Ref. Ares(2024)9255791 - 27/12/2024

CrossEU

D1.4 – CROSSEU IAF and DAFNI functionalities (Version 1)

WP1 - Task 1.4

December 2024

Funded by the European Union

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

UK participants in this project are co-funded by UK Research and Innovation

Project Acronym	CROSSEU
Project Name	Cross-sectoral Framework for Socio-Economic Resilience to Climate Change and Extreme Events in Europe
Grant Agreement Number	101081377
Project Coordinator	Administrația Națională de Meteorologie R.A. (MeteoRo)
Project Duration	January 2024 – December 2026
Website	www.crosseu.eu

Deliverable No / Milestone No	D1.4
Dissemination Level	Public
Work Package	WP1
Lead beneficiary	United Kingdom Research and Innovation (UKRI)
Author(s)	UKRI : Tom Kirkham, Brian Matthews, Esther Turner, Teagan Zoldoske
Reviewed by	MeteoRo: Sorin Cheval, Angela Cristina Plescan, Laurentiu Ciuhu, Vlad Amihaesei Hereon: Paul Bowyer K&I: Gabriele Quinti
Date	December 2024
File Name	CROSSEU_ D1.4-CROSSEU IAF and DAFNI functionalities (Version 1) _V01_27122024_UKRI

Table of Contents

List of Figures

List of Tables

Executive Summary

The CROSSEU Integrated Assessment Framework (IAF) is the point at which multisector risk assessments for the projects use cases take place. The IAF will enable the project's teams to collaborate on the development of workflows combining models and data to assess risk and address specific challenges presented within the projects use cases.

The CROSSEU IAF is built on the DAFNI platform which offers a researcher focused tool capable of supporting data and modelling requirements from multiple domains. The deliverable explains how researchers can use DAFNI to publish, discover and use models and data within workflows to assess climate risk.

This deliverable presents Version 1 of the CROSSEU Integrated Assessment Framework (IAF). It describes the foundations of the IAF, how it is used and the additional work that is taking place to better integrate the CROSSEU modelling requirements into the platform. This will continue to evolve as the requirements of CROSSEU continue to develop.

Keywords

Computational Models, Data, Platform, High Performance Computing, Security, Workflow, Metadata, Risk, Data Analysis, Climate

Abbreviations and acronyms

Acronym	Description
AI	Artificial Intelligence
AgMIP	Agricultural Model Intercomparison and Improvement Project
ARBM	Adaptive Resilience-Based management
BC-IAM	Benefit-Cost – Integrated Assessment Model
CC	Climate Change
CAP	Common Agricultural Policy
CCAP	Climate Change Action Plan
CGE	Computable General Equilibrium model

CCH	Climate Change Hotspot
CS	Case study
CSA	Case study area
DDM	Demographic Distribution Models
DICE	Dynamic Integrated Climate Economy
CSS	Decision Support System
EC	European Commission
ECA&D	European Climate Assessment and Dataset
ECB	European Central Bank
ECHO	European Climate and Health Organisation
ECMWF	European Centre for Medium-Range Weather
EDG	European Green Deal
EEA	European Environmental Agency
EIOPA	European Insurance and Occupational Pensions Authority's dashboard on the insurance protection gap for natural catastrophes
ENGAGE	Environmental Global Applied General Equilibrium model
ERDB	European Bank for Reconstruction and Development
FOLU	Food and Land Use Coalition
FUND	Climate Framework for Uncertainty, Negotiation, and Distribution
GDP	Gross Domestic Product
GHG	Green House Gases
GIS	Geographical Information System
HDR	Harmonised Data Repository

IAF	Integrated Assessment Framework
IAM	Integrated Assessment Model
IGO	Intergovernmental Organisation
IPCC	Intergovernmental Panel on Climate Change
IRENA	International Renewable Energy Agency
ISIMIP	Inter-Sectorial Impact Model Intercomparison Project
JTF	Just Transition Fund
JRC	Joint Research Center
LCA	Life cycle assessment
ML	Machine Learning
NCME	NatCatModelling Engine
NGO	Non-Governmental Organisation
M&A	Mitigation and Adaptation
MAES	Mapping and Assessment of Ecosystems and their Services
P-IAM	Process – Integrated Assessment Model
PAGE	Policy Analysis of the Greenhouse Effect
Plan4EU	European Optimal Electricity System Management Tool
POLES2	Prospective Outlook for the Long-term Energy System model
R&I	Research and Innovation
RCM	Regional Climate Models
SCC	Social Cost of Carbon
SDG	Sustainable Development Goal
SCSS	Spatial Decision Support System

SE	Socioeconomic
SoA	State of the art
STL	Storyline
UNCCD	United Nations Convention to Combat Desertification
UNDP	United Nations Development Programme
UNEP	United Nations Environment Programme
UNHCR	United Nations High Commissioner for Refugees
UNFCCC	United Nations Framework Convention on Climate Change
WAREG	European Water Regulators
WHO	World Health Organisation
WMO	World Meteorological Organisation

Introduction

The objective of the CROSSEU project is to equip stakeholders with the necessary tools and insights to make well-informed decisions based on the assessment of the risk arising from the hazards associated with climate change on different socio-economic sectors, adapt to those climate risks, and reduce their impacts. Deliverable D1.1 (CROSSEU) on the CROSSEU methodology developed the requirements and expertise from multiple sectors converge to present the socio-economic risks of climate change, giving an overview of the co-design and development process as illustrated in Figure 1.

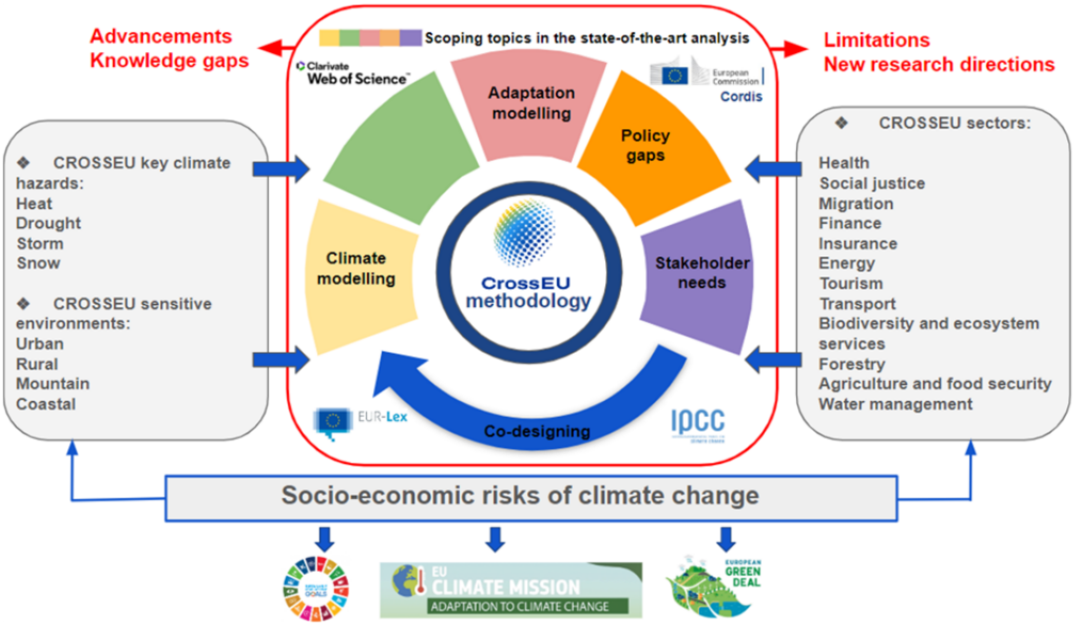

Figure 1: Scoping analysis for the co-design and development of the CROSSEU methodology.

Within CROSSEU, researchers collaborate and share data, knowledge and expertise. Within the project this is enabled by a digital platform that forms a common environment for collaboration between researchers to support decision making. This digital platform forms the foundation of an Integrated Assessment Framework (IAF) for CROSSEU to collaborate across different research groups and applications, to share data and models, to develop integrated workflows within a common data framework, and place to make available results data for sharing with relevant stakeholders and passing onto related systems, including the CROSSEU Decision Support System.

Computationally good decision making is driven by the combination of Data and Models which generate knowledge and data driven insight on which they can be based. Within the project Data and Models are hosted

Funded by the European Union

or linked into the IAF from other platforms. The CROSSEU IAF is delivered by UKRI building on the UKRI DAFNI platform.

A major component of the IAF is the Harmonised Data Repository (HDR). This collects the key data sources which CROSSEU use cases require and makes them available as a collection for access. This should be with a harmonised data framework that allows flexibility of using data standards from across the application domains, also support a common data and metadata framework to enable data interoperability. The HDR represents a significant part of the work of the IAF, and is thus treated within a separate deliverable, D1.6.

This deliverable:

- presents the high-level objectives and technical requirements which the CROSSEU IAF should seek to satisfy;
- outlines the key functionalities within the CROSSEU IAF and how they can be supported within the DAFNI platform;
- proposes a methodology for populating and using the IAF.

We conclude by considering the next steps and future development of the IAF within the project.

1. IAF Objectives

CROSSEU has selected 6 case studies focus on four key climate hazard categories: storms, heatwaves, droughts, and snow, together with two others considering cross-sectoral multi-hazard risks and indirect impact analysis; each case study considers one or more demonstration of risk assessment on a socio-economic sector: agriculture, biodiversity, energy, food security, forestry, health, tourism, transport, water supply, migration, social justice, finance, and insurance. These represent a non-exclusive set of sectors which are all significantly affected by climate change, and where tractable approaches are available. These sectors will be broadly covered by a set of relevant data and models to explore the sectoral risks.

CROSSEU also considers multi-hazard risks combining hazards and the effect on multiple sectors. on the interaction between the themes. Traditional modelling approaches focus on a single feature/theme, while in the real-world different systems interact with one another, e.g., the impacts of future flooding events are dependent on the projected state of the urban development. The modelling framework thus needed to consider these dependencies, using a systems-of-systems approach (J. W. Hall, 2016; C. M. Little 2019) allowing models developed independently to interoperate with each other.

As well as the baseline risks, different adaptation strategies need to be considered. These can take different forms by focusing on lowering the severity of the hazard or reducing the vulnerability and/or exposure. The climate change impact would then be reassessed with these adaptations in place.

Thus, for a complete risk assessment many complex scenarios are formulated and executed representing the multi-dimensional combination of factors for each theme of interest. Managing this large number of executions and their resulting data sets in a transparent, verifiable and reproducible manner represent a major challenge. Further, this system is accessible to a range of stakeholders so that they can access the results and potentially review and rerun analyses.

The CROSSEU IAF provides an open, innovative and flexible integrated data and computational environment to support as required local, national, regional (e.g., Scandinavia, or the Danube catchment) and European-scale computational modelling-based analyses to assess the socio-economic risks or climate change in different sectors, and assess policy responses and adaptation plans to mitigate those risks.

To handle the complexity of analysis, CROSSEU is designed to develop a modelling framework formed by a collection of data, models and workflows presented together within a computational platform and made available to the users of the CROSSEU results. The modelling framework has the several requirements described in the following paragraphs.

In particular, the IAF should support the following objectives.

- Independent Development and Execution. A diversity of models is being developed independently by the CROSSEU research teams working to their own requirements and implementation standards. The common CROSSEU IAF should accommodate the different standalone models and allow them to be executed independently of other models.

- Flexible yet interoperable data environment. CROSSEU considers a diverse set of sectors: agriculture, biodiversity, energy, food security, forestry, health, tourism, transport, water supply, migration, social justice, finance, and insurance integrated with environmental and socio-economic analysis. Each of these sectors applies its own data standards and computational modelling methods; nevertheless, there is the need to share input and results data sets to support a range of climate, socio-economic and adaptation scenarios. However, there are common data foundations (e.g., Geospatial and Environmental data).

  To assess a comprehensive range of sectors within the platform, including relevant interactions between sectors, the IAF needs to retain flexibility and adaptability rather than impose a narrow range of data formats, data dictionaries, and parameters. The metadata should be sufficiently general yet allow sector specific elements to be represented and accessed.

- Multi-systems modelling. The diverse models arising from different sources within the CROSSEU IAF can be linked and executed together within a common data framework in a flexible and meaningful manner within an open modelling environment to support complex, reproducible workflows, exploring the interaction of models of different systems.

- Multiple analyses. To assess different warming scenarios across different regions and to evaluate different policy options across multiple scales so that risks from climate change can be evaluated consistently across a range of socioeconomic futures, model workflows need to be executed repeatedly. The IAF should accommodate the repeated execution of workflows across different parameter options, preferably at set up without repeated user intervention.

- Accessibility to Stakeholders. The IAF should be accessible to stakeholders outside the immediate CROSSEU Consortium to review the analyses undertaken, and to access results data. Further, results data should be accessible to the CROSSEU DSS. Results data should be supplied with provenance information so that the methods used to generate results are transparent and auditable.

- Legacy. The IAF should create a sustainable legacy via a library of results and to better inform future risk assessments. The data, models and workflows should be available after the project to support the continued access to results, with the potential to rerun workflows on updated parameters, and to add further models and workflows to the system. The IAF would have the potential to become an open ecosystem to support continued climate risk assessment, given a suitable sustainability approach.

The CROSSEU IAF is designed to enhance assessment of risk and increase the range of questions that can be posed. It allows interactions between sectors to be considered. For example, it enables better exploration of how changes in land-use can act as an adaptation to climate change, especially in the context of carbon sequestration efforts and land-use changes, potentially required to meet the UK government's policy target of net-zero emissions by 2050.

As the project progresses IAF will develop further incorporating requirements from the CROSSEU community to better support the project.

2. IAF Technical Requirements Supporting Climate and Sector Specific Research

Combining climate data with sector specific research communities presents a wide range of challenges to researchers. Across all sectors models and data have increased in size and complexity; these challenges the compute resources of individual researchers and large-scale resources are costly and can be technically difficult to navigate.

In addition, within some sectors data and models can be difficult to find and access, they are also subject to different licensing agreements and may be commercially or societally sensitive, motivating a common approach to data security.

As well as being difficult to find, data can be hard to share with distributed colleagues; this leads to analysis work taking place by individual institutions and a reduction of vital collaboration and knowledge-sharing. For models to reflect real-world situations more accurately, there is a need for them to capture the interactions between systems in multi-systems models.

The variety and variability of these models presents a significant challenge, as extensive sector expertise is required to exploit each model. Thus, a collaborative platform is needed that can handle models and data. Furthermore, in the context of the CROSSEU IAF a common data integration framework is needed for a flexible multi-system modelling approach. To ensure that results are reliable and repeatable makes it essential to store versioned copies of the underlying data sets, with auditable provenance of results.

Initially there are four main technical requirements of the CROSSEU IAF:

- Independent Development and Execution: Models can be developed independently by research teams within their own requirements and implementation standards and integrated into a common CROSSEU IAF as standalone models and executed independently of other models.
- Systems of Systems: Models within the CROSSEU IAF can be executed as part of a wider workflow consisting of multiple models from the IAF. Thus, delivering a systems-of-systems approach within the IAF.
- Containerisation: A common approach to containerisation of models to enable deployment on the IAF is required, supplemented by documentation and support to end users.

Funded by the European Union

- Integration with External Data Sources: Not all data that models use will be hosted on the platform due to reasons that can include security, volume of data etc. The IAF will integrate with these data source to aid the execution of models on the IAF that use them.

These four objectives state the ambition of the project with respect to the CROSSEU IAF. Further breakdown of these objectives is defined later in the deliverable as an initial set of IAF Requirements which will shape the use of the DAFNI platform in the project.

Expanding the four IAF objectives in order to deliver the CROSSEU on the DAFNI platform, the technical requirements have been captured as in Table 1.

Table 1: Initial CROSSEU IAF Requirements

ID	Requirement Name	Description	Validation
1	CROSSEU Model Deployment	CROSSEU models identified and deployed on the CROSSEU IAF	CROSSEU models are available for use on the CROSSEU IAF
2	CROSSEU Data Access	Data from CROSSEU partners and other sources deployed to the CROSSEU IAF or integrated via APIs	CROSSEU models on the IAF can access required data sources.
3	CROSSEU User Management	Federated identity management enabled on the IAF to support CROSSEU community access	Users can logon to the platform and gain access to models / data they have the rights to use.
4	Support for Collaborative Workflows	Partners from CROSSEU can use the IAF to design and test workflows consisting of multiple models	Multi-model CROSSEU workflows in use on the platform.
5	CROSSEU Compute	Adequate compute resources are available on the IAF to support	Complex models and workflows consisting of multiple models required on

		execution of IAF workflows	the projects can execute efficiently on the platform
6	CROSSEU models support in containers	Enable CROSSEU models to function from containers (i.e., license call outs etc.)	CROSSEU model owners can deploy to the IAF and
7	User Training	Materials to aid user training i.e., deploying models to the platform	Creation of online videos and documents to support users in learning about the IAF.
8	Third party User Integration	Access to parties external to CROSSEU on the platform	Demonstration of external user onboarding process, access policy and identity management / authentication steps.

The requirements of the IAF will be updated as more models are proposed to the IAF along with associated use cases. In the current state, the requirements cover the main steps to deploying models to the IAF and ensuring they execute under different conditions.

Funded by
the European Union

3. Building the IAF on DAFNI

The Data and Analytics Facility for National Infrastructure (DAFNI) platform will provide the foundation on which the CROSSEU IAF will be built to address the core requirements / objectives of CROSSEU with respect to the IAF. DAFNI provides an established platform which can support collaboration between researchers on the project at an early stage to deliver the initial set of requirements.

Within the domain of climate change existing tools and frameworks exist for assessing risk and sharing data and models. For example, LandCaRe DSS is an interactive decision support system for climate change impact assessment and the analysis of potential agricultural land use adaptation strategies (https://www.landcare-ggmbh.de/). In addition, the decision support tool ADAPT2CLIMA (https://tool.adapt2clima.eu/en/home/) aims to improve understanding of climate change and its impacts on agriculture, aiding farmers, policy makers, agronomists, and the agribusiness industry in adaptation planning.

DAFNI has been chosen over these and similar platforms and frameworks due to its ability to link into existing frameworks and support a wide variety of models and data. DAFNI has been founded from a computational model into Digital Twin perspective to simulate real world scenarios. Models and supporting evidential data have increased in size and complexity. This challenges the computing resources available to individual researchers, DAFNI utilising high-performance computing aims to reduce the cost of access to these resources for researchers both financially and in terms of skills required.

Data and models can be difficult to find and access, and are subject to different licensing agreements and may be commercially or societally sensitive, motivating a common approach to data security. Data can also be hard to share openly due to licensing or access issues; this leads to analysis work taking place within individual institutions, reducing collaboration and knowledge-sharing. If models are to reflect the real-world more accurately, there is a need for them to capture interactions within multi-systems models.

The variety of these models presents a significant challenge, as extensive sector expertise is required to exploit each model. Thus, a common data integration framework is needed for a flexible multi-system modelling. The need to ensure that results are reliable and repeatable makes it essential to store versioned copies of the underlying data sets, with auditable provenance of results. The choice of DAFNI was made as it offers a high-performance platform that can support a wide variety of data / model domains and skill levels.

DAFNI (Matthews et. al. 2023) provides a computational modelling platform dedicated to infrastructure systems research. DAFNI provides a high-throughput resource that allows computational models developed independently to be uploaded, executed and shared by research scientists. DAFNI also provides a self-service data repository to upload and access data. DAFNI provides a flexible and user-friendly workflow engine to users, allowing models to be executed either alone or in combination with others, enabling complex system of systems analysis from different models and data. DAFNI also provides a self-service data repository to upload and access data, as well as a library of user-created visualisations.

A diversity of approach lies at the heart of the infrastructure research, with different groups independently developing models to capture the state-of-the-art. However, the need to collaborate means that these different models need to interoperate within the same computing environment. DAFNI uses software containerisation techniques and an object store to provide a service which allows users to upload the executables of their code into a common environment which is agnostic of code language or data format. This format-agnosticism allows researchers to write code in their preferred format and couple this with other models written using different formats and approaches.

Supporting research collaborations is a core ambition of DAFNI and the platform supports this by providing "group areas" for researchers to pool and couple resources, without making them public to the wider platform. To support the diversity of approach that lies at the heart of the infrastructure research community, DAFNI uses containerisation and an object store to provide a service which is agnostic of code language choice or data format.

This format-agnosticism allows researchers to write code in their preferred format and couple this with other models written using completely different formats and approaches. The DAFNI Platform retains models, data and visualisations beyond the lifetime of a given project, providing a central legacy resource which is accessible to all infrastructure systems researchers.

Additionally, DAFNI provides a flexible and user-friendly workflow engine to users, allowing the containerised models to be deployed executed either alone or in combination with others, enabling complex system of systems workflows to be created from different models and data. DAFNI allows user models to be executed within a workflow engine, with workflows built using a graph representation of the model execution sequence and defines which outputs are passed to downstream models. Input files for individual models in a workflow can be selected from DAFNI's data catalogue and suites of input parameters can be defined; these can be changed for each execution of a workflow, allowing users to build a set of runs and their outputs. The details of these runs are retained in the workflow record, providing the provenance of the output.

DAFNI services are bound together by a generalised user front end meaning that no coding experience is required to run models on the platform. This generalisation allows users with different levels of technical background to upload assets and manage their work on the platform, thus supporting groups with different levels of software engineering expertise. In addition, to facilitate more technical users of the platform, DAFNI also supports a Linux-based command line interface and a Python package.

There is, however, considerable diversity in requirements for computational modelling and a single uniform platform may not satisfy all the needs. For example, DAFNI currently has a relatively small hardware base which limits its capacity. Further, reference data sets maybe curated by a variety of external data sources, and the IAF should not seek to replicate that work unnecessarily, storing locally only the data which is needs.

As a consequence, the IAF should also allow the use and linkage to external data sources and compute platforms. As an example, DAFNI has already coordinated with the JASMIN system (Lawrence et al. 2013), a globally unique data intensive supercomputer for environmental science and currently supports over 1500 users on over 200 projects. JASMIN provides the UK and European climate and earth-system science communities with the ability to access very large sets of environmental data and high compute capacity. The links between DAFNI and JASMIN are currently planned to be strengthened. The IAF should build on this to extend to interoperation with other appropriate computing infrastructure.

Similarly, reference datasets are held at dedicated data centres such as the UK's Centre for Atmospheric Data Analysis (CEDA), including significant data sets such as CMIP. The IAF should develop mechanisms to access, query, and download relevant data into its own harmonised data repository as needed.

Supporting research collaborations is a core of DAFNI and the platform supports this by providing "group areas" for researchers to pool and couple resources with collaborators, without making them public to the wider platform. All assets uploaded to or created on the DAFNI platform are private by default, and can be shared with the group or made public on request and become available to all users of the platform. The DAFNI Platform retains models, data and visualisations beyond the lifetime of a given project, providing a legacy resource and maintaining the traceability of the provenance of results.

A significant previous project has demonstrated the value of the DAFNI platform to support multi-scenario climate risk assessment. The UK funded project OpenCLIM (Open Climate Impacts Modelling framework) project (Nicholls et al., 2024) took an integrated assessment approach and implemented it using the computing platform provided by DAFNI into an assessment framework, bringing together models, data and sector processes to facilitate harmonised, geospatial assessments of climate change risks across the UK in the context of socioeconomic change and mitigation targets (e.g., Net Zero for greenhouse gas emissions by 2050). OpenCLIM enables a comparison of the efficacy of various options for adapting to these risks, and includes explicit assessment of interactions between risks, as well as uncertainty analyses. This process is designed to inform the urgency scoring inherent to the UK's Climate Change Risk Assessment adaptation prioritisation and future National Adaptation Plans by analysing priority risks in a consistent, spatially explicit and systematic way, enabling research insights (e.g., interdependence between climatic risks in the UK), and supporting the generation of policy relevant information. We propose to bring a similar approach of the OpenCLIM implementation and the experience of its development and support to the CROSSEU project.

4. IAF Core Functions and Components

Built upon DAFNI the IAF architecture incorporates DAFNI at its core. The main initial architectural components of the IAF can be seen in the diagram below.

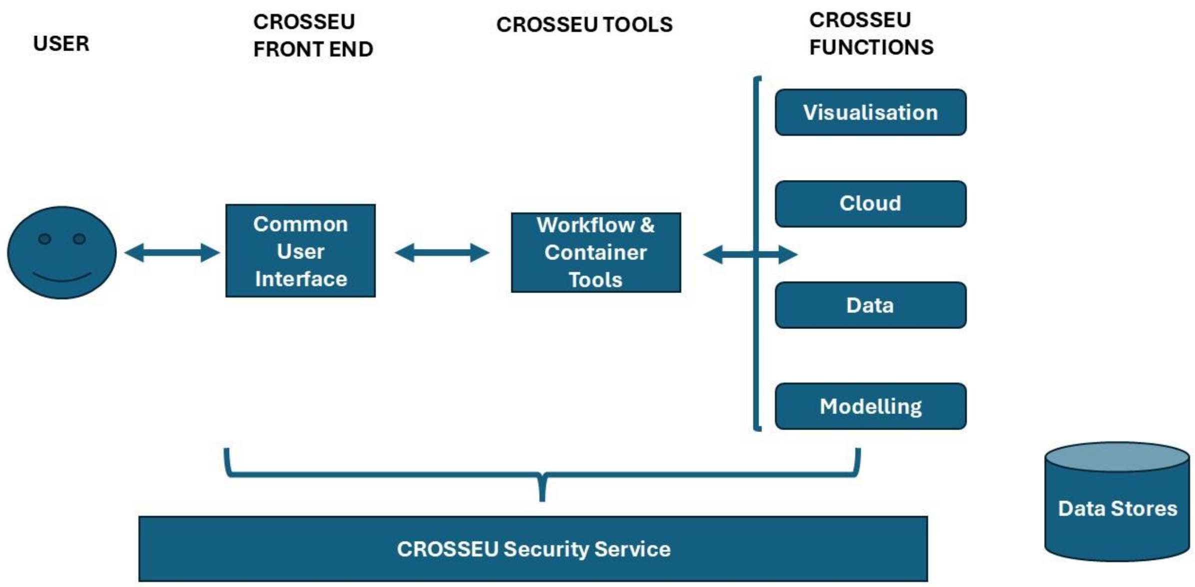

Figure 2: IAF Core Components

From an end user perspective, the main point of interaction with the platform is via the Common User Interface (Common UI). This is the point at which users can register or log-on to the platform and access its key functionalities. The Common UI interacts directly with the CROSSEU Security Service (CSS).

The CSS is the Authentication and Authorisation engine within the platform and is able to grant users access to specific resources depending on the identity tokens that they present. Tokens are linked to either username and password authentication via the CSS or Federated Identity Token presented by the user to the CSS.

At the heart of the IAF are tools and functions. The main tooling within the platform is related to the composition of workflows to utilize data and models in the IAF. The workflow tooling enables users to collaborate on the platform to build workflows spanning data and models to address specific application requirements.

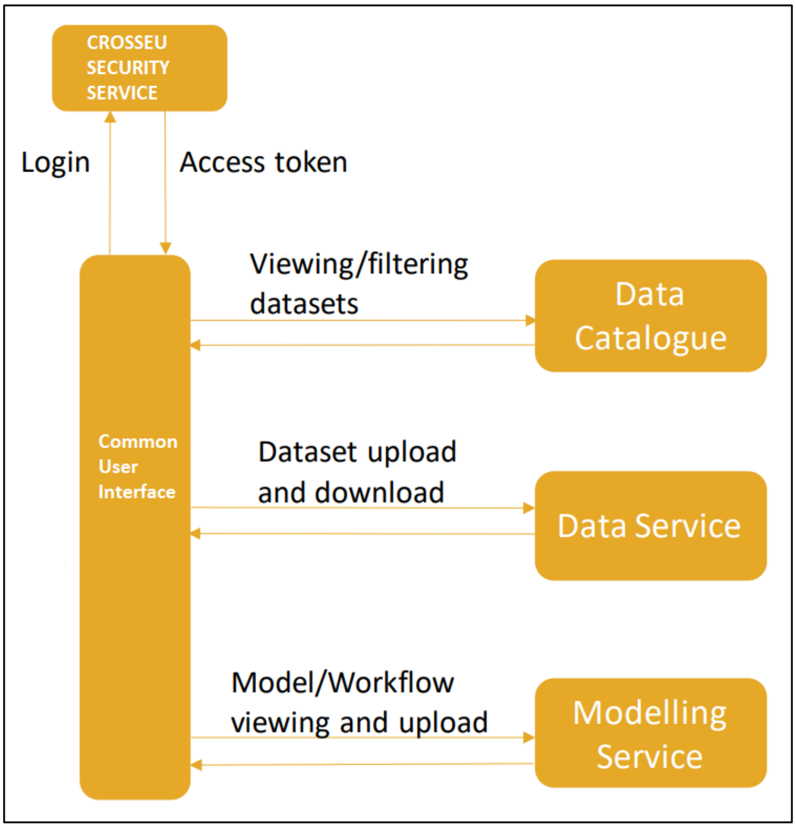

Figure 3: CROSSEU Security Service granting access to models and data

Functionality within the IAF can be split between Compute / Cloud, Data and Models (storage and retrieval / execution) and visualization. The data services include the point at which the metadata associated with models and data on the platform is stored. This metadata enables model and data discovery as part of the process in forming collaborative workflows.

Visualisation of the output from the models and workflows is enabled by the Infrastructure Visualisation Service (IVS). The IVS supports a series of standard graphical data visualization and also enables data to be integrated with third party tools such as Jupiter Notebooks. As part of the CROSS-EU work the TEAL visualisation tool is being integrated into the platforms in CROSSEU WP3.

The Data Store component can be seen as a core database within the platform. This stores data and models locally on the ICE. However, it is expected that most use cases will use data hosted outside of the ICE Data Store. In this case federated data access will be linked into the ICE to enable access to other data. This will be handled at the CSS level.

The hardware for the IAF that supports the cloud and data storage along with workflow execution and visualization is hosted at the Scientific Computing department of STFC / UKRI.

Funded by
the European Union

With respect to hardware, it is made up of 27 server nodes with 792 CPU cores and an additional 10 GPUs. Its Flash Array provides 55Tb of storage capacity to the IAF, the bulk of which is made available to platform users. A series of virtual machines host the development and deployment infrastructure and a separate Kubernetes cluster powers the IAF platform itself.

This infrastructure uses Cloud Native services (Applications) for the greater part, ensuring future flexibility and capacity to operate entirely or partially from the cloud as demand grows. The Kubernetes cluster runs on 12 nodes and – along with sister services such as Vault for secret management - supports the microservices architecture of the IAF Platform.

4.1. Using the IAF

The CROSSEU IAF built on DAFNI makes use of containerisation technology to allow users to upload, run and share models written in any open-source language. The IAF uses a standard 'model definition file' to define the scientific and technical metadata of a model, including default inputs and outputs. By standardising at the container level in this way - rather than the code level – the IAF allows researchers to work in their preferred or sector-standard language and couple their model to models written in other languages. Licenced languages, such as MATLAB, may also be possible to containerise and upload on a case-by-case basis.

Aside from being a simple catalogue of models, the IAF allows these models to be executed in a workflow engine powered by Argo. Workflows are built using an intuitive point-and-click interface builds a graph representation of the models and the sequence in which they should be executed, as well as defining which outputs are passed to models in downstream steps.

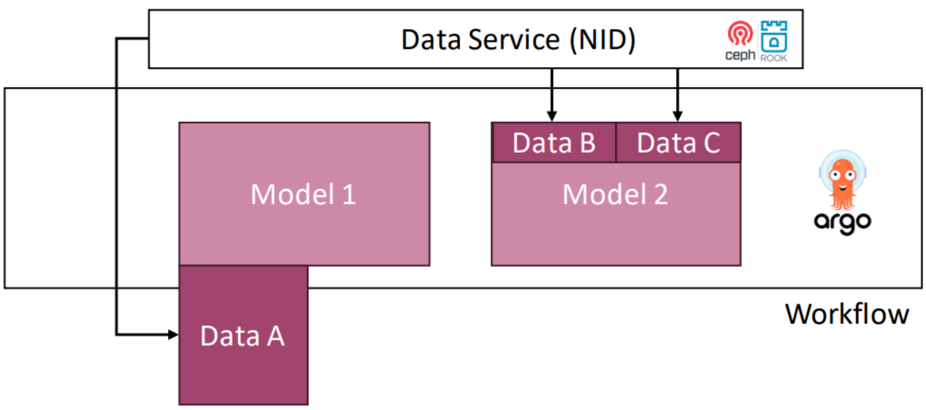

Figure 4: Linking models and data via workflows

Input files for individual models in a workflow can be selected from IAF's data catalogue and suites of input parameters can also be defined for each model. These input files and parameters can be changed for each execution of a workflow, allowing users to build a set of experiments and their outputs. The details of these experiments are retained indefinitely in the workflow pages after each run is made, providing a legacy view of the specific details of the research work undertaken.

The IAF data catalogue is an underpinning user-service. Users upload the datasets required to drive their models to the platform. Data can be any format and is held in object store.

Datasets can be explored visually using a couple of tools supported by the platform. The IAF supports a Drag-and-drop data explorer and Jupyter notebooks. Individual user-created visualisations using web-based frameworks such as Dash, Python Panels or R-shiny can be imported on to the IAF in collaboration with the development team on a case-by-case basis.

All assets uploaded to or created on the IAF are private by default. All assets can be made public on request and become available to all users of the platform. Group spaces can be created to share results with collaborators.

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.

IAF Modelling Capabilities

The IAF builds on the DAFNI modelling capabilities. This enables models' developers to create models using the programming languages that they are most comfortable with. In order to accommodate this heterogeneity, the platform wraps the models into Docker containers.

4.2. Models Stored as Docker Containers

Docker provides a set of tools for packaging and executing applications within an isolated environment known as a container. Within the container, the features of the source computing environment, such as file dependencies, software libraries and versions are captured with the application executable and are available to that application. Hence, when that container is loaded onto a new host, the application can execute as if it were in its source environment.

The user augments the Docker container with a description file, specifying the metadata associated with the application and the input and output parameters and datafiles. This description is used to catalogue the model and provide a template for input and output parameters and data files which are used to execute the application.

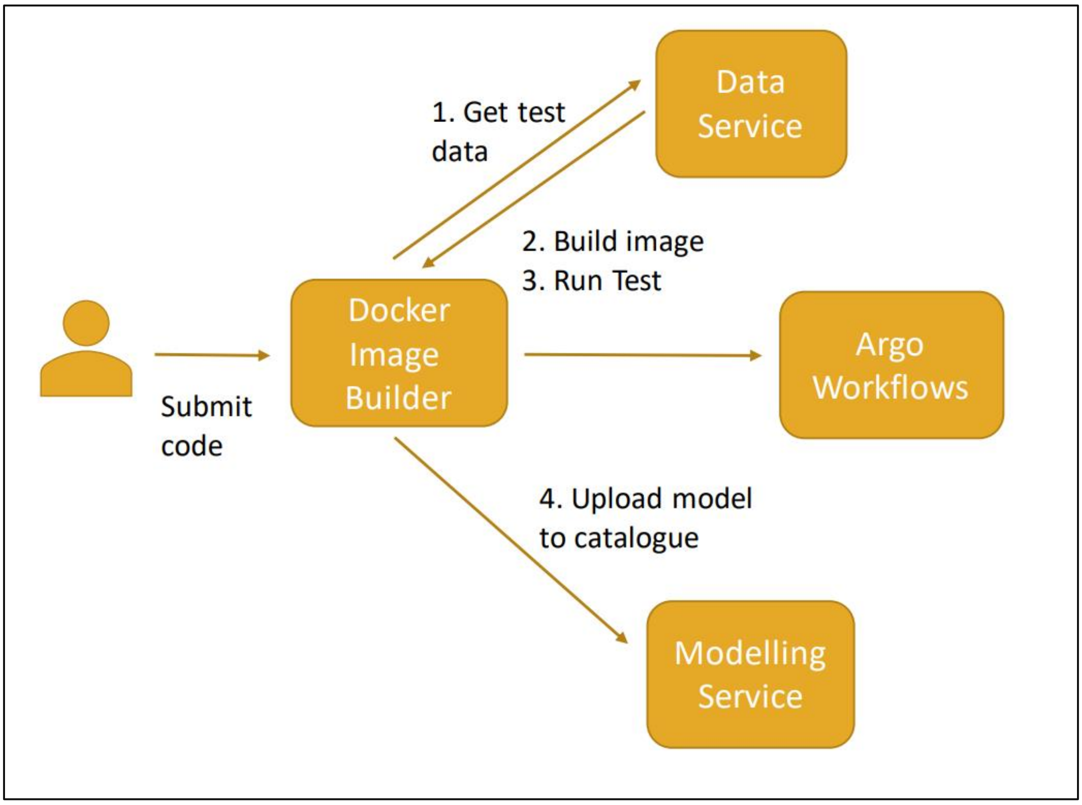

Figure 5: Docker Image Creation

The collection of containerised applications is uploaded to the IAF and shared as a CROSSEU collection within the model catalogue. Note that the framework contains multiple versions of some packages, so they are available for re-running scenarios for validation and comparison versions.

4.3. Model Metadata

As discussed, models are associated with metadata to aid discovery and the integration of them into workflows. The IAF will present a secure, searchable and version-controlled repository co-located with the modelling and execution environment and the developed framework. The nature of IAF data repository system, based on an object store platform, means its file type is agnostic so files of any type can be uploaded and metadata must be provided at the point of upload. Files can be shared by the owner (uploader in the first instance) to pre-assigned groups and various permission levels allow data to be protected from changes by members of one group, but editable by members in another.

Each dataset at the point of upload is assigned and owns two unique identifiers, the first a dataset ID and the second a version ID. This allows traceable versioning while maintaining the ability for any version to be easily accessed. Data versions can only be deleted by the principal owner.

Within CROSSEU collections of data will be created and used to explore the use case scenarios, uploaded by scenario developers. This includes extracts of common data sets, but also data required for particular models, such as habitat coverage for biodiversity, and urban areas. As stakeholders' access and use the IAF for their own analysis, they can add their own data to the collection for their own usage and share it with other stakeholders to access. Note that results data sets are also made available via the same collection.

4.4. Models Linked Together in Workflows

As discussed, IAF workflows link models supporting sector scenarios and their execution to produce the result datasets. This is enabled by the IAF workflow engine, which allows models to be chained together and iterated.

The scenario requires the framework to be scalable with the ability to run multiple workflows simultaneously, be that to run the desired scenarios, or explore the sensitivity of the workflows to parameter changes, perhaps for adaptation measures (e.g., the amount of green space required and allocated in new urban developments and the effect on urban flooding and urban heating).

Treating workflows as objects, different parameter sets can be applied to the same workflow instance and run in parallel, allowing such analysis to be undertaken, where scenarios are limited in number. For larger ensembles a looping function within workflows enables lists of specific parameters to be used to generate ensemble-like outputs.

Managing large sets of runs, ensemble or otherwise, can pose numerous challenges, including recording and managing the outputs and parameters values used in each. To support customisation, input parameters are exposed to allow the automatic generation of the required metadata files when saving outputs This allows for dataset names and descriptions to contain the key information without a user being restricted to a set of defaults or requiring more in-depth knowledge of the DAFNI platform to set metadata details for outputs. In Figure 6, we give an example of a simple workflow from the OpenCLIM project, chaining the HEAT and HARM models for evaluating the impact of extreme heat event on mortality.

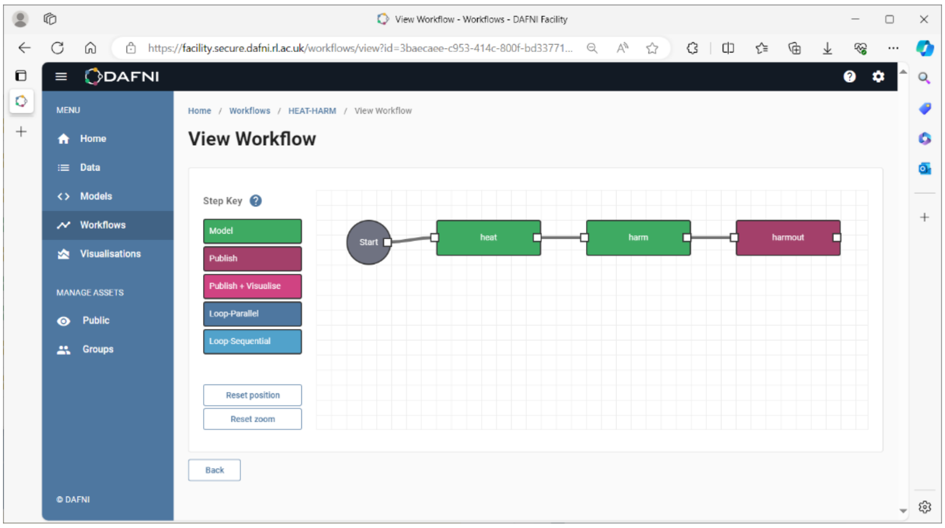

Figure 6: DAFNI workflow chaining the HEAT and HARM models (OpenCLIM, Nov 2023)

5. Methodology for populating the IAF

The CROSSEU IAF will employ a simple underpinning strategy adapted from the OpenCLIM project and designed to be compatible with the principles of open, scalable and trusted research, as shown in Figure 7. This approach allows the development of the framework to adapt to the evolution of the range of scenarios. This strategy defined the process, starting from the uploading and sharing of models developed across the partnership, combining with the appropriate input data sets, integration of models within a computational workflow, and execution to generate and explore results, including curating them for review and reuse.

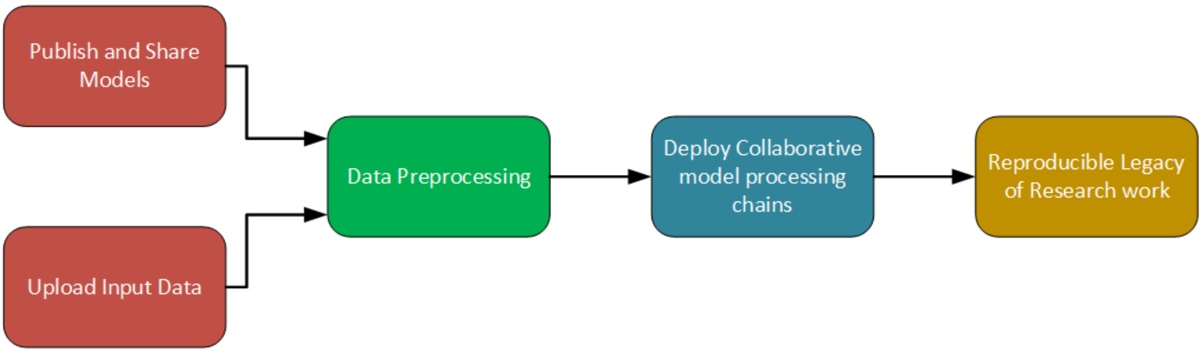

Figure 7: Model and Data Deployment Strategy

5.1. Publish and Share Models and Data

The core of the approach is the deployment and application of a number of key computational models. CROSSEU employs a number of established models which have been used in earlier studies, mostly as standalone entities, while developing extensions and exploring new approaches. This provides the challenge of developing the framework to couple disparate, sector specific models, into an integrated assessment framework, where models vary in spatial resolution, file format and structure used for output data deployed into the hybrid framework.

To identify the relevant computational models that the use cases will use within the Framework, CROSSEU has instigated an Initial Model Survey. This seeks to capture the characteristics of the models that can then be used to tailor the requirements to support those models within the IAF and to be uploaded as metadata to aid the search, discovery and integration of workflows and data within the IAF. Table 2 gives a description of the characteristics which are being collected in the survey.

Table 2: Model Survery Structure

Model Metadata	Description
Name	Name the model is usually referred by
WP and task	The WP and task number that the model is associated with
Status of Model Development	Does this model exist, is it currently in development, is it planned, it is a legacy system?
Model Description	Short textual description of the model
Model Owner/contact	Contact person(s) responsible for the model (with email)
Model Organisation	Organisation which is the primary owner (licensor) of the model.
Support contact	Please give a support contact for model changes if different from the contact above (with email).
Licence	Licence for the model (including URL)
Dependency licences	Is the model dependent on 3rd party code which is subject to other licences? Are they open source?
Access restriction	Any access conditions on using the model?
Model URL	Website describing the model and how it can be used/technical documentation.
Codebase URL	URL of where the code is maintained (e.g. GitHub repo)
Code and platform	
Operating System	Operating system(s) supported - variants on Linux preferred
Programming language	Main programming language; not mandatory, but may be useful if advice is needed to use Docker.
Dockerised	Is a Docker image provided/is there experience in using Docker/will help be required to Dockerise the model?
Parallelised code	Has the code been parallised or multi-threaded?
GPU Usage	Has the code been designed for use with GPUs?
Execution information	
Configuration file	Does the model have a configuration file with setting for input parameters? E.g .ini or .yaml file?

Pre-processing	Will the model typically require pre-processing steps?
Post-Processing	Will model outputs require post-processing to meet standards such as csv/netcdf ?
Runtime	Typical expected runtime in your standard environment - this may vary between platforms.
UI	Does the model have its own dedicated UI? If so, what kind? e.g., a purpose-built client? Web-based?
Interactive	Does the model run in interactive or batch mode?
Links to other models	Are there other models you are aware of that this model is likely to be run in a workflow with?
Additional Information	
Additional information	Any additional information which you feel might be useful.

The survey has been shared to all partners as a collaborative document. We expect more models to be added to the survey as the project progresses. The model survey is not only useful to identify models and team support needs, it also has identified gaps in the DAFNI platform capabilities on which the CROSSEU IAF is being built.

For the models to be made available on DAFNI, the independently produced software packages need to be containerised using Docker. Docker provides a set of tools for packaging and executing applications within an isolated environment known as a container. Within the container, the features of the source computing environment, such as file dependencies, software libraries and versions are captured with the application executable and are available to that application. Hence, when that container is loaded onto a new host, the application can execute as if it were in its source environment. The user augments the Docker container with a description file, specifying the metadata associated with the application and the input and output parameters and datafiles. This description is used to catalogue the model and provide a template for input and output parameters and data files which are used to execute the application. The collection of containerised applications will be uploaded onto DAFNI and made available as a CROSSEU Collection within the platform.

Within DAFNI, input data is stored in the data management and cataloguing environment providing a secure, searchable and version-controlled repository co-located with the modelling and execution environment and the developed framework. The nature of DAFNI's data repository system, based on an object store platform, means its file type is agnostic so files of any type can be uploaded, and metadata must be provided at the point of upload. Files can be shared by the owner (uploader in the first instance) to pre-assigned groups and various permission levels allow data to be protected from changes by members of one group, but editable by members in another. Each dataset at the point of upload is assigned and owns two unique identifiers, the first a dataset ID and the second a version ID. This allows traceable versioning while maintaining the ability for any version to be easily accessed. Data versions can only be deleted by the principal owner.

The CROSSEU IAF will thus include a collection of data used to explore the scenarios, uploaded by scenario developers. This includes extracts of common data sets, but also data required for individual models, such as habitat coverage for biodiversity, and urban areas for heat impact analysis. As stakeholders' access and use the CROSSEU IAF for their own analysis, they can add their own data to the collection for their own usage and share it with other stakeholders to access. Note that results data sets will also be made available via the same collection.

5.2. Data Processing

As with many integrated frameworks, data interoperability between models is a key challenge. CROSSEU has to accommodate a wide range of different application sectors, including pre-existing models that require data to conform to different formats, so a single set of project data standards is not possible. The initial datasets may not be compatible with the model inputs and model data outputs may be in a format that needs changing to work as an input for a subsequent model in a workflow. However, a harmonised approach to data and metadata supported by the project would allow data integration and exchange that would considerably enhance the usability of the IAF. Consequently, the project aims to produce a Harmonised Data Repository, supported by a harmonised data and metadata framework, that allows flexibility of using data standards from across the application domains, while also supporting a common data and metadata framework to enable data interoperability. We propose that DAFNI Data Repository and data service components form the basis of the HDR, augmented by recommendations and support for harmonised data and metadata. The HDR represents a significant part of the work of the IAF, and is thus treated within a separate deliverable, D1.6.

Building on the HDR, data transformation and processing should be supported so that models can execute on the available data, and the output data from one model can be used as input to another. Within the DAFNI platform, this data transformation can be made explicit via the use of auxiliary models which can be included in data analysis workflows and executed independently. Thus, the data processing is not bundled into application code, and therefore can be modified independently of that code, adding another dimension of flexibility to the framework.

Two forms of auxiliary models for data processing are used within the framework: (i) pre-processing of data to prepare it for input into models; and (ii) integration processing, whereby a range of 'utility' models have been developed to perform intermediate transformation steps within workflows.

Pre-processing models enable the input of data from the data repository into models, where data transformation is needed for a dataset to become interoperable, and scenario specific in some cases, for analysis to take place. These models may need to be executed less frequently to prepare a modified input data set that can then be used from the data collection in many executions.

Further, a suite of utility models should be developed for integration processing within workflows as part of the evolving IAF, some for bespoke uses for a single model integration, and others more generic allowing for the same transformations to be used across multiple workflows. The scale of utility models will vary, from a model used to extract files from an archive, to extract date from external datasets, to changing the spatial scale of a dataset with multiple input parameters. This suite of models will enable a consistent and managed approach in resolving interoperability issues, while a thorough and methodological testing and verification process has been undertaken on such tools.

5.3. Collaborative Model Processing

The final component of the CROSSEU IAF is the definition of workflows supporting sector scenarios and their execution to produce the result datasets. This was enabled by the DAFNI workflow engine, which allows models to be chained together and iterated.

The scenario-based nature of the analysis to be undertaken in CROSSEU requires the framework to be scalable with the ability to run multiple workflows simultaneously, be that to run the desired scenarios, or explore the sensitivity of the workflows to parameter changes, perhaps for adaptation measures. Treating workflows as objects, different parameter sets can be applied to the same workflow instance and run in parallel, allowing such analysis to be undertaken, where scenarios are limited in number. For larger ensembles a looping function within workflows enables lists of specific parameters to be used to generate ensemble-like outputs.

Managing large sets of runs, ensemble or otherwise, can pose numerous challenges, including recording and managing the outputs and parameters values used in each. To support customisation, input parameters are exposed to allow the automatic generation of the required metadata files when saving outputs This allows for dataset names and descriptions to contain the key information without a user being restricted to a set of defaults or requiring more in-depth knowledge of the DAFNI platform to set metadata details for outputs.

5.4. Reproducible Legacy of Research

A principle of the IAF is to ensure the legacy of the modelling, both the results presented and the ability for others to run new scenarios or reproduce results beyond the life of the initial project. The framework has therefore been designed with this in mind, with a curated set of identifiable and recorded workflows, stored with provenance information bringing together the models and existing parameter sets to enable these workflows to be more easily reused. Many of the legacy issues associated with traditional research are addressed through the underlying DAFNI architecture, with reported results fully traceable back to specific versions of models, datasets and unique parameter sets.

6. Next Steps and Future Development

This deliverable outlines the requirements and key components of the IAF architecture and how they will be used to support the CROSSEU wider objectives.

Next steps are to:

- Complete model survey to collect a comprehensive view of the models that CROSSEU will support in the platform.
- Work with CROSSEU use case teams to identify the workflows that support those use cases and consider how they can be realised in the IAF.
- Revisit training on the use of DAFNI with CROSSEU implementations teams.
- Extend the DAFNI metadata model to incorporate feature requests from CROSSEU
- Analyse the data requirements from the data survey to identify the core commonalities and features which should be represented in a harmonised data framework and supported in the HDR. Propose a version 1 of the harmonised data framework.
- In light of the data and modelling requirements, revisit the IAF architecture to provide more detail on the coordination with external data sources and third-party computational platforms, as well as the DSS. In addition, as more models and data sets are added we expect to define in greater detail the core services of the IAF particularly around security and privacy.

The next phase of the project will integrate CROSSEU models onto the platform. These models will be tested and deployed into workflows which we will expect raise future areas of work for the components in the IAF. For example, the security requirements of model/data providers could possibly require the integration of specific designs into the security architecture and processes of the IAF.

Bibliography

- Applications, I. C.-N. (n.d.). (https://ieeexplore.ieee.org/abstract/document/8125550.
- CROSSEU. (n.d.). D1.1 – Report on co-design and operationalisation of the CROSSEU methodology .
- Hall, J. W. (n.d.). The future of national infrastructure: A system-of-systems approach. Cambridge University Press, 2016.
- Little, C. M. (n.d.). "The relationship between US East Coast sea level and the Atlantic meridional overturning circulation: A review." Journal of Geophysical Research: Oceans 124.9 (2019): 6435-6458.
- Matthews, B.M. et. al. (2023) DAFNI: a computational platform to support infrastructure systems research, Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction https://doi.org/10.1680/jsmic.22.00007

CROSSEU Partners

Meteo Romania	UEA University of East Anglia	WORLD METEOROLOGICAL ORGANIZATION
Università degli Studi di Padova TESAF	UCL	K&I Conoscenza e Innovazione
Helmholtz-Zentrum hereon	LGi sustainable innovation	eDF
BOKU	DTU	WEMC World Energy & Meteorology Council
UK Research and Innovation	University of Bucharest	CZU