly, the components within a system can be aggregated into systems-of-systems at a higher-

85  scale.  Thus equipment items can be aggregated into models of plants, which themselves can

86  be aggregated with other components into models of organisations, or of geographic localities,

87  which in turn can be aggregated into cities, regions or national or international.  Secondly, the

88  interactions between different infrastructure systems, such as water, transport, energy, waste,

89  communications and the built environment can be integrated into a common infrastructure

90  model, with interactions with the natural, social and economic environments taken into account.

91  This later case is becoming increasingly important, in for example the effects on the power

92  distribution network of the change of transport to electric vehicles (Chaudry et al, 2022),, or the

93  effects on water supply of economic activity resulting from new transport links (ITRC Mistral

94  2020). The variety and variability of these models presents a significant challenge as extensive

95  domain expertise required to exploit each model.  Further, the models themselves need to be

96  interoperable, via programmatic interfaces and common libraries.  Data needs to be shared and

97  exchanged across the models and domains, and across different scales and semantic

98  representations.  Thus a common data integration framework is needed for a flexible multi-

99  systems modelling system.


100  ## 2.2  Objectives of DAFNI

101

102  In response to these challenges, DAFNI has been developed as a shared platform has been

103  developed to provide a dedicated compute resource for the National Infrastructure modelling

104  community. DAFNI has been supported by the UK Collaboratorium in Research on

105  Infrastructure and Cities (UKCRIC, see https://www.ukcric.com/ ) in a 4 year development

106  phase (2017-21) involving a consortium of 12 UK Universities, led by the University of Oxford.

107  The Scientific Computing Department of the UK's Science and Technology Facilities Council

108  (STFC) was commissioned by the consortium as development partner and host.  STFC's role is

109  the support of national scientific research infrastructure and was seen as being well suited to the

110  delivery of the platform.

111

112  The objectives of DAFNI are to provide a common platform to support scalable, collaborative

113  research into infrastructure systems, as follows.

114

115  **A common platform for sharing and combining data and models**.  The DAFNI platform

116  provides a common computing hub for the infrastructure systems research community to store

117  data and models and make them available to trusted collaborators.

118

119  **A shared space to support collaborations and build multi-systems models.**  The shared

120  platform can enable collaborations to build and execute more complex multi-system models at

121  scale, accessing common data and combining models into workflows.

122

123  **A legacy environment.**   Access to models, data and results in the repository can be made

124  available and usable for the long-term, providing a legacy environment, persisting beyond the

125  lifetime of individual research projects.

126

127  DAFNI is intended to improve the opportunities for and quality of research; and reduce the

128  complexity of all aspects related to conducting the research in a performance computing

129  environment, including data access and processing, model execution, security and visualisation.

130  It enables the combination of these features into a functional platform that addresses the data,

131  licencing and scalability challenges identified above.


132  ## 2.3  Analogous Facilities

133

134  Within the broader research infrastructure landscape there are other facilities that have a similar

135  role to DAFNI within their respective domains, including the following.

136

137    The Australian Urban Research Infrastructure Network (AURIN 2022) provides compute

138    Infrastructure and expert support for urban, regional and social science researchers across

139    Australia.  It develops advanced data and analytic capability for the adoption of high-impact

140    research within government and industry, holding reference data sets for long-term availability,

141    and providing simulation and visualisation capability for decision support.  It does not provide a

142    user environment with capability for users to supply their own models and data resources and

143    construct their own workflows.

144

145    Biodiversity and Climate Change Virtual Laboratory (BCCVL) (Hallgren et al. 2016) is an

146    Australian government funded initiative aiming to reduce the barrier to entry into high-resolution

147    climate change and biodiversity impact modelling, utilising high-end HPC infrastructure for non-

148    technical literate researchers. Via the 'virtual data laboratory' users can access over 4000

149    climate datasets and 300 environmental descriptors collocated onto a common geospatial and

150    temporal grid. Further, users can execute pre-validated, managed models and either download

151    results for custom offline post-processing or utilise one of several pre-defined techniques to

152    analyse their results.

153

154    Urban Centre for Computation and Data (UrbanCCD) (UrbanCCD 2022) is a joint initiative at the

155    University of Chicago and Argonne National Laboratory, to support the study of urban science.

156    The UrbanCCD does not provide a dedicated computing facility, but researchers may make use

157    of the Argonne Leadership Computing Facility for batch computing.

158

159    JASMIN (Lawrence et al. 2013) is a globally unique data intensive supercomputer for

160    environmental science and currently supports over 1500 users on over 200 projects. JASMIN

161    users research topics ranging from earthquake detection and oceanography to air pollution and

162    climate science. JASMIN provides the UK and European climate and earth-system science

163    communities with the ability to access very large sets of environmental data, which are typically

164    too big to download and process using their own computers.. This rapidly reduces the time it

165    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 2, and briefly described below.
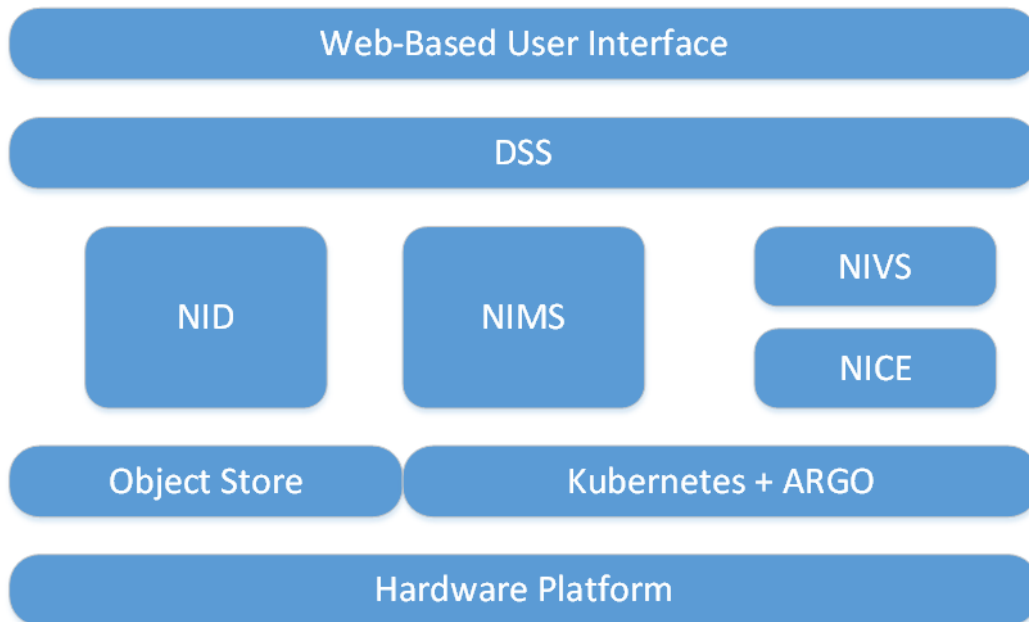


**Figure 2: DAFNI Core Components.**

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 as a Kubernetes cluster.  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

184        includes: a centrally managed data-store; a data catalogue; and a data access and

185        publication service.

186     • **National Infrastructure Modelling Service (NIMS).** The NIMS provides support to

187        improve performance of existing models, reduce the complexity of creating models and

188        facilitate the creation of multi-systems models. It includes a model catalogue and a

189        workflow creation and execution framework based on ARGO (see

190        https://argoproj.github.io/).

191     • **National Infrastructure Cloud Environment (NICE).** The NICE provides scalable

192        cloud environment with a number of Platform as a Service (PaaS) offerings, including

193        Jupyter notebooks (see https://jupyter.org/).  Currently this is used to within the internal

194        architecture of DAFNI, to deploy services within the cluster.

195     • **National Infrastructure Visualisation Suite (NIVS):** The NIVS supports visualisation

196        tools to facilitate understanding of data, models, outputs and translation of findings to

197        decision makers. This includes traditional visualisation as a service (e.g. graph and

198        tabular representations) and user developed analyses using Jupyter Notebooks.

199     • **DAFNI Security Service (DSS):** The DSS manages the security of the platform, which

200        allows users to seamlessly access and use services they have rights to, while at the

201        same time maintaining security and integrity of data. Services include authentication,

202        authorisation, monitoring, and group management.

203  These components have been implemented in a micro-services architecture. This allows the

204  capabilities within DAFNI to be independent with an extensible and flexible delivery of the

205  platform in line with the evolving nature of the National Infrastructure modelling landscape.

206

207  Following a structured design approach, a hierarchical overview of the platform has been

208  derived leveraging the capabilities and functions outlined as part of the core capabilities

209  analysis.  Two central components, the NID, and the NIMS, are discussed in more detail.


210  ## 3.2  The National Infrastructure Database

211

212   The National Infrastructure Database (NID) is the foundation of DAFNI, a core service that

213   allows researchers to upload, access and share datasets which are necessary to their research.

214   It then manages the provision of data to models, workflows and visualisations, with outputs from

215   model executions published back to the NID, allowing the research community access to the

216   latest model outputs.

217

218   The NID uses a MinIO object storage instance with a capacity of up to 900 terabytes. The

219   adoption of object storage allows DAFNI to be flexible and store any data in any format

220   required. MinIO provides a Cloud Native solution which integrates seamlessly into DAFNI's

221   underlying Kubernetes environment. This is supported by databases which store and manage

222   the metadata records for each dataset, providing Data Search and Data Versioning capabilities

223   around the data-store itself.

224

225   DAFNI researchers interact with the data store via the DAFNI Data Repository, illustrated in

226   Figure 3, a tailor-made repository service that allows researchers to upload data to the NID and

227   manage the access to that data, allowing others on the platform to access it either globally,

228   individually or through groups. In addition, researchers can update their datasets and create

229   new versions, and all registered users on the DAFNI platform can access and download the

230   open access datasets.

231



241
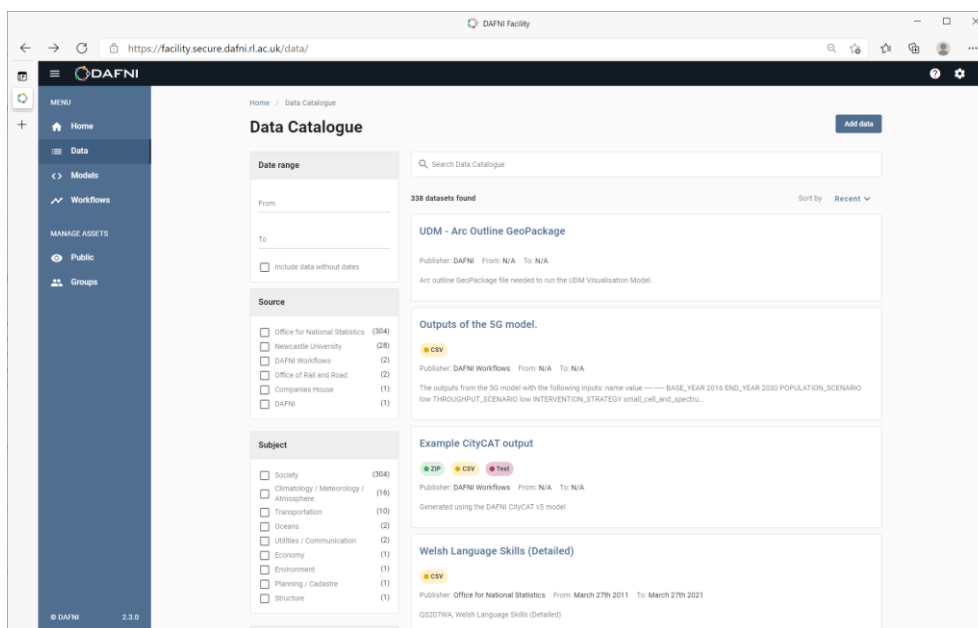
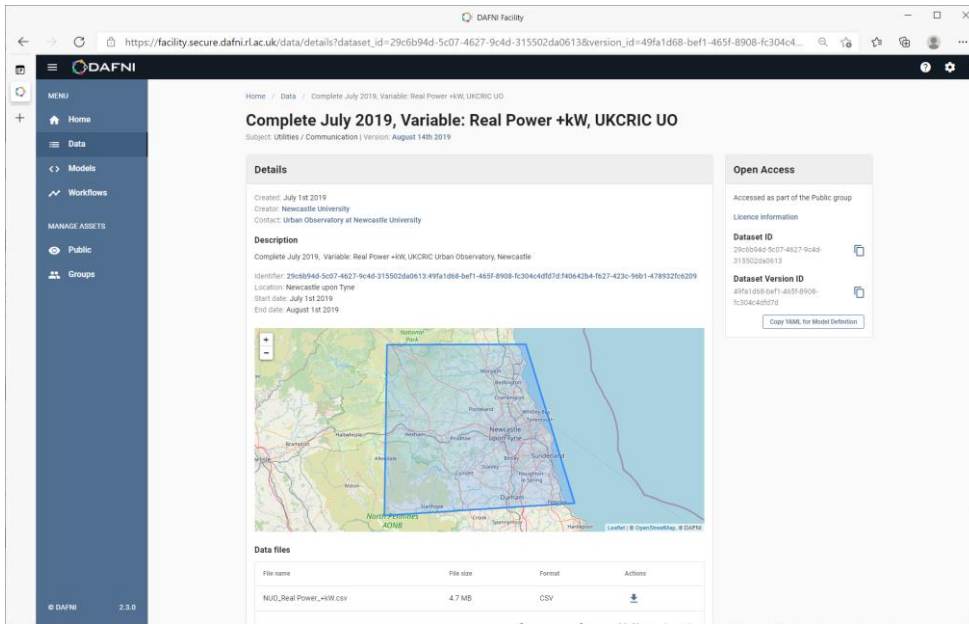**Figure 3: The DAFNI NID a). the data catalogue; b). entry for an example data set.**

**Metadata**. DAFNI has adopted a rich metadata schema, based primarily on DCAT V2 (World-Wide Web Consortium, 2020), a W3C recommendation for interoperability between data catalogues, augmented with additional features supporting geospatial information, 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

271 text search or by spatio-temporal filtering. Filters by data source, theme and file format are also

272 supported.


## 273  3.3  The National Infrastructure Modelling Service

274
275 The National Infrastructure Modelling Service (NIMS) encompasses both the model catalogue

276 and model workflow systems on DAFNI. The purpose of the NIMS is to allow DAFNI users to

277 run user supplied models through the use of workflows without specialised HPC knowledge.

278

279 **Containerisation**.  The execution of user generated models and their combination into multi-

280 systems models is challenging because of necessary interoperability between models. Each

281 model, developed by independent groups of researchers and software engineers, has a set of

282 dependencies on programming language, packages and libraries.  These dependencies make

283 porting models onto a common platform a complex and time consuming process, a significant

284 barrier to the use of high-performance computing.  Further, coupling models together requires

285 the sharing of data in interoperable formats and access to APIs for models to communicate.

286

287 To simplify these challenges, the DAFNI NIMS utilises containerisation using the Docker

288 packaging system (see https://www.docker.com) to encapsulate functionality and dependencies.

289 Docker builds self-contained packages encapsulating the model executable together with its

290 execution environment, also bundling configuration and libraries files. A model definition file in

291 YAML Ain't Markup Language (YAML see https://yaml.org/) format is also provided to

292 accompany the "dockerised" model, specifying the interfaces, input parameters and data sets,

293 and outputs to the model, together with metadata that will be displayed about the model

294 catalogue.  Dockerised models can then be uploaded onto the platform and can be deployed

295 and executed via the Kubernetes system.  Thus DAFNI can execute user code independent of

296 their dependencies.

297

298

299
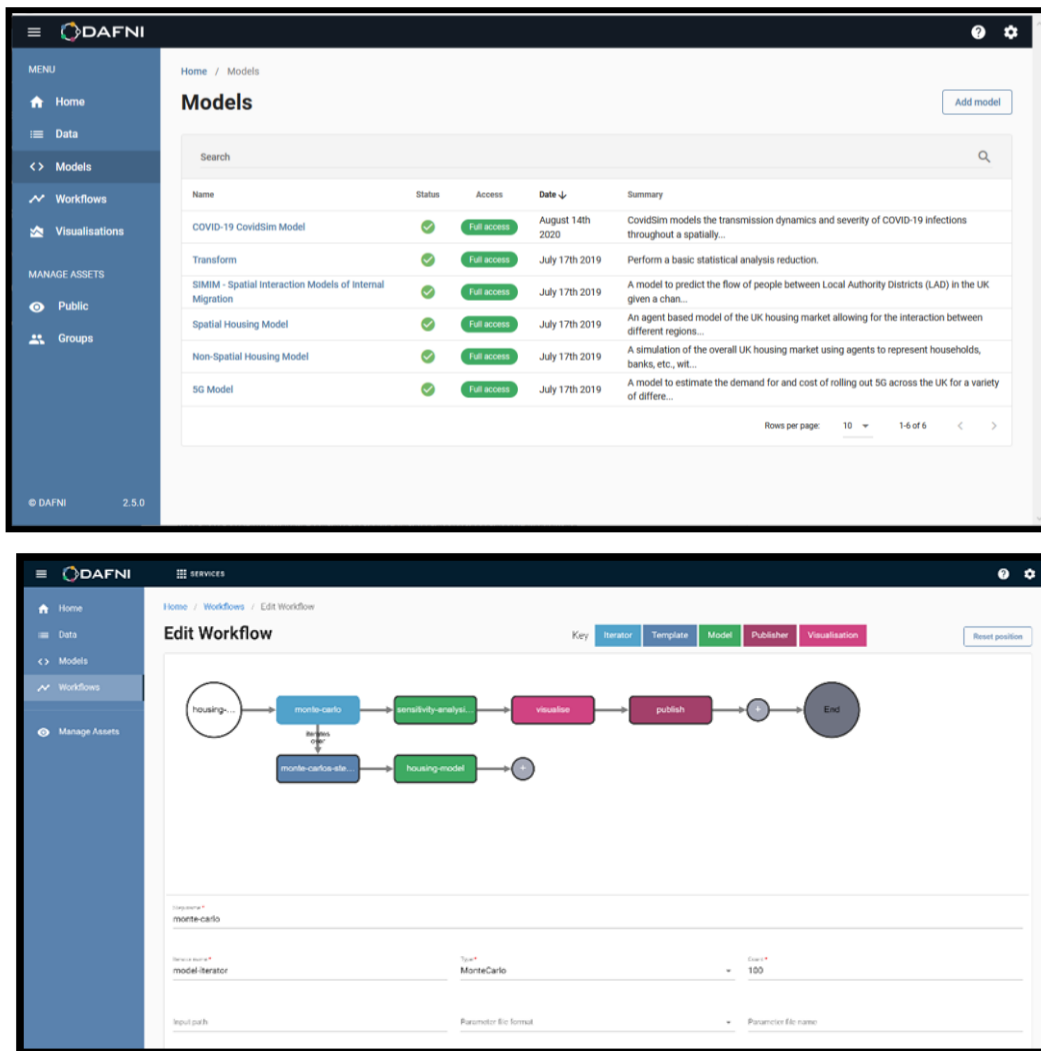
**Figure 4: The DAFNI NIMS a). The Model Catalogue; b). Building a workflow**

**Model Catalogue**. Models are uploaded into a Model Catalogue, illustrated in Figure 4a, a repository of models, based on the Harbor, an open-source system to provide 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 output the results of these workflows to share with other users. 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. On execution, the Kubernetes

330 orchestration engine allocated resources and deploys the workflow into "pods" across a number

331 of nodes in the cluster where each can be executed on their own resources.  This flexibility can

332 allow for more dynamic allocation of resources within DAFNI and allows for use of any operation

333 that can be containerised to be used within the workflows (e.g. data transformation and

334 visualisation).To build a workflow, users add a series of interconnected steps, as shown in

335 Figure 4b. The step types are described below.

336

337 *Model.*  The model step facilitates the execution of a model. Users can choose the model from

338 the model catalogue and set any input parameters for the model, with data selected from the

339 NID. Models can also be chained together in the workflow, passing output data from model into

340 the inputs of the next to allow for multi-systems modelling. For example, a model which

341 simulates population growth can be chained to a model which relied on population numbers to

342 predict house prices, thus allowing the exploration of the effect on house prices of different

343 demographic scenarios.

344

345 *Iterator.*   Iterators allow the same step in the workflow to be repeated multiple times whilst

346 changing parameters within a given range either randomly or with a pre-defined increment. This

347 allows multiple executions of the same model to be completed in parallel to one another where

348 possible, so many runs of the same model can be completed across a range of values, or

349 across random values in Monte-Carlo simulations where the same model can be run multiple

350 times with different parameters.

351

352 *Publisher.* The publisher step takes outputs from a model and ingests them into the NID. The

353 user supplies metadata about the resultant dataset which will be displayed in the data

354 catalogue.

355

356 *Visualisation.* The visualisation step takes the outputs of a model and creates a visualisation

357 builder containing those outputs using the NIVS. This allows the user to go directly from the

358 results of a finished workflow into generating graphs or charts from those results in a

359 visualisation builder, or via a user programmable Jupyter notebook.

# 4   Using DAFNI

The initial phase of the DAFNI construction programme (2017–2021) was detailed 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 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 (ITRC, see https://www.itrc.org.uk/ ) NISMOD system, a key example of where collaborative environments have been established 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 is a collection of codes that currently run 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-I 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

389    sensitivity analyses.  As a result, the NISMOD-1 team have successfully run a number of

390    sensitivity analyses on the Water Supply models and achieved a speed of up to 10 times faster

391    than the original service.  This demonstrates the benefits that can be derived by moving

392    existing, proven infrastructure models onto a high throughput cluster; moving the data as well as

393    the software to the DAFNI system is key to obtaining scalable performance.   The work on the

394    NISMOD pilot has continued through the development of NISMOD-2 and its implementation on

395    DAFNI as the platform has evolved.  Workflows supporting the NISMOD scenarios are now

396    available on DAFNI, which provides NISMOD users with a long-term execution environment.

397

398    Further projects are now using the DAFNI platform.  The Open Climate Impact (OpenClim see

399    https://gtr.ukri.org/projects?ref=NE/T013931/1 ) project is developing a modelling framework to

400    explore the impact of future climate change scenarios on infrastructure, exploring such factors

401    as flood events in urban environments, the effect of extreme heat events on the population, and

402    the effect on agriculture.  The project has particular emphasis on adapting the environment to

403    climate change, and the mitigating effects those adaptations might have.  DAFNI is being used

404    in the project as a common modelling framework to connect the different models and to provide

405    a legacy space so that the workflow can be accessed in the long term.  The Centre for Greening

406    Finance and Investment (see https://www.cgfi.ac.uk/) is also planning to use the DAFNI platform

407    similarly to host and develop a share data and modelling framework to explore how

408    environmental change will impact the risks on investment, insurance and other activities within

409    the finance industry.

## 410   5   Future developments

### 411   5.1 Enhanced Data Framework

412    Data remains the central driver for the future research and exploitation of computational models

413    of infrastructure systems, and richer handling of data would enhance the power and range of

414    DANFI for researchers. The following extensions and enhancements to the NID are being

415    explored.

416

417     ●  **FAIR Data publication and Data Curation.**  DAFNI supports a metadata description for

418        data and models, and thus partially satisfies the FAIR principles.  In order for DAFNI to

419        support reusable reference data within the research community, this needs to be enhanced

420        to support a data publication pipeline, underpinned by data curation processes to update

421        and maintain data for the long term.

422     ●  **Scaling and querying large data.**  The handling of large data sets within workflows can be

423        inefficient as data copying and transfer is a high-latency exercise, and frequently queries

424        are applied early in the workflow to extract the relevant data-slice suitable for processing.

425        Large-data immutable datasets can be treated as static objects which can be accessed in a

426        common manner across different processes, with data-slicing taking a "data-cube"

427        approach.

428     ●  **Interoperability Framework.**  The data and modelling framework in DAFNI has the

429        advantage of being generic and thus can accept data in any format.  However, in linking

430        models into workflows, there still remains the need to undertake data manipulation tasks,

431        such as queries, format transformation, projections between scales and other data

432        transformations.  By providing enhanced support for particular data formats and providing a

433        suite of "data adaptors" or "transforms" in an interoperability framework, the process of data

434        manipulation can be simplified.

435     ●  **Semantic framework.**  A further extension to the interoperability framework would be to

436        introduce an ontology framework.  By supporting a selected suite of ontologies, rich data

437        enhanced mappings can be supported within workflows as well as enhancing the search

438        and discovery service.  As part of the DAFNI Champions programme, a report on the

439        selection and use of suitable ontologies with recommendation for future development

440        (Varga et al. 2021).


441 ## 5.2  DAFNI and Digital Twins

442

443 The concept of Digital Twins (Batty, 2018; Callcut, 2021) has emerged over the last decade as a

444 key technology for the future planning delivery and operation of infrastructure systems.  There

445 has been a high level of interest from government and industry in investing in Digital Twins as a

446    tool to predict, optimise and control the outcomes infrastructure investment.  Initiatives such as

447    the Digital Twin Hub (see https://digitaltwinhub.co.uk/) have been developing framework for

448    combining digital twin models into a 'National Digital Twin' (NDT), "a digital model of our national

449    infrastructure which will be able both to monitor our infrastructure in real-time, and to simulate

450    the impacts of possible events" (National Infrastructure Commission, 2017), via the sharing of

451    data and computational resources into a common digital twin ecosystem.    DAFNI can play a

452    role in the development and deployment of Digital Twins for infrastructure systems by providing

453    support for features that a national digital twin ecosystem would require to be effective.

454

455    A NDT would require an ecosystem of models from a wide range of sources, which can be

456    combined into large-scale, multi-system digital twins.  Running multiple twins at scale would

457    require high-performance computing platform which allows models to be executed rapidly,

458    scaled up in resolution, and geographical and temporal range.   The platform independent

459    approach of DAFNI offers the basis of such an environment.  Further, a NDT would need to

460    support the combination of models into new workflows to support connected digital twins, and

461    provide visualisations of results for human decision support, again supported within DAFNI.

462

463    Digtial Twins also bring significant challenges for data management and integration, and a NDT

464    would require a wide range of different data sources to be brought together into a shared trusted

465    data space, in a common information management framework.  For the sustainability of the

466    NDT, this would need to be maintained and curated for the long-term.  Again DAFNI already

467    offers the NID, which could form the basis of such a data space.

468

469    Thus, DAFNI could provide a hub for a digital twin infrastructure, supporting the research and

470    development required to explore the opportunities of deploying digital twins within infrastructure

471    systems development and operation.

472

473    DAFNI is working with initiatives such as the UKCRIC Urban Observatory programme (see

474    https://urbanobservatory.ac.uk) and interacting with key stakeholders, including the Centre for

475    Digital Built Britain (see https://www.cdbb.cam.ac.uk/).  It is developing some pilots DTs,

476 including on traffic management with Sheffield University.  Further development on DAFNI is

477 exploring how to provide additional functionality to support Digital Twins, including extending the

478 information management framework in DAFNI, as discussed above, interacting with real-time

479 input and streaming data systems, and working with Machine Learning to form adaptive models

480 for decision making from historic data.

481 # 6    Conclusions.

482 DAFNI is an infrastructure platform to support the development of sharing of multi-systems

483 models of national infrastructure.    The data and model sharing allows access to models across

484 infrastructure-systems modelling, and across collaborations.  DAFNI thus offers the

485 infrastructure systems engineering community a space to leverage their research into wider and

486 deeper applications.

487

488 Further, DAFNI also has the potential to support collaborations between researchers,

489 government and industry.  Data and computation is seen as central to the infrastructure

490 engineering practise, and as the NIC asserts "Data is part of infrastructure and needs

491 maintenance in the same way that physical infrastructure needs maintenance" (National

492 Infrastructure Commission, 2017).  DAFNI provides the basis for a trusted, common, vendor-

493 neutral hub for data sharing and exchange with support for maintaining the value of these

494 assets for the longer term.  Further, it is recognised that while large-scale computing is valuable

495 to solve new business and research challenges, it remains hard to access and use for non-

496 specialists (see for example Government Office for Science, 2021).  DAFNI provides a user

497 environment that seeks to overcome some of these technical barriers.

498

499 DAFNI has transitioned from a development project to a service phase.  This phase will enable

500 DAFNI's operational growth to increase usage and capability to support research in EPSRC's

501 Engineering Programme and related fields, so that the UK's national infrastructure research

502 effort can remain at the cutting edge. It also allows DAFNI to build on its commitment to

503 supporting changing and sustainable infrastructure needs through working with government and

504 industry.

505

513

514 **References**

515

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

518

519 Batty M. (2018) Digital twins. *Environment and Planning B: Urban Analytics and City Science.*
520 45(5):817-820. https://doi.org/10.1177/2399808318796416

521

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

524

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

528

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

531

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

536

537    Dobson B, Coxon G, Freer J, Gavin H, Mortazavi-Naeini M and Hall JW (2020). The spatial

538    dynamics of droughts and water scarcity in England and Wales, Water Resources Research, 56

539    (9): e2020WR027187. https://doi.org/10.1029/2020WR027187 .

540

541    Ford A, Barr S, Dawson R, Virgo J, Batty M, Hall J. (2019) A multi-scale urban integrated

542    assessment framework for climate change studies: a flooding application. Computers,

543    Environment, and Urban Systems, 75, 229-243.

544    https://doi.org/10.1016/j.compenvurbsys.2019.02.005

545

546    Global Infrastructure Hub, (2017). Global Infrastructure Outlook: 2017.

547    https://www.oxfordeconomics.com/recent-releases/99f4fa86-a314-4762-97c6-fac8bdcbe40a

548    (accessed 24 January 2022)

549

550    Government Office for Science (2021). Large-scale computing: the case for greater UK

551    coordination.

552    https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/fi

553    le/1018875/UK_Computing_report_-_Final_20.09.21.pdf (accessed 24 January 2022)

554

555    Hall JW, Tran M, Hickford AJ and Nicholls RJ. (eds.) (2016). The Future of National

556    Infrastructure: A System of Systems Approach, Cambridge University Press. ISBN: 978-1-

557    10758-874-5

558

559    Hall JW. Thacker S, Ives MC, Cao,Y, Chaudry M, Blainey SP and Oughton E.J (2017). Strategic

560    analysis of the future of national infrastructure, Proceedings of the Institution of Civil Engineers:

561    Civil Engineering, 170(1): 39-47. https://doi.org/10.1680/jcien.16.00018

562

563    Hall, J.W. (2019) Using system-of-systems modelling and simulation to inform sustainable

564    infrastructure choices. IEEE Systems, Man and Cybernetics Magazine, 5(3): 10-20.

565    https://doi.org/10.1109/MSMC.2019.2913565.

566

567    Hallgren W, Beaumont L, Bowness A, Chambers L, Graham E, Holewa H, Laffen S, Mackey B,

568    Nix H, Price J, Vanderwal J, Warren R and Weis G. (2016). The Biodiversity and Climate

569    Change Virtual Laboratory: Where ecology meets big data. Environmental Modelling &

570    Software, vol. 76, pp. 182-186, https://doi.org/10.1016/j.envsoft.2015.10.0251364

571

572    Horni, A., Nagel, K. and Axhausen, K.W. (eds.) (2016) The Multi-Agent Transport Simulation

573    MATSim. London: Ubiquity Press. https://doi.org/10.5334/baw.

574

575    INSPIRE (2022). Topic categories in accordance with EN ISO 19115.

576    https://inspire.ec.europa.eu/metadata-codelist/TopicCategory (accessed 24 January 2022)

577

578    ITRC Mistral (2020).  A sustainable Oxford-Cambridge corridor? Spatial analysis of options and

579    futures for the Arc https://www.itrc.org.uk/wp-content/uploads/2020/01/arc-main-report.pdf

580    (accessed 24 January 2022)

581

582    Lawrence B N, Bennett V L, Churchill J, Juckes M., Kershaw P,  Pascoe S, Pepler S,. Pritchard

583    M. and Stephens, A. (2013) Storing and manipulating environmental big data with JASMIN,

584    *2013 IEEE International Conference on Big Data*, 2013, pp. 68-75,

585    https://doi.org/10.1109/BigData.2013.6691556 .

586

587    Lomax N, and Smith AP, (2020). DAFNI Pilot 4: SPENSER - Synthetic Population Estimation

588    and Scenario Projection model. https://dafni.ac.uk/wp-content/uploads/2020/05/dafni-pilot-4-

589    dafni-hosts-population-forecast-model.pdf (accessed 24 January 2022)

590

591    National Infrastructure Commission (2017). Data for the Public Good. NIC, London, UK

592    https://nic.org.uk/app/uploads/Data-for-the-Public-Good-NIC-Report.pdf (accessed 24 January

593    2022)

594

595    Pianosi F, Rougier J, Freer J, Hall JW, Stephenson DB, Beven K, Wagener T (2016). Sensitivity

596    analysis of environmental models: a systematic review with practical work. Environmental

597    Modelling and Software, 79:214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

598

599    Schooling J, Enzer M and Broo DG (2021).  Flourishing systems: re-envisioning infrastructure

600    as a platform for human flourishing. Proceedings of the Institution of Civil Engineers – Smart

601    Infrastructure and Construction, https://doi.org/10.1680/jsmic.20.00023

602

603    Thacker S, Adshead D, Fay M, Hallegatte S, Harvey M, Meller H, O'Regan N, Rozenberg J and

604    Hall JW (2019). Infrastructure for Sustainable Development, Nature Sustainability, 2 324–331.

605    https://doi.org/10.1038/s41893-019-0256-8 .

606

607    Urban Center for Computations and Data (2022) About UrbanCCD. see

608    http://www.urbanccd.org/about/#about-urbanccd (accessed 24 January 2022)

609

610    Varga L, McMillan L, Hallett S, Russell T, Smith T, Truckell I, Postnikov A, Rodger S,  Vizcaino

611    N, Perkins B, Matthews B, Lomax N (2021). Infrastructure Research Ontologies.

612    http://dafni.ac.uk/wp-content/uploads/2021/05/IRO-final-report-31-03-2021.pdf (accessed 24

613    January 2022)

614

615 Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* (2016). The FAIR Guiding Principles for

616 scientific data management and stewardship. *Sci Data* **3,** 160018

617 https://doi.org/10.1038/sdata.2016.18

618

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

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

621

622 Young MA & Blainey SP (2018) Developing railway station choice models to improve rail

623 industry demand models, Transportation Planning and Technology 41(1):80-103.

624 https://doi.org/10.1080/03081060.2018.1403745

625

626 Young MA, Blainey SP, Gowland T & Nagella S (2019).  An Automated Online Tool to Forecast

627 Demand for New Railway Stations and Analyse Potential Abstraction Effects. Transport

628 Practitioners' Meeting, 10-11 July 2019, Oxford. https://eprints.soton.ac.uk/432493/ (accessed

629 24 January 2022)

630