ForestNavigator

D9.1 Concept of the ForestNavigator Portal

Date
18.07.2023

Grant Agreement	101056875
Call identifier	HORIZON-CL5-2021-D1-01
Project full title	ForestNavigator: Navigating European forests and forest bioeconomy sustainably to EU climate neutrality
Work package	WP9
Due date	31/07/2023
Deliverable lead	IIASA
Authors	Andrey Lessa Derci Augustynczik, Rupesh Singh, Fulvio Di Fulvio, Petr Havlík, Quentin Lejeune, Andyara Callegare

Abstract

This report presents the conceptualization of the ForestNavigator portal and its components, namely the "Data Repository", "Model Repository" and "Computing module". These three components (together referred to as the platform) will host the data, models, and source codes from the ForestNavigator project, as well as workflows for data processing and model deployment. This platform will allow ForestNavigator researchers and external researchers to interoperate shared data and models, to reduce modelling cycles, and to increase the outreach of the project.

The ForestNavigator portal will also allow visualisation of spatially explicit data and pathways via two webtools, the "EU Data Explorer" and the "EU Pathways Explorer". Both publicly accessible visualization tools will contain the main output of the ForestNavigator project. This report describes the technical infrastructure around the architecture of the components, as well as their functionalities. This includes the definition of interfaces for exchange between models, data and visualization tools employed on the ForestNavigator portal.

Keywords

Cyberinfrastructure, computing platform, interoperability, reproducibility, data management

Internal Technical Auditor	Name (Beneficiary short name)	Date of approval
Task leader	Andrey Lessa Derci Augustynczik	17/07/2023
WP leader	Quentin Lejeune	14/07/2023
Coordinator	Petr Havlík (IIASA)	17/07/2023
Project Office	Sophie Rau (ARTTIC)	18/07/2023

Dissemination level

PU	Public, will be published on CORDIS	✔
SEN	Sensitive, limited under the conditions of the Grant Agreement	

Nature of the deliverable *   R

Table of Contents

List of Figures

Abbreviations

API	Application Programming Interface
CA	Climate Analytics
DAG	Directed acyclic graph
DFS	Distributed file system
IIASA	International Institute for Applied Systems Analysis
EC	European Commission
EU	European Union
GFZ	Geo Forschungs Zentrum (German Research Center for Geosciences)
HTC	High-throughput computer
HPC	High-performance computing
WP	Work Package

1. Introduction

This report presents the conceptualization of the ForestNavigator portal, which serves as a comprehensive platform for integrating technical infrastructure requirements and key components necessary for an efficient data exchange and web-based visualization of the results from the ForestNavigator project.

To leverage the increasing availability and frequency of datasets for modelling, a cyberinfrastructure able to efficiently integrate novel data and models is key to support the further development of analytical tools at the service of policy assessments and to improve models' projections. Thereby, automated data assimilation procedures may be established, streamlining inputs, outputs and removing redundancies in data harmonization efforts, increasing the scalability of simulation efforts (Fer et al., 2021). Furthermore, an important component of such modelling system relates to the capabilities to communicate and disseminate modelling results, following FAIR principles, via visualization tools that facilitate the use of the outputs for decision-making (Souffront Alcantara et al., 2018).

The ForestNavigator portal is designed to provide a cyberinfrastructure with the following aims:
- foster model-data integration in near real time
- shorten modelling cycles
- provide access to the main analytical tools and results developed in the project
- enable stakeholders to navigate through the main project outputs through data and pathways explorers, and retrieve the underlying data, for supporting their decision-making processes.

This report describes the technical infrastructure needs and its main components, including the definition of interfaces for exchange between models and visualization of ForestNavigator main results.

2. Overall structure of the Portal

The ForestNavigator Portal includes five operational components. Figure 1 describes the components of the Portal and its different likely users. The first three components of the Portal are referred to as the "Data Repository", "Model Repository", and "Computing module". The first two will contain the data and models from the ForestNavigator project, and the last the workflows for data processing and model deployment. In conjunction, these three components are referred to as the platform, which will allow ForestNavigator researchers and external researchers to interoperate shared data and models.

The two remaining components of the Portal will be the "EU Data Explorer" and the "EU Pathways Explorer". These components visualise spatially explicit data and the policy pathways/scenarios, respectively. Both visualization tools will be publicly accessible, together referred to as Web Explorers, and contain the main results of the ForestNavigator project.
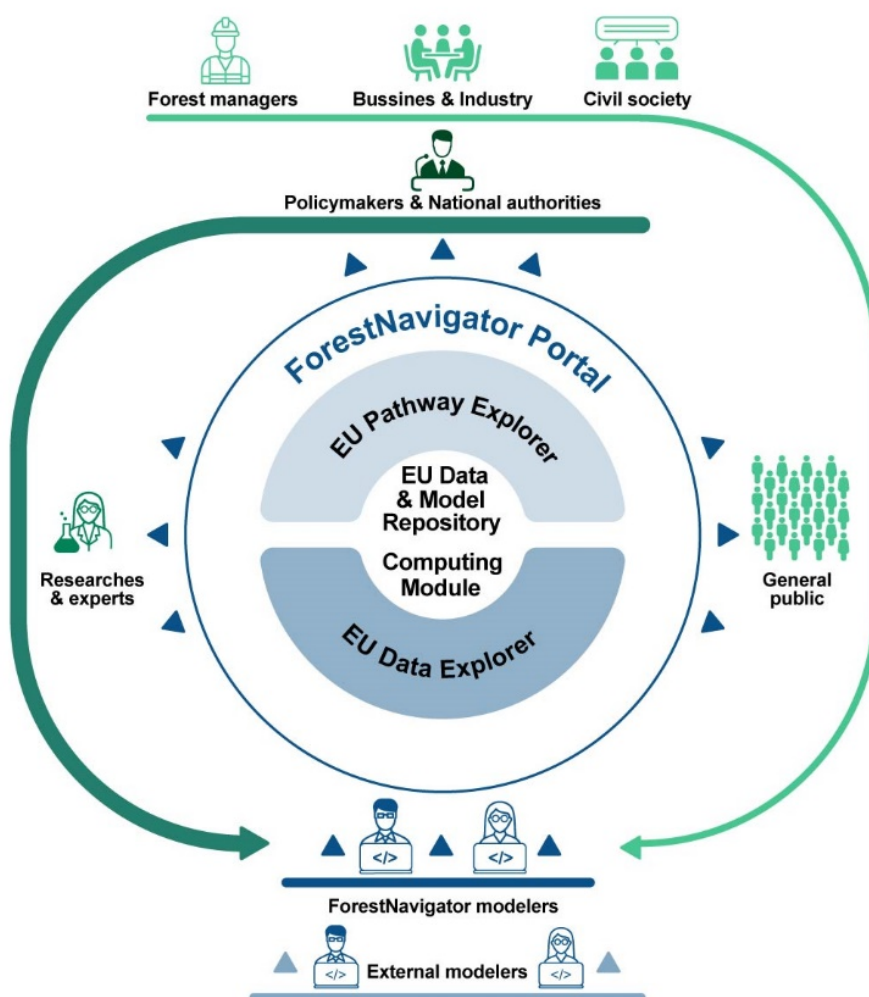

Figure 1. The ForestNavigator Portal and its main components.

2.1. Data Repository, Model Repository and Computing module

Data repository

The data storage hubs (Data Repository) will be capable of storing petabytes of data and, when required, the capacity can be extended by including additional hardware. This repository will be physically present close to a computing cluster and network to enable running jobs or models and perform read and write operations efficiently.

The Data Repository will contain multiple shared datasets that will be accessed by the models for producing the pathways and the georeferenced results, stored in standard data formats, sourced from the outputs of ForestNavigator Work Package (WP) 2 to WP7.

The Data Repository will comprise a distributed file system (DFS) along with a platform control server. The input / output operations will be performed on the DFS by a computing module. In this context, several actions can be performed on data included in the Data Repository via a platform control server (Figure 2).

Users: Researchers in ForestNavigator and outside the project
Hosting: IIASA cloud space DFS

Model repository

This component will contain the models used to compute the mitigation pathways and indicators, and two other subcomponents, namely a git repository and a container image repository.

The git repository uploaded to the cloud will host source codes of models or jobs in a versioned and documented manner. The container image repository will host container images of the models, which will be executed directly on a computing cluster. Container image building can be automated on every major release or tag of the source code of a model or job, drawing on the most updated model versions (Figure 2). Container image repositories are also called container image registries (e.g., Docker registry).

The main models hosted on the git repository and docker registry will be G4M-X-GLOBIOM and MESMER-L-X. Further models will be identified and added to during the project lifetime.

Users: Researchers in ForestNavigator and outside the project
Hosting: IIASA cloud space, GitHub, Docker registry

Computing module

This module will be responsible for dispatching and running a job or jobs in a pipeline to a computing cluster. Jobs are usually organized in workflows, represented as directed acyclic graph (DAG).

This module will contain a client application connected to a server, which is close to the computing cluster to improve performance. The client application will have the interface for users to dispatch jobs by pointing to the proper job container image from registry, as well as input and output

locations in the data repository. After the job dispatch, actions are made on the client application, job instructions would be passed to the computing cluster via a platform control server. The computing module is represented schematically in Figure 2.

Users: Researchers in ForestNavigator and in the Policy Modelling Forum
Hosting: IIASA cloud space

2.2.   Structure and processes

Having scalable data store and scalable computing capacity are essential when dealing with big data and big data analysis, as in the case of the ForestNavigator project. The ForestNavigator portal will feature a DFS for data scalability and a high-throughput computer (HTC) cluster for computing power. A platform control server stays at the top of this architecture and acts as a controller for the DFS and HTC cluster. Requests to the platform control server are made from a platform client application and are usually responsible for exploring scenarios or simulations results and dispatching workflows related to simulations to the HTC (Figure 2).

Jobs, including models and data processing scripts, are expected to be containerized and saved as a container image on a container image registry. When the jobs are dispatched, the HTC will pull the job container image from the registry and execute it, while all the input/output operations are performed on the DFS.

To handle spatially explicit datasets, such as maps and gridded modelling results, a geo server can directly read the data from the DFS and serve it to GUI clients efficiently. This enables exploration by the users. Hereafter, we describe each of these components in detail according to Figure 2.
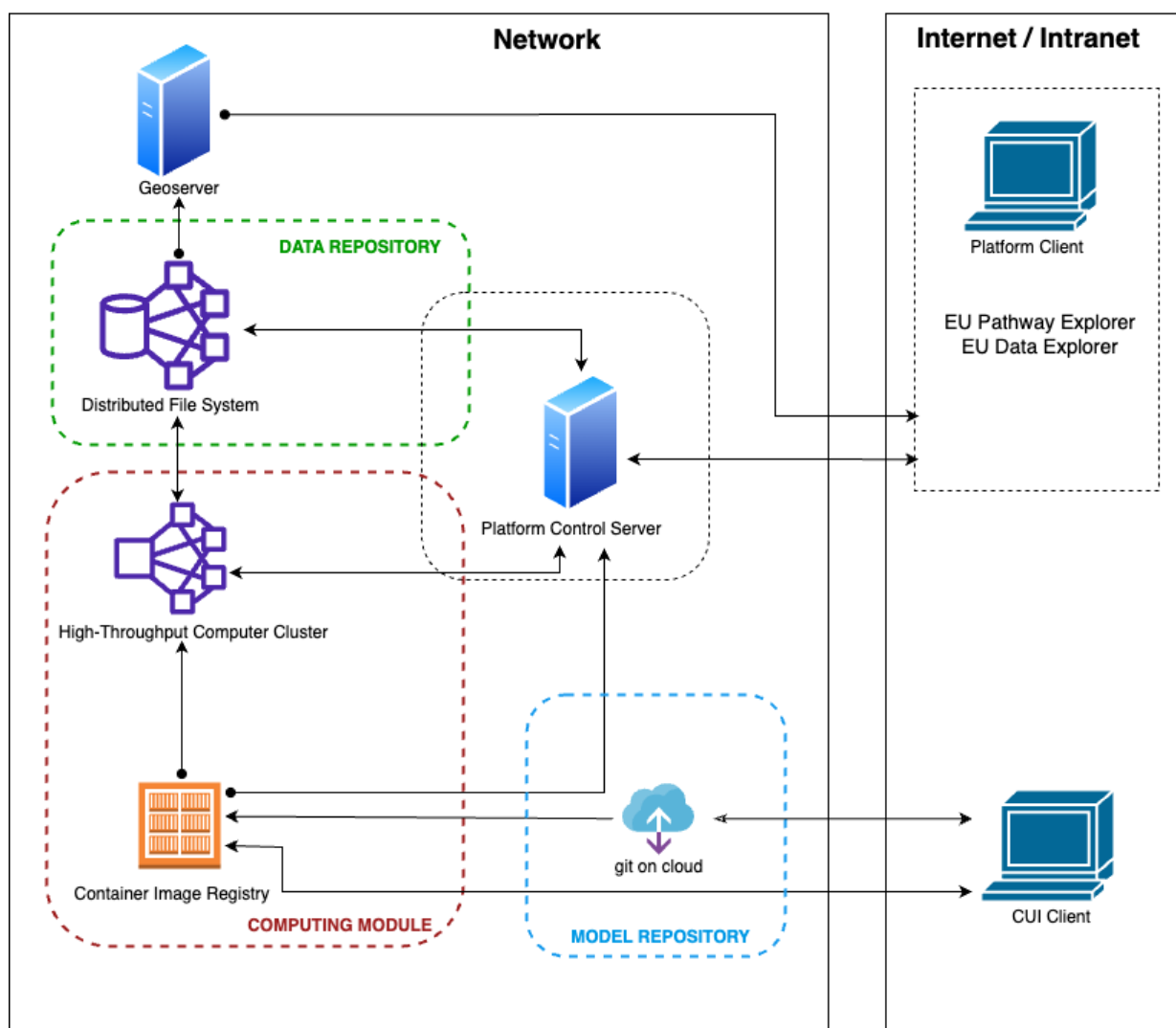
Figure 2. The ForestNavigator platform infrastructure architecture diagram.

Distributed file system

A DFS is a client/server-based network file system that allows multiple users to access and share files across a geographically dispersed network. The files are stored on multiple servers in different locations, and data access is coordinated by a central server or set of servers. This enables better data availability, reliability, and scalability, as well as improved performance and fault tolerance. Examples include Hadoop Distributed File System (HDFS), Google File System (GFS), and Amazon Web Services (AWS) Elastic File System (EFS).

Advantages:
- Transparent local access — Data is accessed as if it's on a user's own device or computer.
- Location independence — Users do not need to know where file data physically resides.
- Scaling — Teams can add as many machines as they want to a DFS to scale out.
- Fault tolerance — A DFS will continue to operate even if some of its servers or disks fail.

High-throughput computer cluster

An HTC cluster is a type of computing system that is designed to process many parallel jobs in an efficient and cost-effective way. It typically consists of multiple interconnected computers, or nodes, that work together to achieve a high level of processing power and throughput.

Unlike high-performance computing (HPC) clusters that are optimized for running a few large, complex simulations or calculations, high-throughput clusters are optimized for running many smaller, independent jobs simultaneously. This makes them well-suited for data-intensive applications such as the exploration of multiple alternative scenarios and pathways, where large quantities of data need to be processed quickly and efficiently.

High-throughput clusters often use specialized software frameworks, such as Apache Hadoop or Apache Spark, to distribute and manage the processing of data across the cluster. They may also incorporate other technologies, such as distributed file systems, high-speed networking, and scalable storage systems, to support their high-throughput capabilities.

Container image registry

A container image registry is a central repository for storing and sharing container images. Container images are the packaged software artifacts used to create and run containers, and they typically include all the necessary components (e.g., code libraries, runtime, dependencies) needed to run an application in a containerized environment.

Container image registries allow developers to store and share their container images with others, making it easier to deploy applications across different environments. They also provide version control, access control, and security features, enabling organizations to manage and secure their container images throughout the software development lifecycle.

Some popular container image registries include Docker Hub, Google Container Registry, Amazon Elastic Container Registry (ECR), and Red Hat Quay. Many of these registries offer both public and private repositories, depending on whether developers want to share their images publicly or keep them private within their organization.

Geoserver

GeoServer is an open-source server that allows users to share and edit geospatial data through standard protocols, such as Web Map Service (WMTS), Web Feature Service (WFS), and Web Coverage Service (WCS). In the context of platform, we will use geoserver to serve geographical maps from DFS to the client software.

Platform Control Server

The control server acts like an entry point or the gateway to the platform infrastructure. All the client requests for HTC and DFS are passed via a control server. The control server will also be responsible for managing and serving project level information along with the user level access policy.

Platform Client

This is the platform client application for the end user. Two major features of this client application are: 1) to explore the scenarios and results 2) to run the job or jobs in a pipeline in the computing module.

2.3. Functionalities required for the orchestration platform

The orchestration platform will provide several functionalities to the users, to increase the project's impact and outreach:

1. Users inside and outside IIASA can sign up and use the platform.
2. The platform will have proper access control policy for its users. Only the authorized users will have access to the resources and actions present on the platform.
3. Environment with the proper namespace can be created for the project and users affiliated to the project could be added to access project's resources and dispatch actions.
4. A container registry with proper access control will be provided for the researchers to upload model container images.
5. The platform will have an interface to dispatch job(s) in pipeline to the executing computer cluster.
6. Users should be able to set parameters required to run the job from the platform.
7. Datasets can be uploaded to the project data repository.
8. The platform will provide a graphical user interface, namely the scenario explorer, where users can analyze and visualize the model results.
9. Users should be able to explore all the data related to the project and mark datasets for visualization and analysis in a graphical user interface provided by the platform.
10. Some datasets in the project data repository will be marked as public for visualization purposes for the public users.

2.4. Data to be hosted in the Web Explorers

The data specifications are differentiated according to the sections of the Portal, given the specific technical requirements, namely spatially explicit datasets, containing maps and regionally aggregated outputs and indicators of interest, as well as mitigation pathways.

EU Data Explorer

The EU Data Explorer will provide functionalities to visualize selected geospatial data products obtained from WP2 to WP5 and retrieved from the IIASA cloud repositories, as well as implementing workflows to retrieve tailored aggregated information. A Beta version of the EU Data Explorer will be available online by March 2025.

Main data output could be sourced from (to be further refined/prioritized based on stakeholder preferences):
- Multilayered forest geodatabase at service of monitoring and modelling carbon and biodiversity (D2.1, led by GFZ)

- Forest carbon stock state and change in dependency of climate and management (100 m to 5 arcmin) (D2.2, led by BOKU)
- Yearly update of data on forest cover and disturbances events mapped at high-resolution and at short time intervals (100 m to 5 arcmin) (D2.3, led by GFZ)
- Yearly updates in EU-wide and country GHG inventories (D2.4, led by GFZ)
- Datasets on extreme events from the MESMER-L-X emulator, and their impacts on G4M-X mitigation potentials (5 arcmin) (D3.4, led by IIASA)
- Dataset containing the high-resolution EU scale mapping of G4M-X biophysical potentials for carbon sequestration from afforestation, restoration, and changes in managements, (5 arcmin) (D3.5, led by BOKU)
- Report on EU forest management costing module and green jobs calculator (D4.4, led by IIASA)
- WP4 MESMER-L-X climate feedback calibration scenarios (D4.5, led by LMU)
- Report on recreational and cultural value of EU forests under alternative management options (D4.6, led by ENEA)
- Biomass demand trajectories for energy and materials (National) (D5.1, led by UHE)
- HWP contribution potentials to EU forest sector mitigation (National) (D5.2, led by TI)
- Substitution potentials and impacts of wood use in the EU (National) (D5.3, led by UHE)
- Green growth and welfare potentials in the EU (NUTS2/National) (D5.4, led by WR)

Note: (D) Deliverables will be released automatically on www.ForestNavigator.eu/resources once submitted to the EC. Partner descriptions are available here: https://www.forestnavigator.eu/about/who-we-are/

Users: primarily National/EU decision makers and other stakeholders (researchers, experts, forest managers, NGOs, business and industry, and civil society)
Format of data sources: Geotiff/netcdf (this last format will be preferred), .csv files
Spatial resolution: 5 Arc-minute and finer (down to 100m) for unprocessed data, coarser resolutions for aggregated results (NUTS2, National levels, etc. to be refined based on stakeholder preferences in WP1)
Time: Dynamic monthly, yearly time steps to 5 years-time steps depending on the indicator
Data Hosting: Data repository at the IIASA cloud.

EU Pathways Explorer

The EU Pathways Explorer will provide a platform to visualize data output (aggregated nationally or at NUTS2) obtained from WP6,7,8, especially on climate mitigation pathways and associated socioeconomic and environmental indicators. The explorer will allow the users to navigate through the pathways, obtaining background information for each time-step, as well as compare different scenarios, in terms of key indicators related to forest ecosystems and bioeconomy. A Beta version of the EU Pathways Explorer will be available online by September 2025.

Main data output could be sourced from (to be further refined/prioritized based on stakeholder preferences):
- National and EU baseline pathway for forest sector (D6.3, led by IIASA)
- EU mitigation pathways under alternative woody biomass uses (D7.1, led by E3M)
- EU mitigation pathways under biodiversity goals and other environmental targets (D7.2, led by SLU)
- EU mitigation pathways under adaptation to environmental change (D7.3, led by LMU)
- EU pathways in the global context (D7.4, led by WR)
- Assessment of robustness for the EU explorative pathways (D7.5, led by ETH)

- EU reference policy scenario (D8.1, led by IEEP)
- Reconciled pathways towards medium/long term EU policy objectives (D8.2, led by IIASA,)
- Socioeconomic and environmental indicators associated with each of the pathways.

Note: (D) Deliverables will be released automatically on www.ForestNavigator.eu/resources once submitted to the EC. Partner descriptions are available here: https://www.forestnavigator.eu/about/who-we-are/

Users: primarily National/EU decision makers, and other stakeholders (researchers, experts, forest managers, NGOs, business and industry, civil society)
Format of data sources: .csv files
Spatial resolution: NUTS2, National, EU Regions, EU
Time: Dynamic 5 years-time steps to 1 year.
Data Hosting: data server in IIASA cloud based on GitLab technology and deployment for visualization though Kubernetes orchestrated Application Programming Interface (API).

2.5. Functionalities required to the Web Explorers

The Web explorers developed within the context of the ForestNavigator portal will provide several functionalities to allow a seamless retrieval of outputs and relevant information by the users. Below functionalities are listed separately for the EU Data Explorer and EU Pathways Explorer.

EU Data Explorer

1. Operational web-based service for user friendly visualization of geospatial datasets output of the project.
2. Maps selectable for various years, changes compared to a reference year/period can also be displayed.
3. Visualization of time dynamics using time series plots.
4. Customizable visualization of georeferenced data at different scales (national, NUTS2 levels, grid cell, watershed etc., depending on the indicator).
5. Display of aggregated pre-calculated statistics associated with each 5arcmin/NUTS2/National border.
6. Dashboard accessible online associated to the geospatial grid, summarizing key statistics for the spatial scale selected by the user.
7. Comparison of options/years, for example via a split screen.
8. Data download for web users from the IIASA back-end clouds.
9. Secured data access (where needed).
10. Extra pages featuring results for specific indicators and regions (stories) or providing information on the datasets and underlying methodologies.
11. Links to the EU Pathways Explorer to be able to investigate projected changes in some indicators covered by the Data Explorer under various scenarios/pathways.

Formats of data download: .csv, .tif, .nc files
Tool physical location: The CA geospatial data browser will be Web accessible with its scripts cloned on the IIASA cloud.

EU Pathways Explorer

Operational web-based service for user friendly visualization of pathways output of the modelling work carried out in project:

1. Visualization of non-georeferenced data (pathways) including two dimensions (x-axis = time) and magnitude (y-axis = carbon sequestration, socioeconomic and environmental indicators)
2. Customizable aggregation/split of forest sector mitigation categories included in the pathways (for example possibility to visualize separately the contribution of 'forest management' and 'afforestation' to 'forest mitigation' contribution.
3. Filtering function to select and display relevant pathways and their components of interest for the user story boards describing the scenarios included in the pre-computed pathways and guiding the selection of pathways by the user visual comparison of different scenarios/pathways (split-screen function).
4. Dashboard accessible summarizing key characteristics of socioeconomic and environmental indicators associated to the pathways, for example using qualitative (e.g., traffic light) or quantitative (e.g., scores) assessments of how much the pathways focus on some key dimensions/policy priorities (e.g., bioeconomy, biodiversity conservation).
5. Possibility to build new pathways based on pre-computed ones by allowing to select different policy priorities in various regions and visualize the outcomes of these new combinations.
6. Graph and data download for web users from IIASA back-end clouds.
7. Links to the EU Data Explorer to be able to investigate the evolution of some key indicators covered by the EU Pathway Explorer.
8. Secured data access (where needed).

Formats of data download: .csv files
Tool physical location: The CA data browser will be Web accessible with its scripts cloned on the IIASA cloud

2.6. Interface between the IIASA Repositories and the CA Web Explorers

CA will have authority to access datasets released by the project on distributed file systems via a web API. Each type of released datasets will follow a specific pattern of metadata which can be used to detect the released dataset properties and the location of dataset itself. Thus, CA can use the web API to find all releases with its metadata and files and use it in their web interface. *Figure 3* shows a schematic of the flow of information between the various components of the interface between the IIASA Data Structure and the web explorers leading to the visualization of data corresponding on a specific user selection.

The interface will be distributed across the following containers:

1. **Web Explorer**: The user-facing part of the application that encompasses design, layout, and functionality. Users directly interact with it on their devices, such as web browsers. This platform user client application will contain both the EU Pathway Explorer and the EU Data Explorer.

2. **Application Programming Interface (API)**: The API serves as a bridge between the IIASA data structure (Platform Control Server, Distributed File System, and Geoserver), and the Web Explorer. It manages communication, processing, and data exchange between different system components. After receiving the request from the Web Explorer, the API will query the IIASA data structure for the relevant data, if necessary, it will process it (i.e.: calculating averages, clipping region of interest, etc.), and then return the relevant data or responses to be displayed in the Explorer. Additionally, it also serves as a back-end component, responsible for handling data processing, logic, and communication with the database and/or data repositories.
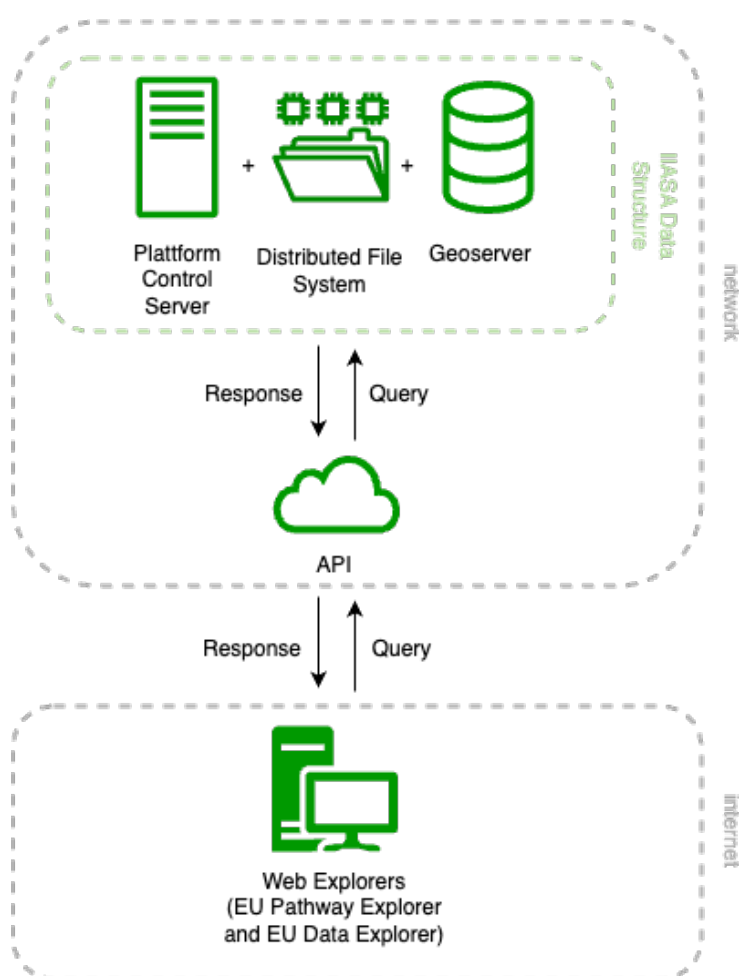

Figure 3. Information flow and visualization process of the IIASA data structure and web explorer's interface

2.7. Data and model updates

IIASA will expose the REST API on http, which will list all the releases made in ForestNavigator and will also provide details of each release. These releases will contain updated data or model versions, which can be leveraged by the Portal to update input data and model outputs. *Figure 4* exemplifies the capabilities of the REST API provided.

Sample release api

List all available releases made until now.

http method: *GET*
url: *<forest-navigator-api-domain>/releases*
successful response status: *200*
successful response data:

```
{
    "total_count": 30,
    "limit": 20,
    "offset": 0,
    "results": [
        {
        "id": "dc80ca-16c0-11ee-be56-0242ac120002",
            "name": "Accessment X",
            "released_datetime": "2012-01-26T13:51:50.417-07:00",
            "type": "IAMC"
        },
        {
        "id": "dc80ca-16c0-11ee-be56-0242ac120002",
            "name": "Accessment Y",
            "released_datetime": "2012-01-26T13:51:50.417-07:00",
            "type": "NETCDFX"
        },
        ...
    ]
}
```

Get IAMC release detail

http method: *GET*
url: *<forest-navigator-api-domain>/releases/<release-id>*
successful response status: *200*
successful response data:

```
{
    "id": "dc80ca-16c0-11ee-be56-0242ac120002",
    "name": "Accessment X",
    "released_datetime": "2012-01-26T13:51:50.417-07:00",
    "type": "IAMC",
    "iamc_csv_data": "https://forest-navigator-s3-bucket-
url/0242ac120002/accessmentx_scenarios_result.csv",
    "data_index_sqlite_db": "https://forest-navigator-s3-bucket-
url/0242ac120002/accessmentx_scenario_result_fields_index.sqlite"
}
```

Get NETCDFX release detail

http method: *GET*
url: *<forest-navigator-api-domain>/releases/<release-id>*
query_params:
    from_date: filter temporal data by from date
    to_date: filter temporal data by to date
    dimx: filter temporal data by dimx (NETCDFX specific)
successful response status: *200*
successful response data:

```
{
    "id": "dc80ca-16c0-11ee-be56-0242ac120002",
    "name": "Accessment Y",
    "released_datetime": "2012-01-26T13:51:50.417-07:00",
    "type": "NETCDFX",              # example type
    "raw_model_data": "https://forest-navigator-s3-bucket-
url/0242ac120023202/accessmentx_scenario_result.nc",
    "temporal_data": [
        {
            "date": "2012-01-26",
            "dimx": "dimxValue",       # dimension native to NETCDFX
            "dimy": "dimyValue",       # dimension native to NETCDFX
            ...                        # and other NETCDFX native dimensions
            "geotiff": "https://forest-navigator-s3-bucket-
url/0242ac120023202/accessmentx_scenario_result_2012-01-26.tif"

        },
        ....
    ]
}
```

Figure 4. Capabilities of the REST API

2.8. Maintenance

The Data Repository, Model Repository, and Computing module are set to be developed by IIASA and the EU Data and Pathway Explorers by CA. After the initial release of the Web Explorers, it is the responsibility of IIASA to maintain and further develop them in the future. Clean architecture is set as a common way of designing the system, so that it is highly reusable, testable, and maintainable by the maintaining party. Thus, the development of the Web Explorers will be based on clean architecture principles. In addition, to enable maintenance and further development, IIASA will have access to the API and CA Web Explorers' source codes (bridging between the IIASA data infrastructure and the Web Explorers, see section 2.6) on github or gitlab.

Maintenance of the project code (in a well-documented and versioned manner) and infrastructure will be conducted will the help of an issue tracker. IIASA will maintain an issue tracker where all the issues will be reported by users and testers. Issues can be created for bugs, security issues, new features, and enhancement recommendations for the platform.

IIASA will further investigate the issues in a periodic manner with developers and stakeholders to set a priority. Once the priority is set, the issues will be handed to developers and researchers to get it resolved. When the issue is marked as resolved by developers, it will again be subject to testing by curators responsible for quality control. Once passed by the quality control team, the developed resolution would be released to the production system.

3.   References

Fer, I., Gardella, A. K., Shiklomanov, A. N., Campbell, E. E., Cowdery, E. M., De Kauwe, M. G., ... & Dietze, M. C. (2021). Beyond ecosystem modeling: A roadmap to community cyberinfrastructure for ecological data-model integration. *Global Change Biology*, *27*(1), 13-26.

Souffront Alcantara, M. A., Kesler, C., Stealey, M. J., Nelson, E. J., Ames, D. P., & Jones, N. L. (2018). Cyberinfrastructure and web apps for managing and disseminating the national water model. JAWRA Journal of the American Water Resources Association, 54(4), 859-871.