FIREURISK - DEVELOPING A HOLISTIC, RISK-WISE STRATEGY FOR EUROPEAN WILDFIRE MANAGEMENT

Grant Agreement Number: 101003890 **Call identifier:** H2020-LC-CLA-2018-2019-2020	
Topic:	LC-CLA-15-2020 Forest Fires risk reduction: towards an integrated fire management approach in the E.U.
Instrument:	RIA

D4.3 - Internal cooperative platform of models, data and knowledge

Deliverable Identifier:	D4.3
Deliverable Due Date:	31/03/2022
Deliverable Submission Date:	31/03/2022
Deliverable Version:	1
Lead partner:	SATWAYS
Authors:	SATWAYS: Leonidas Perlepes, Katerina Valouma, Aggelos Aggelis, Antonis Kostaridis
Work Package:	WP4 - Multi-level integration
Task:	Task A4.1.3 FirEUrisk internal cooperative platform of models, data and knowledge
Dissemination Level:	☐ PU: Public ☒ CO: Confidential, only for members of the Consortium (including the Commission Services)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003890.

Revision History

Version	Date	Edited by	Description
v.0.1	10/01/2022	Antonis Kostaridis (SATWAYS)	Document Content outline / structure
v.0.4	23/02/2022	Aggelos Aggelis (SATWAYS), Leonidas Perlepes (SATWAYS)	DataHub Description
v.0.5	01/03/2022	Antonis Kostaridis (SATWAYS)	Overall design of the platform
v.0.6	04/03/2022	Leonidas Perlepes (SATWAYS)	Consolidated Version
v.0.7	14/03/2022	Katerina Valouma (SATWAYS)	Ready for review
v.0.8	27/03/2022	Vassiliki Varela (KEMEA) Dimitris Stavrakoudis (AUTH)	Review Comments
v.0.85	28/03/2022	Vasileios Kazoukas (KEMEA)	Security Control
v.0.9	30/03/2022	Katerina Valouma (SATWAYS), Leonidas Perlepes (SATWAYS)	Review Comments Consolidated
v.1.0	31/03/2022	Emílio Chuvieco (UAH)	Final version submitted

Quality Control

Type	Date	Reviewed by	Approved/Comment
Peer	23/03/2022	Vassiliki Varela (KEMEA)	Minor changes suggested
Peer	27/03/2022	Dimitris Stavrakoudis (AUTH)	Minor changes suggested
Security	28/03/2022	Vasileios Kazoukas	Approved
SC	31/03/2022	Emílio Chuvieco (UAH)	Approved

Copyright © FirEURisk Consortium. All rights reserved.

Executive Summary

The main objective of the FirEURisk project is to develop, test and disseminate an Integrated and Science-Based Strategy for wildfire risk management in Europe. The various models, tools and data, that will be defined, developed and produced within the project, are considered crucial for this strategy and are expected to be used by the proposed solutions to address the multiple challenges of wildfire management.
According to the project's plan, these models, tools and data will be resulted mainly from the activities of three work packages:

- WP1: "Fire risk assessment to improve prevention"
- WP2: "Risk reduction to improve protection"
- WP3: "Adaptation to future fire regimes"

The efficient sharing of respective information not only among the above FirEURisk work packages but also with the other WPs of the project, will be imperative to the success of the project, being a high priority for the consortium.
In this direction, the internal cooperative platform will ensure the dissemination and availability of the information either produced within the project or existing datasets and information brought in by FirEURisk partners. The management and sharing of that information follow the guidelines presented within the Data Management Plan, as has been developed in the framework of Task A4.1.2 (and planned to be further refined).
The current deliverable (D4.3) presents the main architecture and capabilities of this internal cooperative platform. It is structured as a manual that accompanies the platform, acting as a supportive handbook for the users. This is planned to be updated during the next period of the project, taking into consideration the future refinements of the Data Management Plan and the various data management and integration requirements that will be identified during the upcoming period.

Copyright © FirEURisk Consortium. All rights reserved.

Disclaimer

The content of the publication herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

While the information contained in the documents is believed to be accurate, the authors(s) or any other participant in the FirEUrisk consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the FirEuUrisk Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the FirEurisk Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

Copyright message

© FirEURisk Consortium, 2021-2025. This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged.

Table of Contents

1 Introduction 11
1.1 Purpose of the document 11
1.2 Structure of the Document 11
2 Internal cooperative platform 12
2.1 Platform Deployment 13
2.1.1 Authentication mechanism 14
2.1.2 Open-source frameworks 14
2.2 Platform management / integration 14
3 DataHUB Data Repository 16
3.1 User Authentication 18
3.1.1 User Sign-up 18
3.1.2 User Login 21
3.2 DataHUB UI 21
3.3 “Explore” menu 22
3.4 Spaces 22
3.4.1 Manage Users 24
3.5 Collections 25
3.6 Dataset 27
3.7 Manage Metadata Terms & Definitions 29
3.7.1 Data Visualization 31
3.7.1.1 Shapefiles 31
3.7.1.2 Images 33
3.7.1.3 PDF files 34
3.8 Web Interface / API 36
4 DataHUB Code Repository 38
4.1 Code Repository creation / Code submission 40
5 DataHUB Execution Service 43
5.1 Adding a Tool 43
5.2 Creating Workflows 44
5.3 Executing Workflows 45

Copyright © FirEUrisk Consortium. All rights reserved.

6 Conclusions 46
7 References 47

Copyright © FirEURisk Consortium. All rights reserved.

List of Figures

Figure 1: High-Level architecture 13
Figure 2: DataHUB Data Repository main page....16
Figure 3: The hierarchical structure of the information....17
Figure 4: An example of the FirEUrisk dataset structures....18
Figure 5: Sign up page....18
Figure 6: User verification email example....19
Figure 7: Sign Up Form 19
Figure 8: Sign Up completion notification email 20
Figure 9: User account needs Activation by the administrator....20
Figure 10: User account activation email notification....20
Figure 11: Login page....21
Figure 12: Main navigation pane....21
Figure 13: Logged user's Dashboard 22
Figure 14: "You" dropdown menu....22
Figure 15: A typical Space layout....23
Figure 16: New Space fields....24
Figure 17: User Management in Space....25
Figure 18: Create New Collection menu....26
Figure 19: Collection Layout....27
Figure 20: Dataset creation 27
Figure 21: Add files to the Dataset page 28
Figure 22: Upload progress bar 28
Figure 23: Metadata Terms & Definitions page 30
Figure 24: Geographic data, zip file....31
Figure 25: Shapefile preview on DataHUB 32
Figure 26: Geographic Data, WMS Layer URL 33
Figure 27: Sample Dataset with images 33
Figure 28: Dataset with an image file....34
Figure 29: Dataset with a PDF file....35
Figure 30: Dataset with a PDF file....36
Figure 31: View Profile link....37
Figure 32: API Key creation....37
Figure 33: DataHUB Code Repository main page....38
Figure 34: DataHUB Code Repository user registration 39
Figure 35: DataHUB Code Repository user login....39
Figure 36: DataHUB Code Repository logged user Dashboard....40
Figure 37: DataHUB Code Repository New Repository menu....40
Figure 38: DataHUB Code Repository New Repository form part 1/2 41

Copyright © FirEUrisk Consortium. All rights reserved.

Figure 39: DataHUB Code Repository New Repository form part 2/2 41
Figure 40: DataHUB Code Repository New Repository initial state 42
Figure 41: DataHUB Code Repository Sample project page 42
Figure 42: DataHUB Execution Service Main UI 43
Figure 43: Main menu buttons 43
Figure 44: Upload a new tool dialog 44

Copyright © FirEURisk Consortium. All rights reserved.

List of tables

Table 1: List of Acronyms.....10

Copyright © FirEURisk Consortium. All rights reserved.

List of Acronyms

Table 1: List of Acronyms

List of Acronyms	
API	Application Programming Interface
DataHUB	FirEUrisk internal cooperative platform
DMP	Data Management Plan
HTTP	Hypertext Transfer Protocol
REST	Representational state transfer
UI	User Interface
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UUID	Universally unique identifier
GIT	A free and open-source distributed version control system

Copyright © FirEUrisk Consortium. All rights reserved.

1 Introduction

The FirEURisk project is planned to take advantage of several models, tools and data, that will be produced by WP1, WP2 & WP3, in order to propose solutions, recommendations and an overall science-based strategy for wildfire risk management in Europe. The management of this information internally in the project is planned to be supported by a cooperative platform. Through this platform, partners will be able to store and share the models and tools that have been produced within the project and be informed about existing information shared by other partners.
The platform provides a user-friendly graphical user interface, where users will be able to execute all provided functionalities. Additionally, an interface (REST API) is provided, enabling the execution of the functionalities by external authenticated systems.
The current deliverable and the internal cooperative platform that is documented have been resulted by the activities of Task A4.1.3. This is part of the WP4 activities that are related to the integration of the different phases and scales of fire risk management. The information sharing and data management mechanism follows the Data Management Plan (DMP) that has been defined in deliverable D4.2 – "Data Management Plan".
Additionally, the internal cooperative platform is planned to be the basis of the open platform of models, data and knowledge that is planned to be implemented within Task A4.3.2.

1.1 Purpose of the document

The internal cooperative platform has been designed and deployed on a cloud infrastructure. It is ready for use. The purpose of this deliverable is to accompany the deployment, presenting the high-level architecture and the capabilities of the platform. Moreover, it is planned to be used as a training material that could be used by the users (scientists, stakeholders etc), in order to be informed about the platform and the provided functionalities and to be used as a guide for the execution of the functionalities.

1.2 Structure of the Document

The structure of the document has been selected to fulfil the twofold scope of the deliverable. Its content is separated into two main sections: Section 2 provides a high-level presentation of the platform, whereas the next sections are planned to be used as manual provided to the users, in order to be informed about the provided functionalities and how to execute these. In more detail, Section 3 presents the characteristics of the data repository. Section 4 provides details about the code versioning functionality (GIT service) and Section 5 presents the execution service.

Copyright © FirEURisk Consortium. All rights reserved.

2 Internal cooperative platform

FirEURisk project has to deal with a number of scientific models, data and tools that have to be designed, analysed and used by the partners in order to produce the science-based results of the project. A cooperative platform has been designed and developed by Satways Ltd, in order to manage the storing and sharing of this information internally in the project. The scope of this internal cooperative platform is threefold:

- The management of the scientific data/models/knowledge that will be used within the project. This information will be produced either internally by the project activities, or by external systems.
- The management of the tools that will be used within the project.
- The execution of some of these procedures.

The name of this platform is **DataHUB** and it comprises three subcomponents, each of these supports one of the required capabilities. The three subcomponents are:

- DataHUB Data Repository: it enables the management of scientific information.
- DataHUB Code Repository: it supports the management and versioning of the tools/programming codes.
- DataHUB Execution Service: it enables the analyses' execution.

These subcomponents are combined using a loosely coupled architecture, providing the platform capabilities, such as extensibility and adaptability to the users' needs, that are required by the project. The functionalities are available to the users through an intuitive graphical user interface. Additionally, several functionalities are available for execution by external systems, using the API interface that is provided.

The platform has been deployed on the cloud using several open-source technologies. It is available on the following URL:

https://fireurisk.satways.net

The following figure depicts the high-level architecture of the platform.

Copyright © FirEURisk Consortium. All rights reserved.

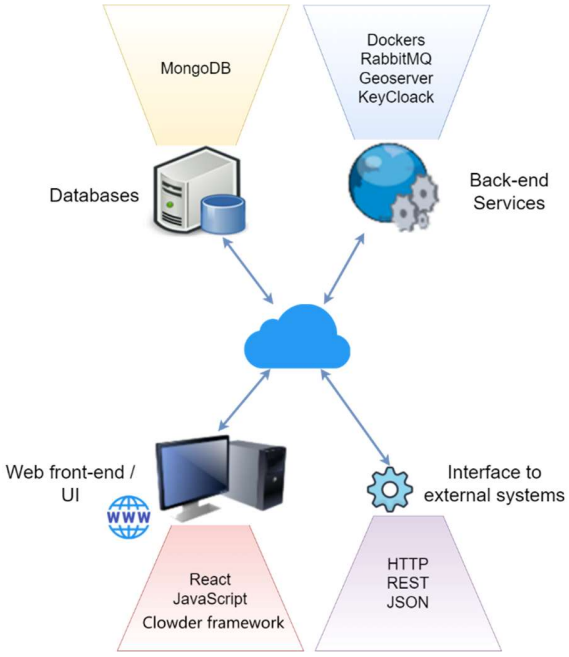


Figure 1: High-Level architecture

The design of the platform is separated into four main layers, according to the type of operations that are provided:

- The database: it manages the storing of the data/codes/information. An instance of the Mongo database [1] is used for that purpose.
- The Back-end services: it provides the core functionalities of the platform. The main logic of the platform as well as the users' authentication mechanism are supported by this layer. It is based on a number of technologies, such as Dockers [2] (the main computational unit that implements the main logic of the platform), Geoserver [3] (management of the geospatial information), KeyCloack [4] (management of users' authentication) etc.
- The web front-end UI: it provides the web graphical user interface that makes available all provided functionalities to users. It is based on the open-source Clowder framework [5]. Several web technologies, such as JavaScript [6], React [7][7] and Scala [8], have been used.
- The interface to external systems: it enables the execution of the functionalities by external systems. An HTTP REST API [9] is provided in order to support this.

2.1 Platform Deployment

The platform has been deployed on the cloud. The Docker technology has been used for that purpose. Docker is an open application development framework that's designed to benefit DevOps and developers. Using Docker, developers can easily build, pack, ship, and run applications as lightweight, portable, self-sufficient containers, which can run virtually anywhere. Containers allow developers to package an application with all of its dependencies and deploy it as a single unit [10]. Using this technology, the management and the configuration of the platform are efficient, enabling

Copyright © FirEURisk Consortium. All rights reserved.

the easy update/alteration of the platform capabilities, in terms of available memory, hard disk capacity, computation capabilities etc. Thus, the deployment of the platform will be easily adapted to the requirements (memory, hard disk capacity etc.) that will be identified during the next periods of the project.
Currently, the characteristics of the deployment have been set to 3 CPUs, 8 GB of RAM and 2 TB of storage. However, during the upcoming period, this configuration will be updated, improving the characteristics of the deployment according to the requirements of the partners and the tools/models/data that will be developed. For example, it is envisaged that within the next period more than 10 TB of storage will be required.

2.1.1 Authentication mechanism

The platform supports various authentication mechanisms, in order to prevent unauthorized users to have access to the stored information. Users are free to register to the platform using unique credentials (username/password). Platform administrators must approve (or not) their registration. Additionally, authentication protocols, such as OAuth1, OAuth2 and OpenID are supported, enabling the users' authentication through external applications, such as Twitter, Facebook, Google etc. These authentication protocols will be useful to the open version of the platform (Task 4.3.2), where external users should be authenticated and take access to the platform.
Specific access rights should be assigned to each authenticated user, taking into consideration their role and responsibilities, in order to control the access to the information by users with limited access rights.

2.1.2 Open-source frameworks

The design and development of the platform took advance of two open-source frameworks, providing the platform with the required functionalities and capabilities. These two frameworks are:

1. The Gitea framework [11] was used as the basis for the code repository. It provides the GIT functionality [12] that is required to manage and have version control on the code of the various tools, models and algorithms that will be uploaded to the platform.
2. The Clowder framework [5] was used as the basis for the data repository. It provides data management functionalities (upload, sharing, storing etc.). The provided are available to the users through a web UI and an API.

The platform has been built on top of these frameworks, extending these and developing new functionalities in order to fulfil the project's requirements and the Data Management Plan that has been defined.

2.2 Platform management / integration

The configuration and management of this platform are supported by the Satways Administrative Team. Members of this team are:

- *Dr. Leonidas Perlepes: l.perlepes@satways.net*
- *Mr. Aggelos Aggelis: a.aggelis@satways.net*
- *Dr. Antonis Kostaridis: a.kostaridis@satways.net*

The management of information and its metadata is based on the Data Management Plan that has been prepared by Task A4.1.2 "Interoperability and data management". As this plan is a living document, an update that may be defined

Copyright © FirEURisk Consortium. All rights reserved.

during the upcoming period will be further analysed in order to check if it is compliant with the platform operations. Minor changes to the platform operations are envisaged to be requested during the upcoming period.

In addition, the platform is planned to provide input to Task A4.3.1 “Integrating FirEURisk resources with other EU platforms and legacy systems”. Within this task, a tool will be built to display outputs from the internal platform on the web and also on most mobile devices.

Finally, the internal platform is planned to be used as the basis for the “Open platform of models, data and knowledge” that will be built within Task A4.3.2.

Copyright © FirEURisk Consortium. All rights reserved.

3 DataHUB Data Repository

The DataHUB Data Repository enables the sharing, organization and analysis of large datasets, models and knowledge. Using their preferred browser, users can log in and be informed about the datasets that have been uploaded by other users or manage their list of datasets. Additionally, a fully functional API is provided, enabling the execution of the framework functionalities and the usage of the information that has been uploaded to it, through any external tool. The DataHUB Data Repository is based on the Clowder open-source framework.

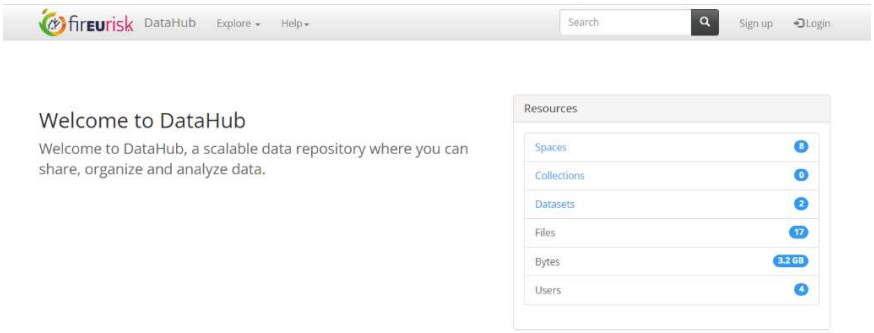


Figure 2: DataHUB Data Repository main page

The component follows a hierarchical structure in order to organize the information that has been uploaded to it. Three main groups are used in order to organize the information. These three groups are:

- **Spaces**: This is the main entity that groups the information/datasets. Internally, it hosts a number of collections and data sets with defined user access rights.
- **Collections**: Collections are used to hierarchically structure the information that is contained in a Space. Multi-level collections (child collections) are supported inside Spaces. This way, the collections could be used in order to further group the datasets. Each collection could host either a number of child collections or/and datasets.
- **Datasets**: This is the entity that can directly host files uploaded and managed by the users. Each dataset is coupled with a number of metadata definitions that specify its characteristics. Each dataset can include folders in order to better organize the hosted data.

This hierarchy is very flexible and allows the sharing of Datasets among other groups. More precisely any Dataset can be shared among multiple Collections, and even among different Spaces. Thus, avoiding the duplication of the same datasets and minimising storage requirements.

Copyright © FirEURisk Consortium. All rights reserved.

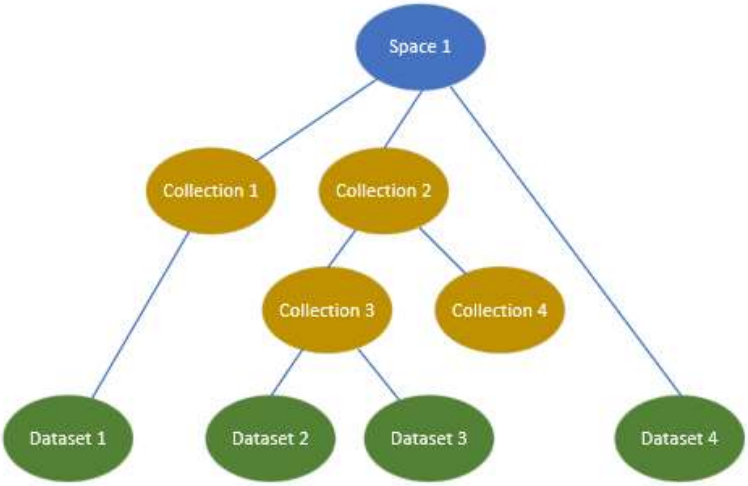


Figure 3: The hierarchical structure of the information

In the context of the FirEUrisk project, an initial hierarchical structure has been proposed. According to this initial structure:

- Each Space, groups the datasets that are related to each Pilot Site (PS). As a result, five different Spaces have already been created in order to host the data of the 5 pilots.
  - One additional Space has been created, in order to host the datasets that will be common to all pilots (e.g., at the pan-European level)
- Each Collection, groups the datasets that are related to a specific tool/model. Different cases/scenarios per tool could be supported through the flexible hierarchic structure of the Collections.
- Datasets should group the actual files that are required or produced by each tool.

The following figure depicts an **example** of the FirEUrisk dataset structures.

Copyright © FirEUrisk Consortium. All rights reserved.

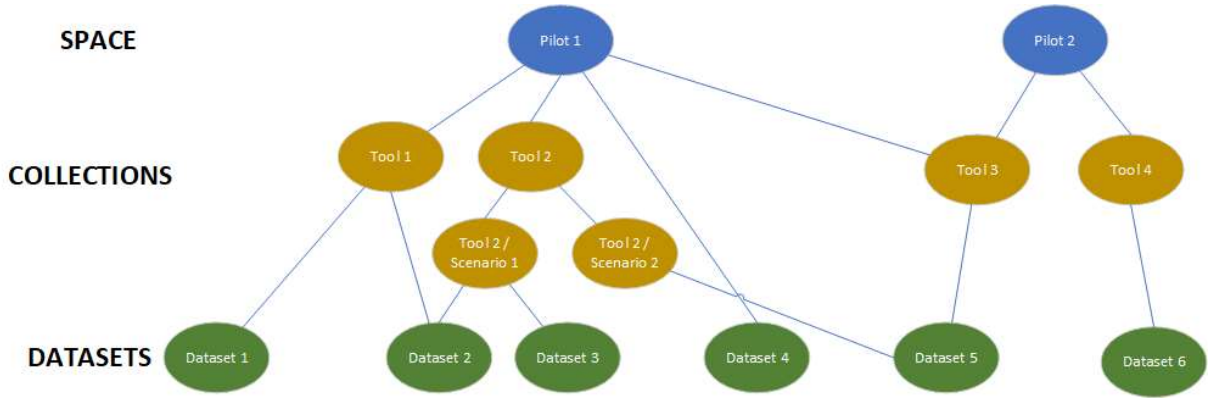


Figure 4: An example of the FirEURisk dataset structures

The hierarchical structure of the information will be further analysed and refined during the upcoming period of the project. The results of various activities (such as A1.3.1, A5.1.2 & A5.1.3, as well as the ongoing integration and cross-packages activities) will be taken into consideration in order to conclude the final structure of the information.

3.1 User Authentication

In order to access the platform, each user has to be registered. The platform supports a three-level authentication mechanism, allowing only verified users to access the data.
The first step that each user must perform is registration.

3.1.1 User Sign-up

To register on the platform a user must access the top right part of the navigator, and select the “Sign up” link. Using the following page (Figure 5), users must provide their emails.

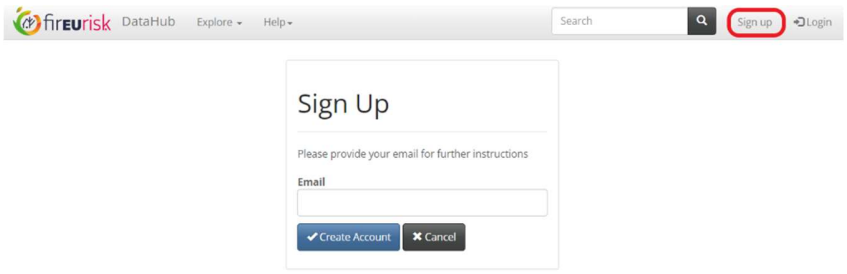


Figure 5: Sign up page

Upon the selection of the “Create Account” button, an email will be forwarded automatically to the email account.

Copyright © FirEURisk Consortium. All rights reserved.

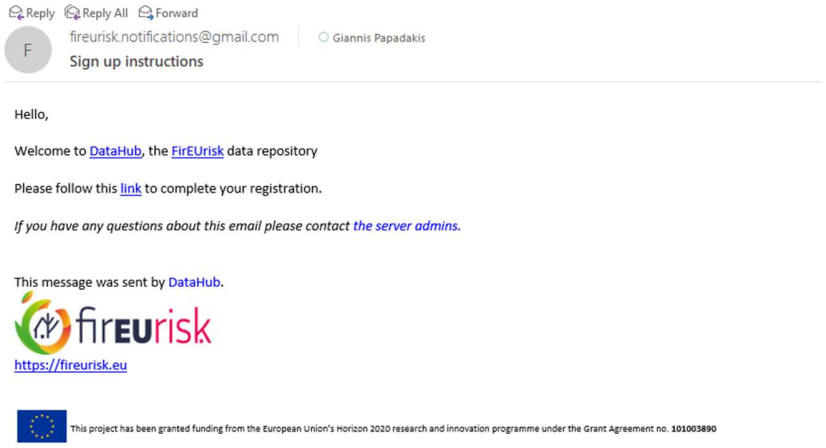


Figure 6: User verification email example

Look for the "Please follow this link to complete your registration" line and click the provided link. The final registration page should appear in the user's browser (Figure 7).

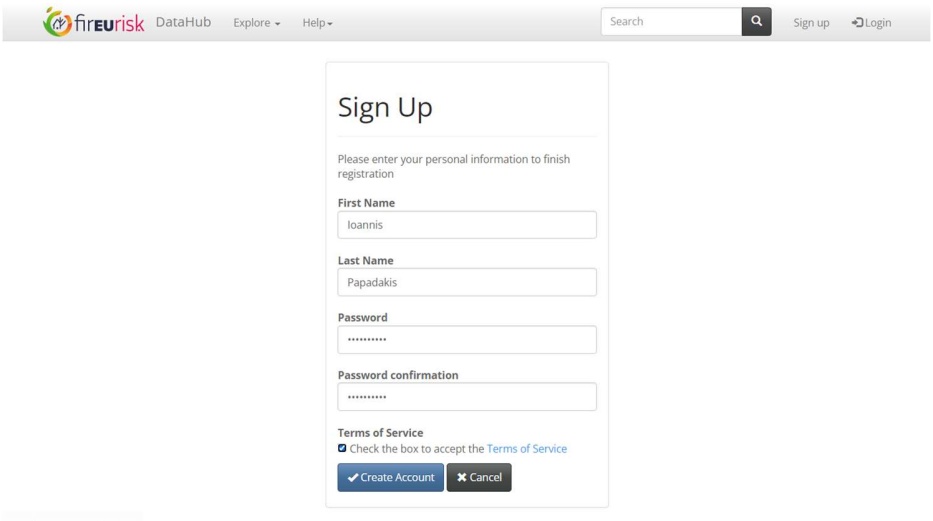


Figure 7: Sign Up Form

The users must fill in the form with her/his first and last name, as well as a valid password, check the Terms of Service checkbox and then press the Create Account button. This button is activated only if the Terms of Service are accepted. A new email (Figure 8) will be automatically sent to the user, verifying the completion of the registration process.

Copyright © FirEUrisk Consortium. All rights reserved.

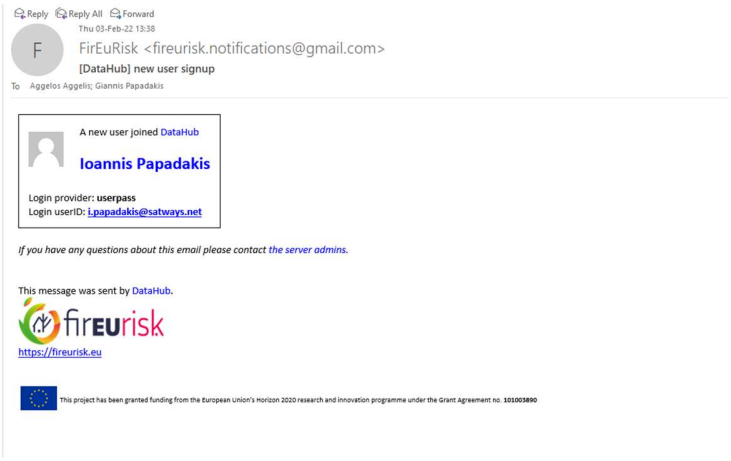


Figure 8: Sign Up completion notification email

One final step remains, and this must be performed by the DataHUB system admins. At this point, the account is successfully created but not activated (Figure 9). Therefore the user has reached a page like the following.

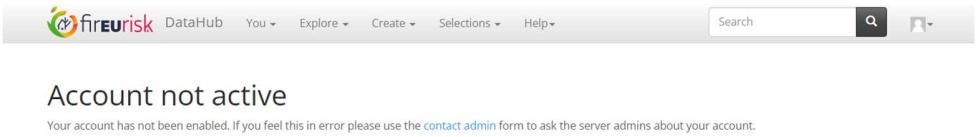


Figure 9: User account needs Activation by the administrator

As soon as the system admins activate the account, a confirmation email (Figure 10) will be automatically sent to the user and then it is possible to be logged into the platform.

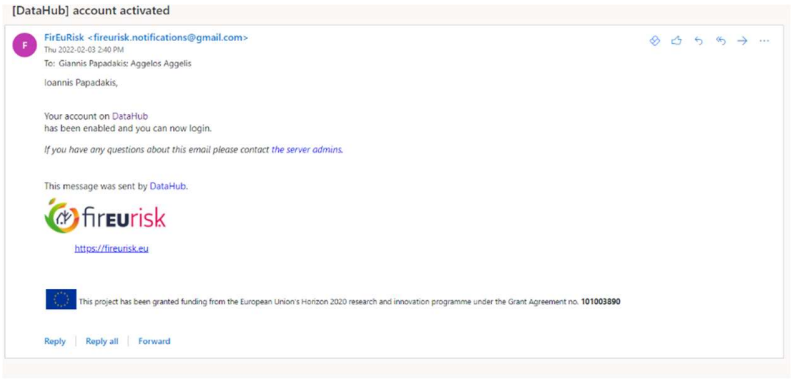


Figure 10: User account activation email notification

The user will receive the following message: "Congratulations! Now you can fully use the DataHUB Data Repository!"

Copyright © FirEURisk Consortium. All rights reserved.

3.1.2 User Login

Once a new user has completed successfully her/his registration, she/he can log into the platform. This is feasible by selecting the "Login" link on the top right part of the navigator. Then, the login page will be displayed (Figure 11), waiting for the user's credentials. After filling in the credentials, the user should press the "Login" button to get authorized and proceed to the logged user's home page.

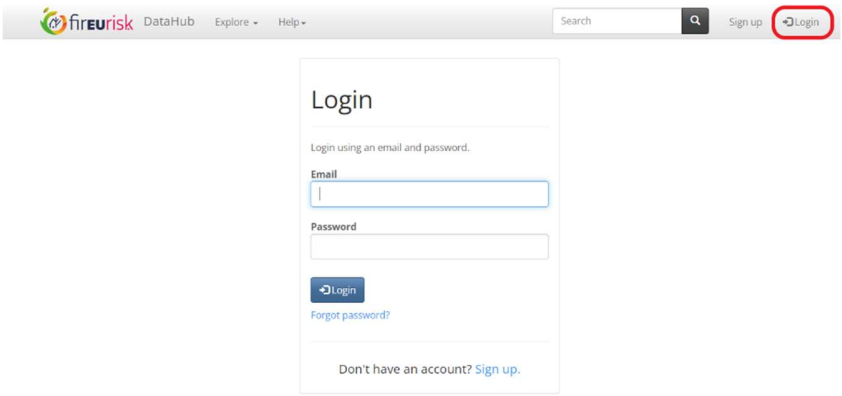


Figure 11: Login page

3.2 DataHUB UI

After logging in, the user is landed on the DataHUB Data Repository home page. User can return to this page, anytime by clicking on the DataHUB link in the top navigation. The navigation pane (Figure 12) offers the main navigation links for any action on the platform and the data stored.

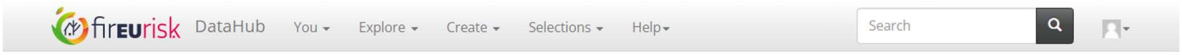


Figure 12: Main navigation pane

Initially, a dashboard is provided to the user, informing her/him about the history of her/his actions and the groups of information that have rights to access. Through this dashboard, the user is also able to create new Spaces/Collections/Datasets and upload new files.

Copyright © FirEUrisk Consortium. All rights reserved.

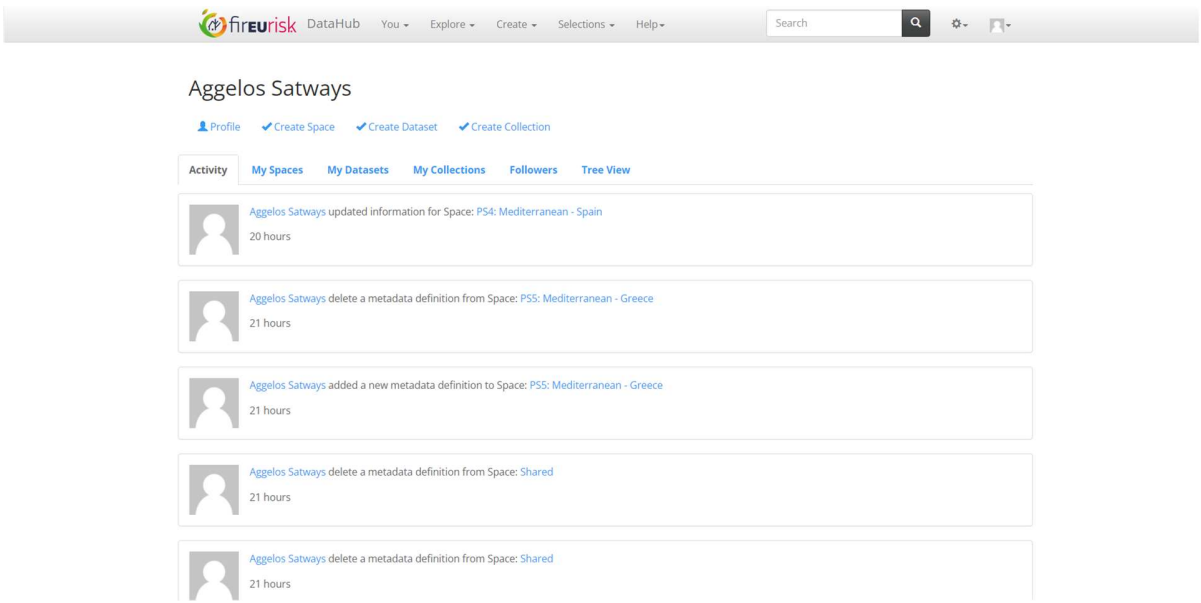


Figure 13: Logged user's Dashboard

This dashboard is available to the user through the "You" menu.

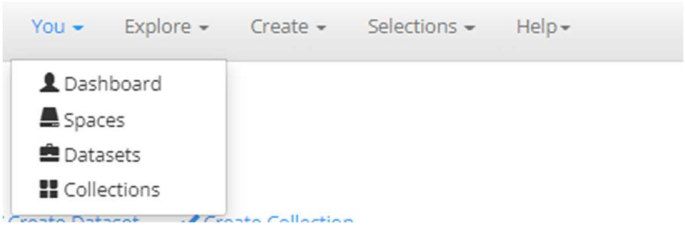


Figure 14: "You" dropdown menu

The Dashboard displays events on Datasets, Collections, Spaces, files and users that the user follows. Any comments added, or any file added or removed from a dataset is noted here.

3.3 "Explore" menu

Additionally, all the Spaces, Datasets and Collections that a user has access to can be seen through the "Explore" menu. A user can have access to Space, Collection or Dataset, either due to ownership (e.g., the user himself/herself has created a Space) or because the data (Space/Collection/Dataset) administrator has invited the user in.

3.4 Spaces

Each Space page presents apart from the Space title a logo and a short description of the Space contents. Next, the Datasets and Collections that are available in the Space are listed. By selecting any of them, the corresponding Collection and Dataset pages will load. On the right part of the page, some statistics are provided regarding the data

Copyright © FirEURisk Consortium. All rights reserved.

stored in the selected Space, and the users that have access to them. Additionally, several management functionalities (manage users, edit Space characteristic, etc.) are provided only to users with the related access rights.

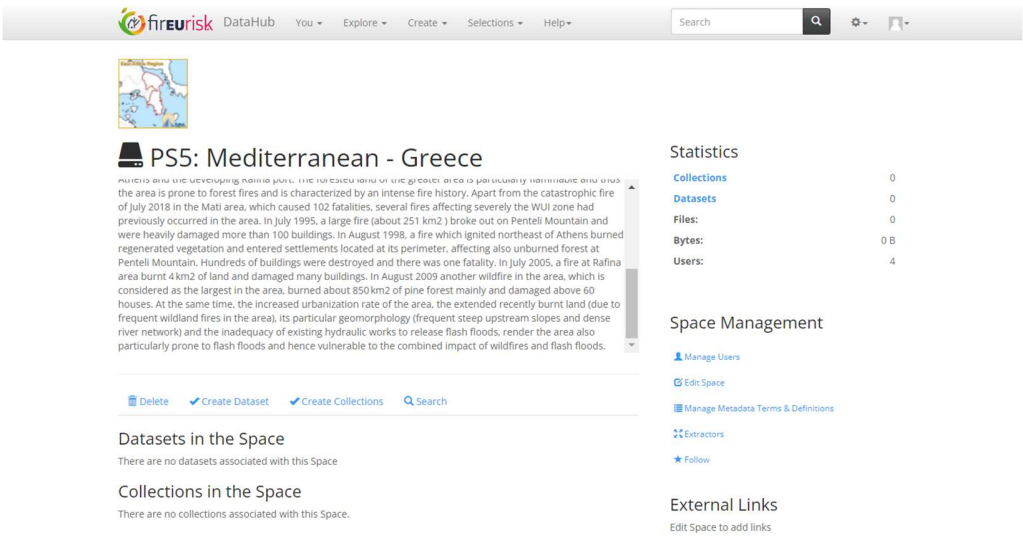


Figure 15: A typical Space layout

By selecting the creation of a new Space, users can fill in all the information that is required in order to describe the purpose and content of the specific Space.

Copyright © FirEUrisk Consortium. All rights reserved.

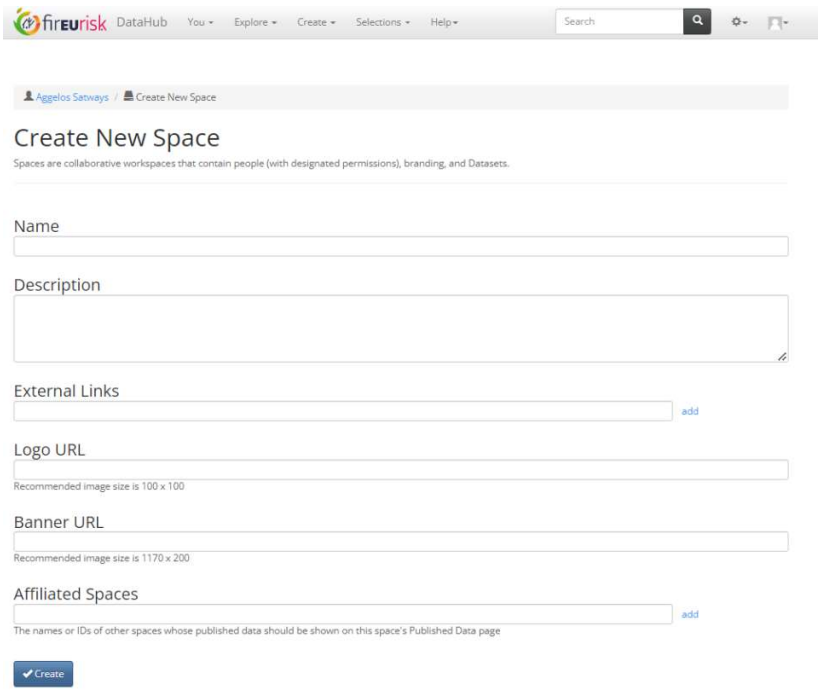


Figure 16: New Space fields

3.4.1 Manage Users

Any user with admin rights on a Space can permit other users to access the data stored in the Space, or to provide them admin access rights. The access right of the users is defined by the role that has been assigned to them per Space. There are three roles available:

- **Admin**: A user with admin rights on a Space is free to perform all operations on the Space and the Collections and Datasets stored there.
- **Editor**: As an Editor, a user can upload/download data from the existing Space but cannot either invite users or delete entities.
- **Viewer**: As a viewer, a user is just able to see the Space data hierarchy as well as the files that are uploaded. No further actions like download can be performed by a user with this role.

Using the “Manage Users” menu, the administrators of each Space can assign specific roles to users.

Copyright © FirEURisk Consortium. All rights reserved.

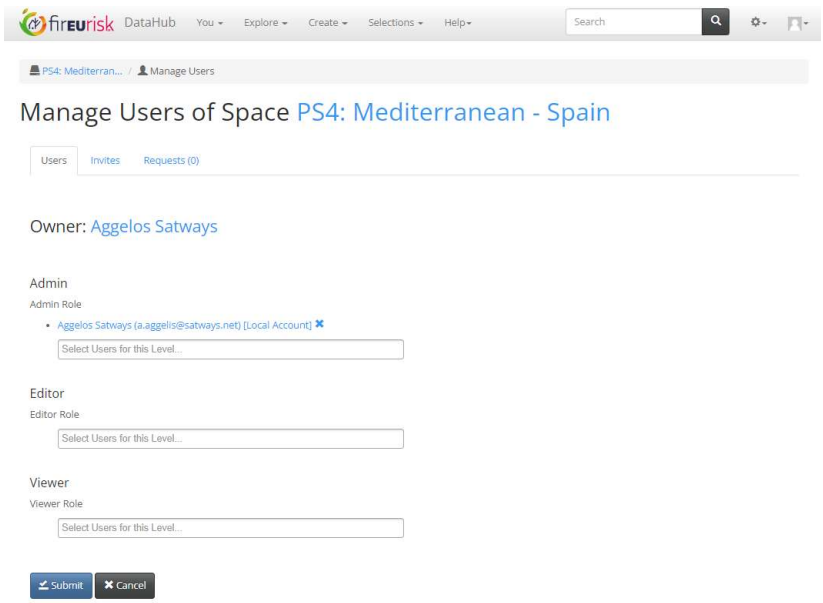


Figure 17: User Management in Space

3.5 Collections

The “Create New Collection” menu enables the creation of a new Collection. The following figure depicts the form that requests the initial information about the new Collection. Users are called to fill in the name of the new Collection, a short description and the Space that the new Collection could be attached. In fact, more than one Spaces can be selected, in order to allow the new collection to be shared among many Spaces.

Copyright © FirEURisk Consortium. All rights reserved.

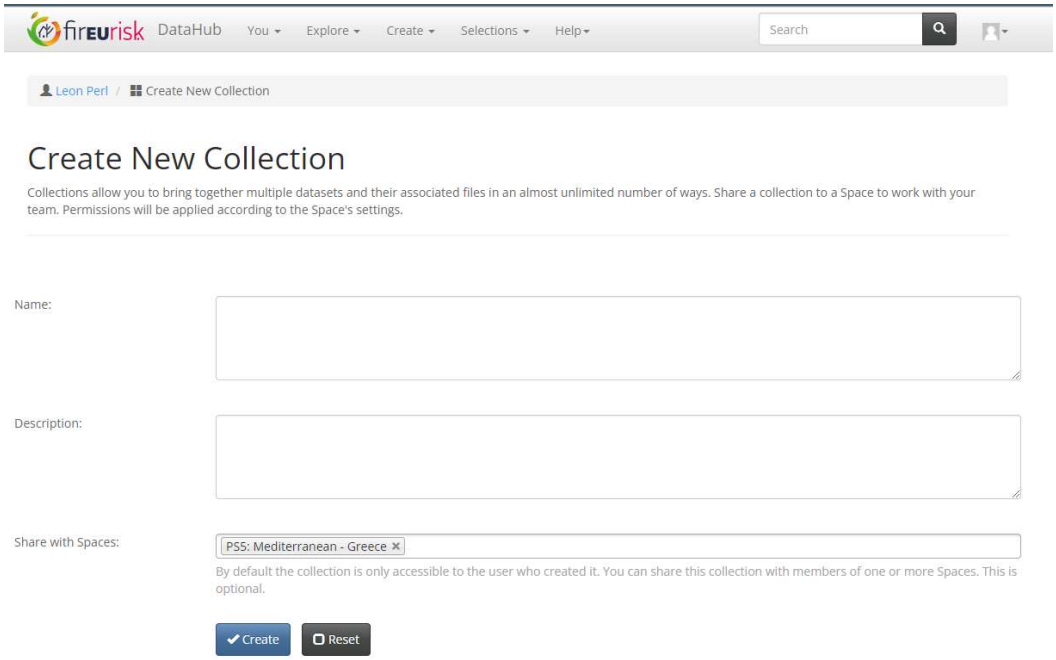


Figure 18: Create New Collection menu

By selecting the "Create" button, the new Collection is automatically created. The following figure depicts the layout that is used in order to depict the information of a Collection. Through that, users (with the related access rights to the Space where the Collection has been created) can be informed about the characteristics of the Collection and the data that are attached. Moreover, users can manage the Collection, either by attaching new Collections or Datasets to it or by assigning this to one (or more) alternative Spaces.

Copyright © FirEUrisk Consortium. All rights reserved.

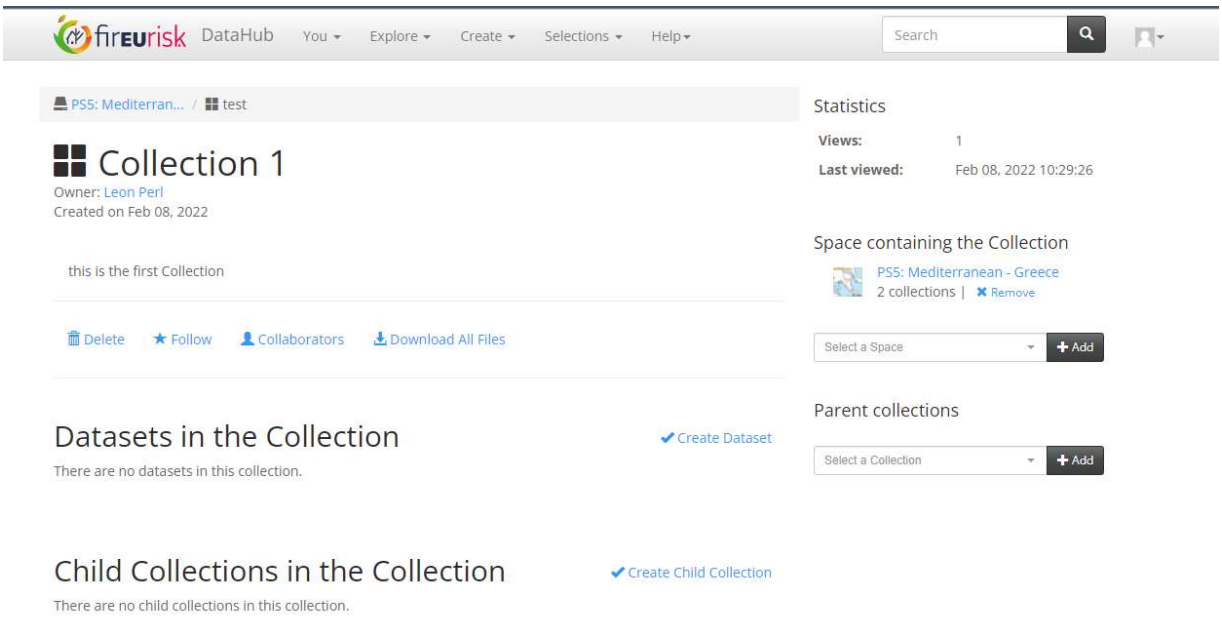


Figure 19: Collection Layout

3.6 Dataset

On a Dataset creation, the user must define a name and a short description. A Dataset can be shared among more than one Space if needed.

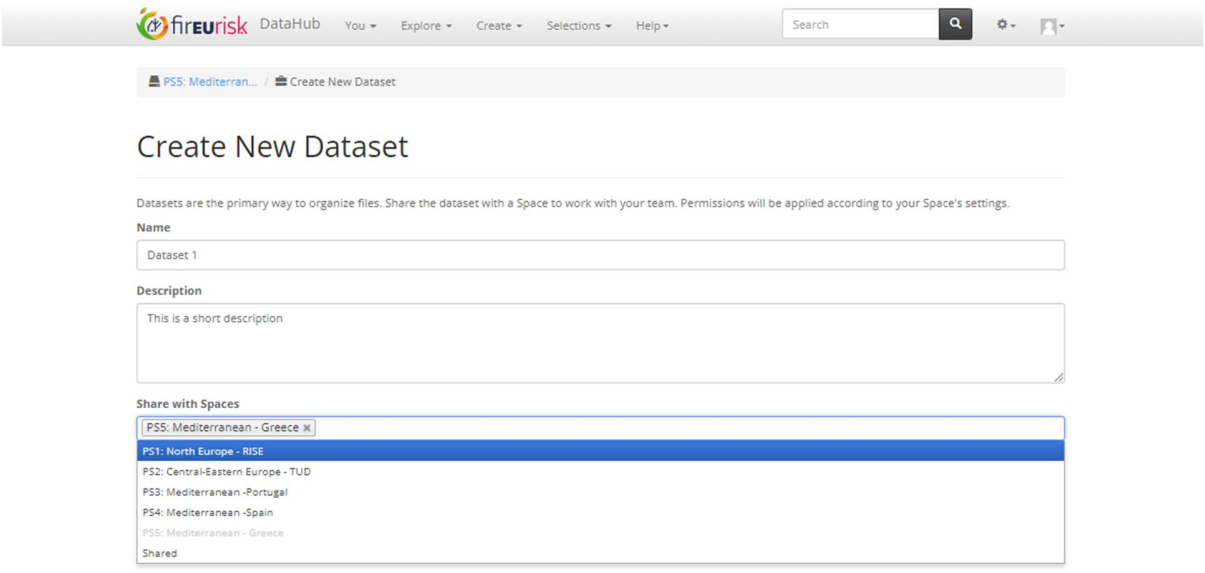


Figure 20: Dataset creation

Copyright © FirEURisk Consortium. All rights reserved.

To upload new files on a Dataset, the user must select the + Add Files link, and then the following page (Figure 21) will be displayed. From this page, the user can select the files that would like to upload, either by Drag & Drop any files from its own filesystem explorer directly on this page, or by selecting the "Select Files" button, to trigger the file, selection wizard. When the files are added, they can either be uploaded one by one by selecting each "Start" button or all together, by selecting the "Upload" button.

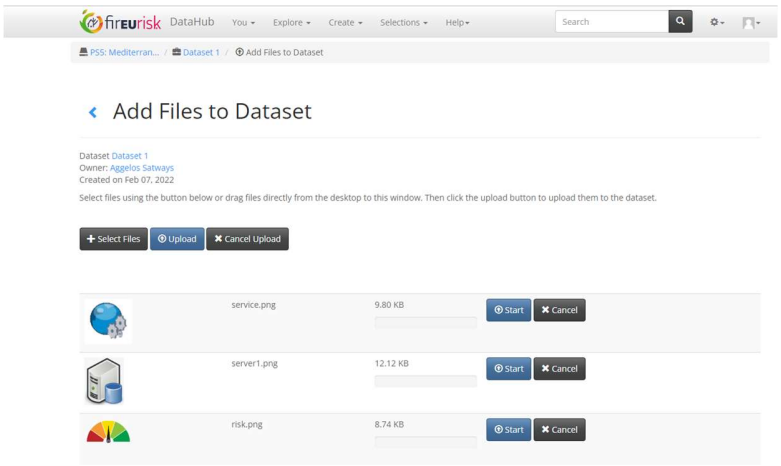


Figure 21: Add files to the Dataset page

While the files are uploaded a green progress bar is displayed. It is obvious that the uploading time depends on the file size and the connection speed. In any case, the browser tab **must** remain open while the progress bar is visible, even if the completion percentage reaches 100%. If the user closes the page, before the progress bar disappears, then the file will most likely fail to get uploaded!

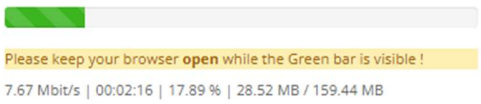


Figure 22: Upload progress bar

Copyright © FirEURisk Consortium. All rights reserved.

Info

In the following table there are some examples of total time to upload calculations, according to some sample file sizes and upload speeds.
With an internet connection of 100 Mbps, and an upload speed of 10 Mbps, a file with 1GB size would take approximately 15 mins to upload!

File Size	Upload speed	Total time
1GB	5 Mbit/s	00:28:36
1GB	10 Mbit/s	00:14:18
3GB	10 Mbit/s	00:42:56
3GB	100 Mbit/s	00:04:17

3.7 Manage Metadata Terms & Definitions

Spaces and Datasets group of information can be coupled with a set of metadata (characterizing each one of these). The list of available metadata fields that could be used per group of information can be dynamically updated either by the platform administrators or by the admin users of each group of information. In the first case, the list of the available metadata fields could be managed centrally in order to be available to all groups of information that has been created for the platform. In the second case, each user can manage the list of metadata fields that could be used for the Spaces and Datasets that the user is the owner of.
On the Metadata Terms & Definitions page, any user with admin rights over a Space can define the metadata terms that will be available for the current Space's Datasets and files. The already existing metadata terms, if any, are presented at the top of the page. A user can add a new term by filling the corresponding fields at the bottom of this page. For any new term, the following fields can be filled:

- **Label**
- **Description**
- URI
- Type
- Defined By
- Definitions URL
- Query Parameter

Among these attributes, only the Label and the Description are obligatory in order to define a new metadata term.

Copyright © FirEURisk Consortium. All rights reserved.

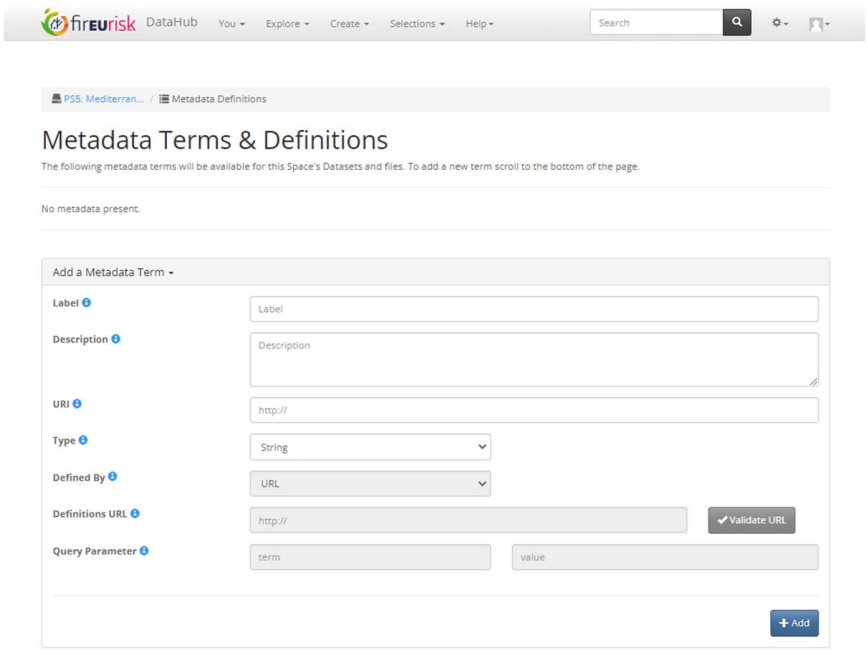


Figure 23: Metadata Terms & Definitions page

According to the current version of the DMB, each dataset must be accompanied by the proper formal metadata, otherwise, the data will not be stored on the platform. All spatial datasets will have INSIRE-compliant metadata [13], whereas non-spatial ones will comply with other proper metadata schemas.

The DMB is investigating all possible solutions for automating the metadata generation process, wherever that is possible. This will be achieved by defining a static metadata template for each FirEUrisk product or dataset, with the platform filling in the metadata elements that change across different files by appropriately parsing the dataset filename. For datasets where a fully automated metadata generation mechanism cannot be set up, the uploader will be responsible for decorating the dataset with the proper metadata. When uploading data through the DataHUB UI, the user will be redirected to a metadata form after uploading the dataset, mandating to fill at least the mandatory elements. When uploading data via the API, a dedicated API endpoint for attaching dataset metadata will be used. Metadata elements that do not change across different files of the same FirEUrisk product/dataset (e.g., keywords) will be prefilled by the platform. The exact mechanisms of these processes will be implemented after their definition by the DMB and will be documented in the updated version of this deliverable. In all cases, the metadata will be validated against the equivalent metadata schema. If the validation fails, the dataset will not be stored in the repository and an appropriate message will be displayed to the uploader (or an appropriate response if using the API).

Due to the above metadata requirements, any partner needing to upload data on the DataHUB Data Repository must first communicate with the DMB, in order to agree on the metadata format and requirements. If no metadata is already defined by the partner, the DMB will assist in the selection of the appropriate metadata schema and the identification of static metadata elements or template for the automatic generation, after discussion with the partner.

Copyright © FirEUrisk Consortium. All rights reserved.

The process of metadata definition is highly dependent on the content of the Data Management Plan. So, any update of the plan during the next phase of the project may trigger an update also in this specific functionality.

3.7.1 Data Visualization

The DataHUB Data Repository is agnostic to file types, that can be uploaded. For some well-known file types, their automatic visualization on the page has already been implemented.

3.7.1.1 Shapefiles

Uploading geographical data in shapefile format can be supported on the DataHUB. The platform, in order to automatically support the visualization of a shapefile, requires these files to be contained in a .zip file. In more detail, the .zip file should contain at least the following four file types:

1. .dbf file
2. .prj file
3. .shp file
4. .shx file

In the following example, a shapefile with the Greek borders is used. A zip file (Figure 24), including 4 required files, is prepared and uploaded to a sample dataset.

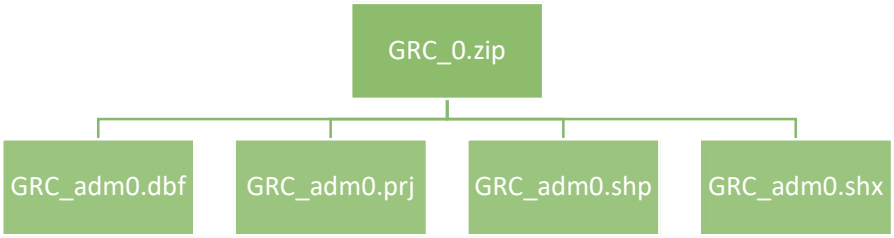


Figure 24: Geographic data, zip file

Copyright © FirEURisk Consortium. All rights reserved.

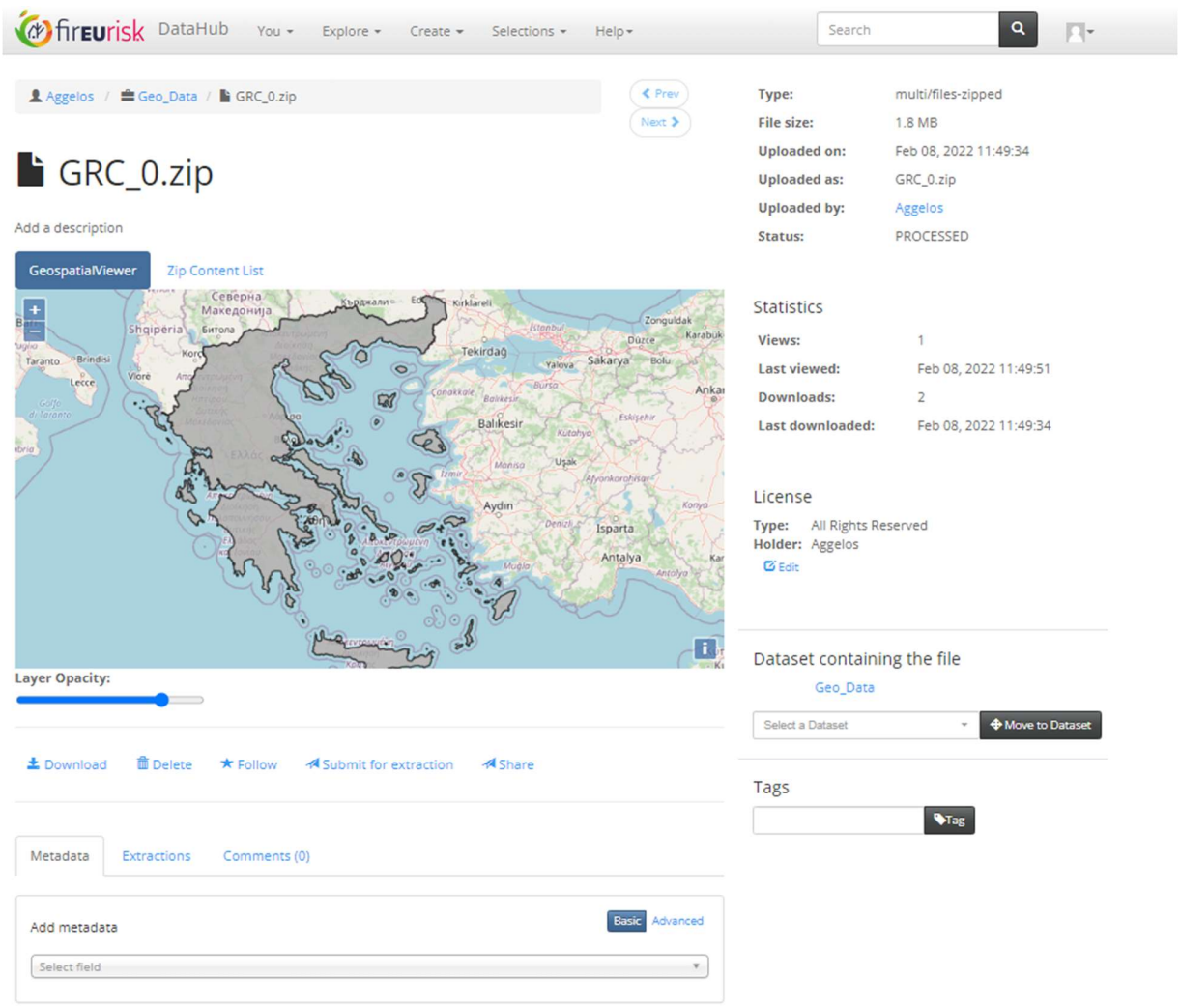


Figure 25: Shapefile preview on DataHUB

Upon the submission of the .zip file, the platform automatically parses the files and creates a representation of that information on the map. The information regarding geographical layers is uploaded on a Geoserver that is available at the following URL: https://geoserver.fireurisk.satways.net

Copyright © FirEUrisk Consortium. All rights reserved.



Figure 26: Geographic Data, WMS Layer URL

3.7.1.2 Images

In this section, a dataset with sample image files uploaded is presented. Following the typical process for upload, the user can upload multiple image files. Once the upload finishes, the Dataset page shows all the contained files. For this specific case of image files, small thumbnail previews are available for each file (Figure 27).

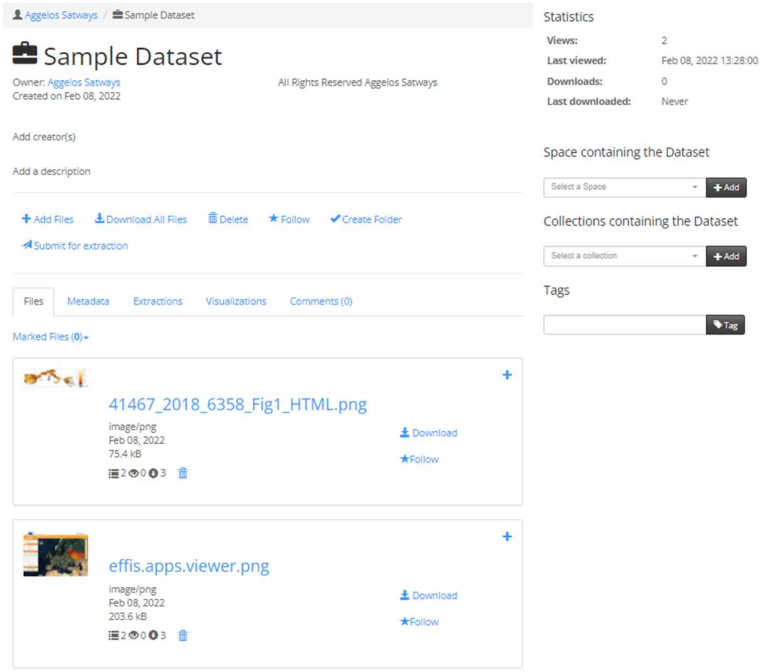


Figure 27: Sample Dataset with images

When the user selects any file of the dataset, a new page is loaded, presenting a preview of the file on the top of the page, among other details and metadata.

Copyright © FirEURisk Consortium. All rights reserved.

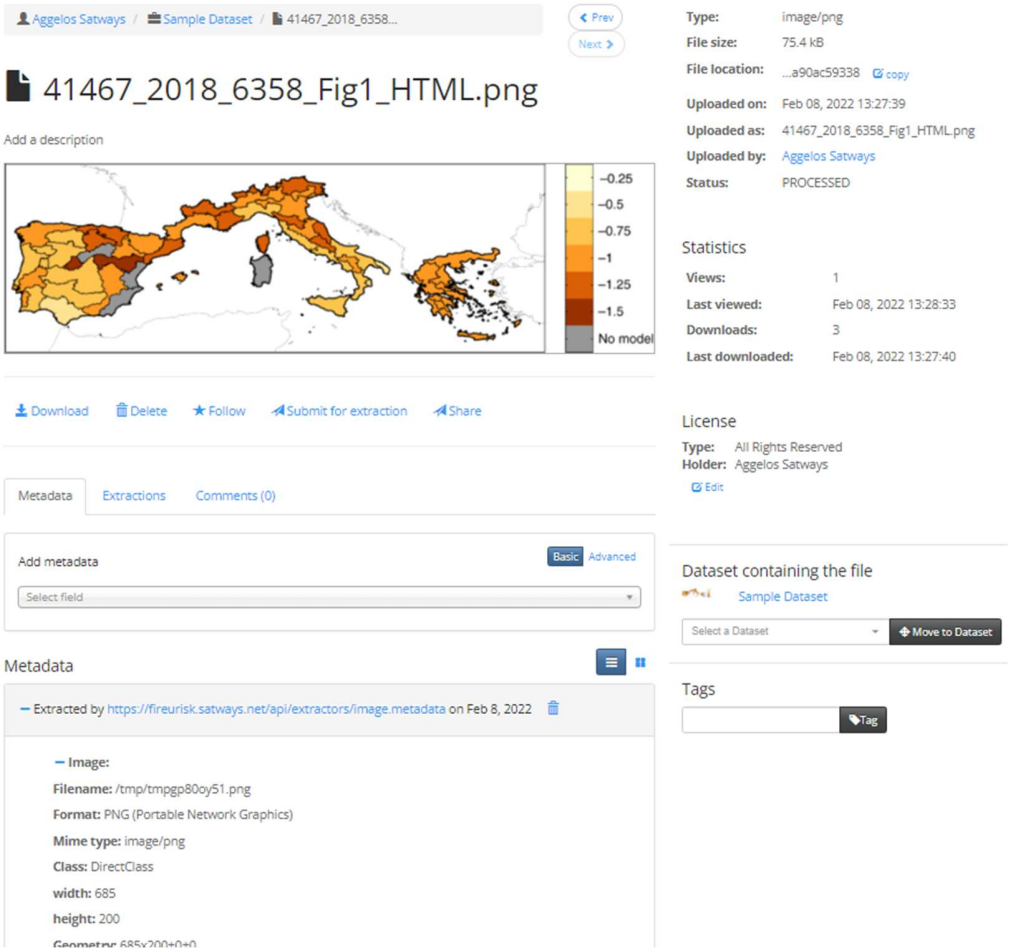


Figure 28: Dataset with an image file

3.7.1.3 PDF files

Another file type that is already recognized by the DataHUB is the PDF type. Once a PDF file is uploaded on the platform, a preview is prepared to make the uploaded data more distinguishable by the user. In Figure 29, the page of a Dataset with a PDF file is presented.

Copyright © FirEURisk Consortium. All rights reserved.

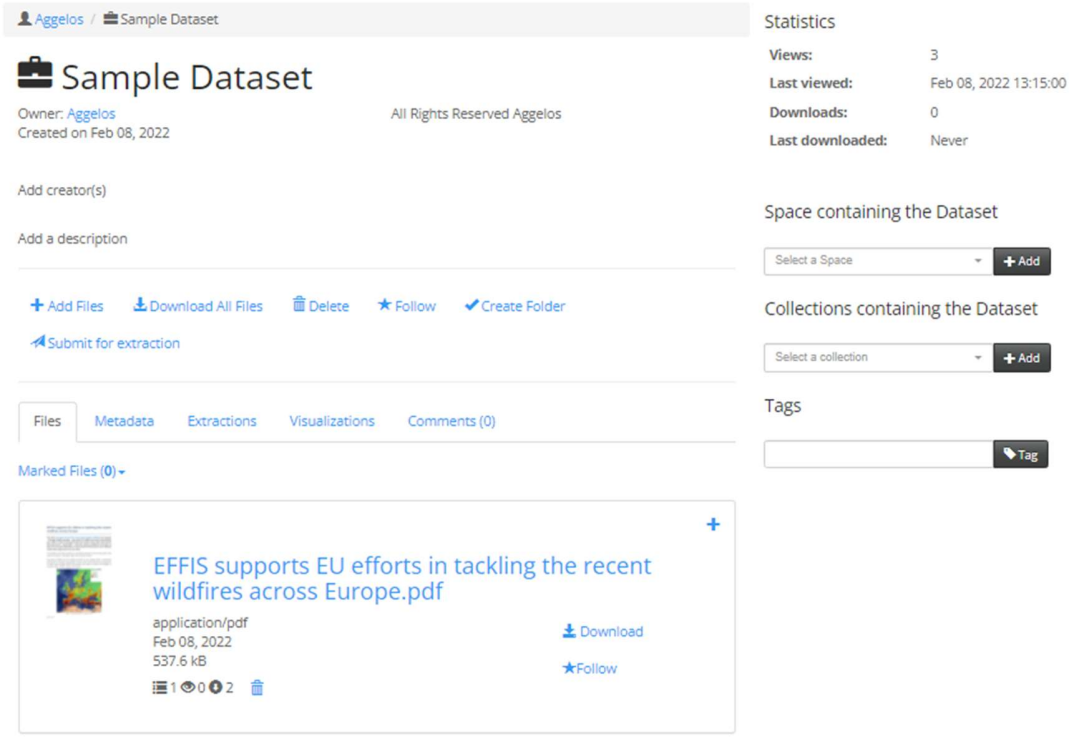


Figure 29: Dataset with a PDF file.

When a user selects the pdf file from the files list of the dataset, a dedicated file, page, is loaded. At the top of this page (Figure 30) a preview of the PDF file is displayed.

Copyright © FirEURisk Consortium. All rights reserved.



Figure 30: Dataset with a PDF file.

3.8 Web Interface / API

The DataHUB Data Repository provides a REST API, in order to enable the management of the information that is uploaded on the platform and the execution of the provided functionalities. The details of the API are available on the following URL:

https://fireurisk.satways.net/swagger/ui/?url=http://fireurisk.satways.net/swagger

Copyright © FirEURisk Consortium. All rights reserved.

In order to call the web API provided by DataHUB, proper authentication is required. Each user is able to create one or more tokens that will be used for this authentication process. To create a new API key, the logged user has to open his profile by selecting the user icon on the top right and the "View Profile" link as presented in the following figure.

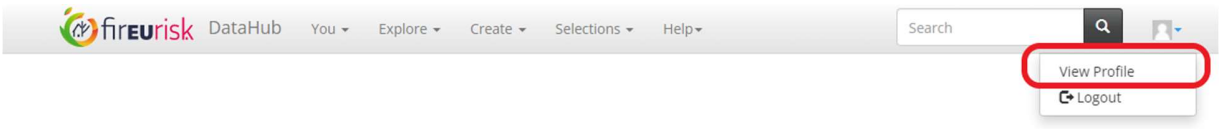


Figure 31: View Profile link

Through that page, the user can handle its own API keys. The tab "API Keys" enlists the list of active keys. To create a new API key, the user must provide a key name. By clicking the "Add" button, the key (UUID number) will be generated automatically.

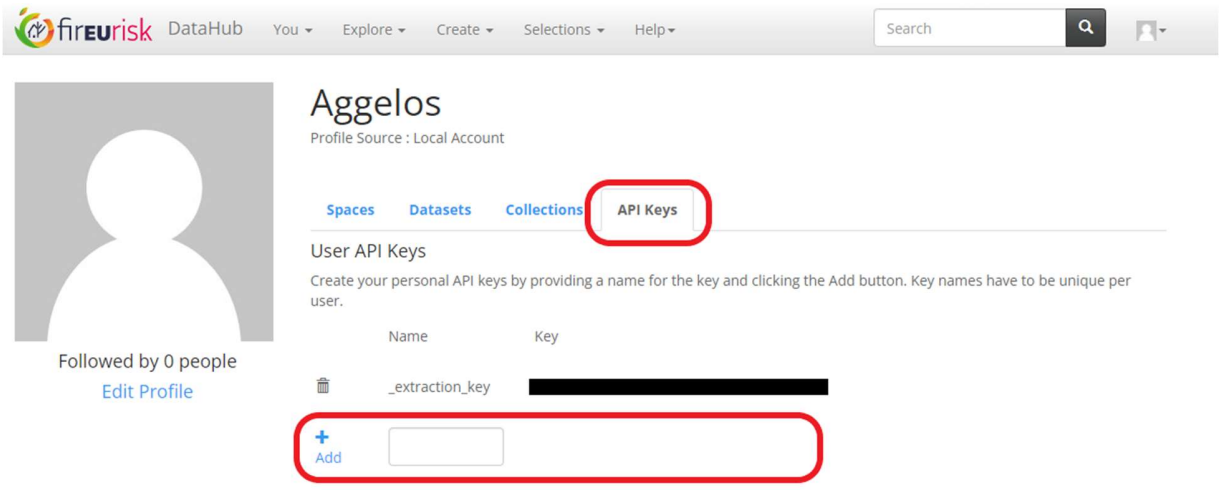


Figure 32: API Key creation

Copyright © FirEUrisk Consortium. All rights reserved.

4 DataHUB Code Repository

The DataHUB Code Repository is a version control system able to host tools and the related programming codes. It is deployed on the cloud and it is accessible through the following URL:

https://git.fireurisk.satways.net

This system is based on an opensource lightweight code hosting system called “Gitea” [11]. This is quite similar to the most popular Git service, “Github” [14] and can be deployed with minimal requirements of system memory.

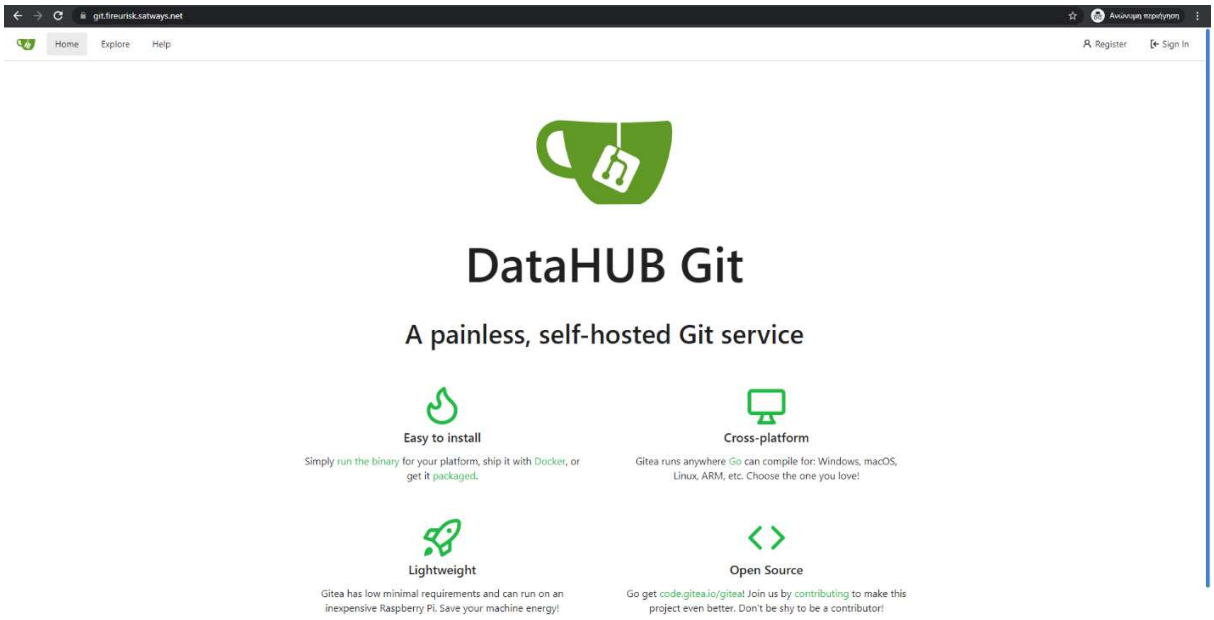


Figure 33: DataHUB Code Repository main page

In order to use the system, new user has to register themselves. This is a quick operation, initiated by selecting the “Register” menu on the top right of the navigation bar. Then the new user must fill in her/his username, email address and provide his password twice. By selecting the “Register” button (Figure 34), the registration process is completed and the new user can log into the system.

Copyright © FirEURisk Consortium. All rights reserved.

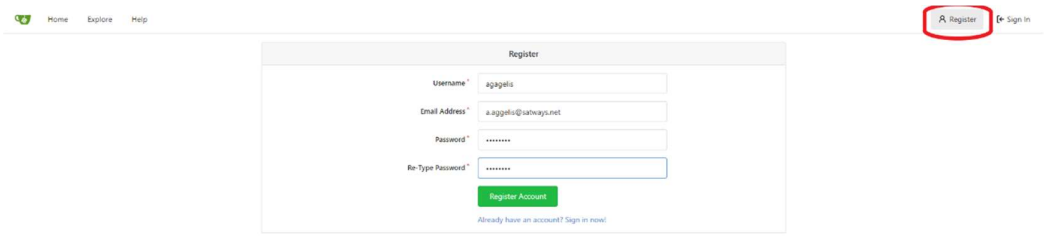


Figure 34: DataHUB Code Repository user registration

To log into the system, the user must select the "Sign In" menu, located just next to the "Register" menu, on the top navigation bar and provide his credentials (Figure 35). Finally, the green "Sign in" button must be pressed.

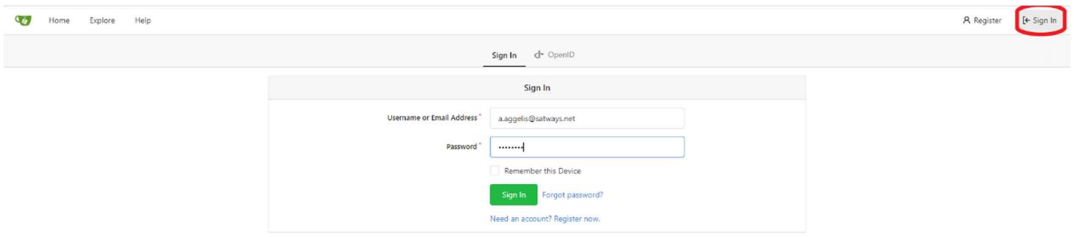


Figure 35: DataHUB Code Repository user login

Immediately after logging in, the user's main Dashboard is loaded (Figure 36). The latest actions and the user's latest activities are presented in the centre part of the page. On the top, a calendar/heatmap presents the activity level of the user for the last months. Just below the calendar, the full history of the user's actions is provided.

Copyright © FirEURisk Consortium. All rights reserved.

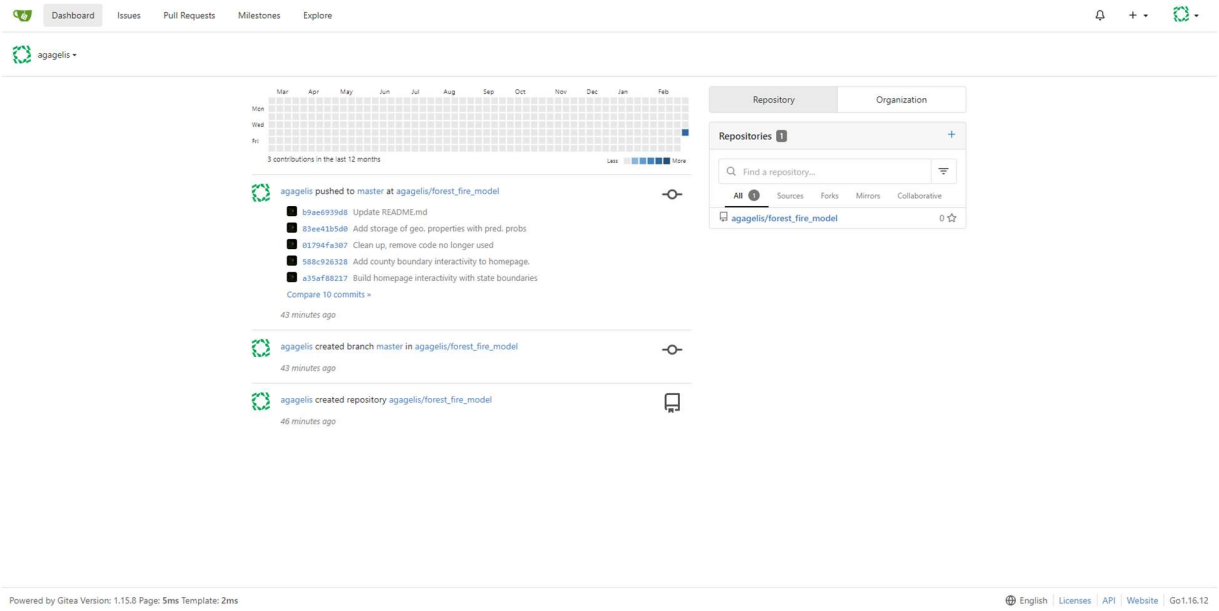


Figure 36: DataHUB Code Repository logged user Dashboard

On the right part of the Dashboard page, information regarding the repositories that the user has uploaded is presented.
Any user that has used GitHub in the past, will find it very easy to navigate on this platform too.

4.1 Code Repository creation / Code submission

To create a new Repository, the user must select the "+" menu next to the logged user's avatar, on the top right of any page. Next, the "New Repository" menu must be selected.

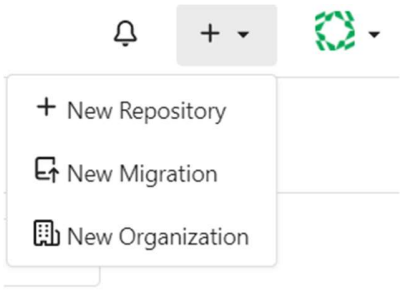


Figure 37: DataHUB Code Repository New Repository menu

The platform now presents the New Repository form (Figure 38 & Figure 39). By filling out this form, the user can configure all the available parameters regarding the ownership, accessibility and the rest of the typical parameters for a source code repository.

Copyright © FirEURisk Consortium. All rights reserved.

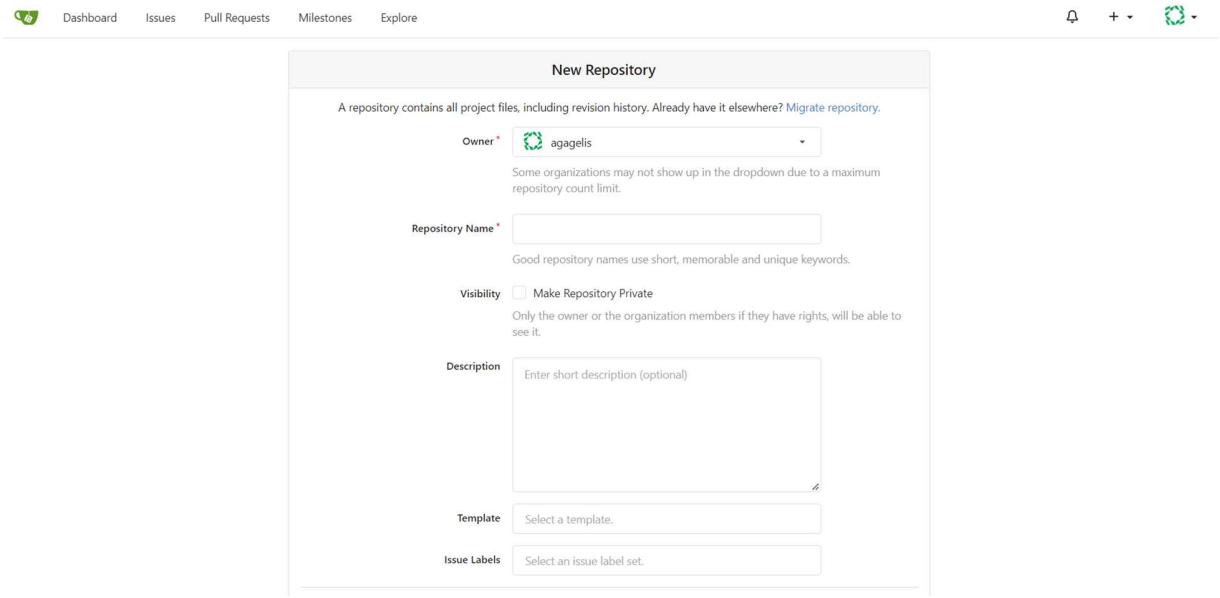


Figure 38: DataHUB Code Repository New Repository form part 1/2

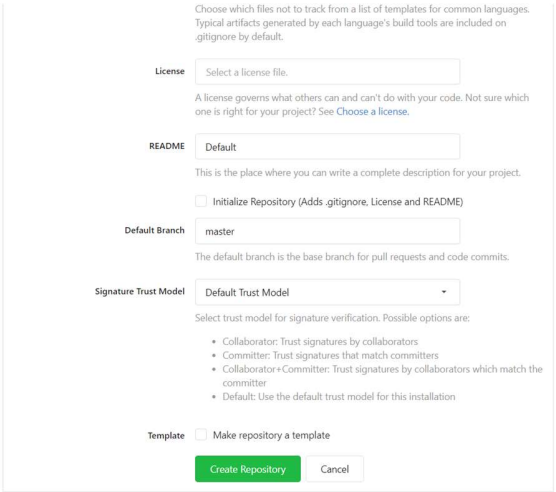


Figure 39: DataHUB Code Repository New Repository form part 2/2

When the new repository form is completed, the green “Create Repository” button (Figure 39) must be pressed, to finalize the creation process. Finally, the new repository is created, and the repository’s page presents a set of instructions (Figure 40) to help the user to upload the source code.

Copyright © FirEURisk Consortium. All rights reserved.

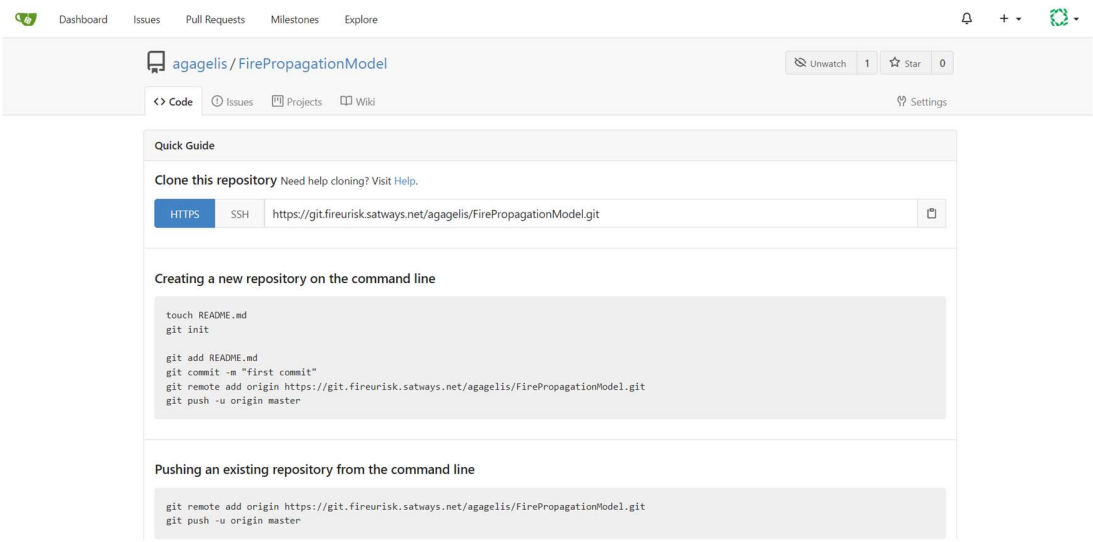


Figure 40: DataHUB Code Repository New Repository initial state

Currently, one open-source project was cloned from GitHub and uploaded to DataHUB Code Repository, as an example. You can find this project (Figure 41) under the following URL: https://git.fireurisk.satways.net/agagelis/forest_fire_model

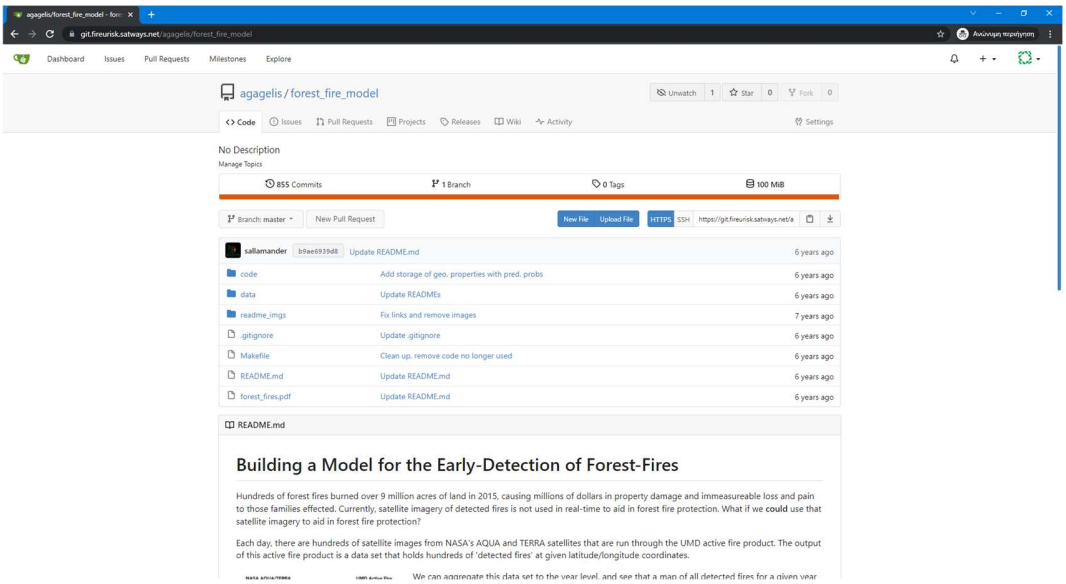


Figure 41: DataHUB Code Repository Sample project page

Copyright © FirEUrisk Consortium. All rights reserved.

5 DataHUB Execution Service

The DataHUB Execution Service enables the execution of several tools and analyses by the users. Users can set up and execute specific execution tasks. These execution tasks are mentioned as execution workflows, where users can define the tools that have to be executed, their sequence of execution and the data that must be fed as input to these tools.

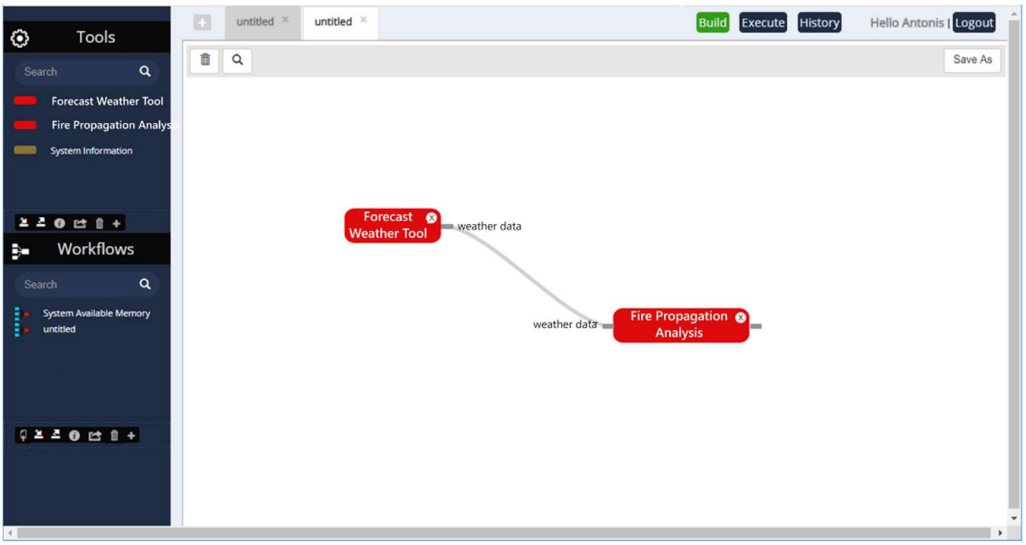


Figure 42: DataHUB Execution Service Main UI

A web editor is provided to the users, in order to define and execute these workflows. Access to this service/editor is allowed only to authenticated users. When the user logs in, he/she gains access to the main functionalities through Build, Execute and History sections. These sections are accessible through the main toolbar menu which contains three buttons that enable navigation to different pages:



Figure 43: Main menu buttons

On the left part of the page (Figure 42), the Tools and Workflows sections are presented. The Tools section contains all the available tools that can be executed on the platform. The tools can be interconnected among each other, shaping workflows. Their main concept is that one tool's output can be used as input by another tool. The Workflows section contains all the stored workflows.

5.1 Adding a Tool

The platform supports the uploading and the remote execution of Java tools. Thus, if a user needs to use another type of application, then this should be wrapped into a java jar application. In order to add a Tool, the user has to click the

Copyright © FirEURisk Consortium. All rights reserved.

button to open the Select Tool type dialog. From the dropdown menu, the user must select the Java Tool option. The following dialog will appear:

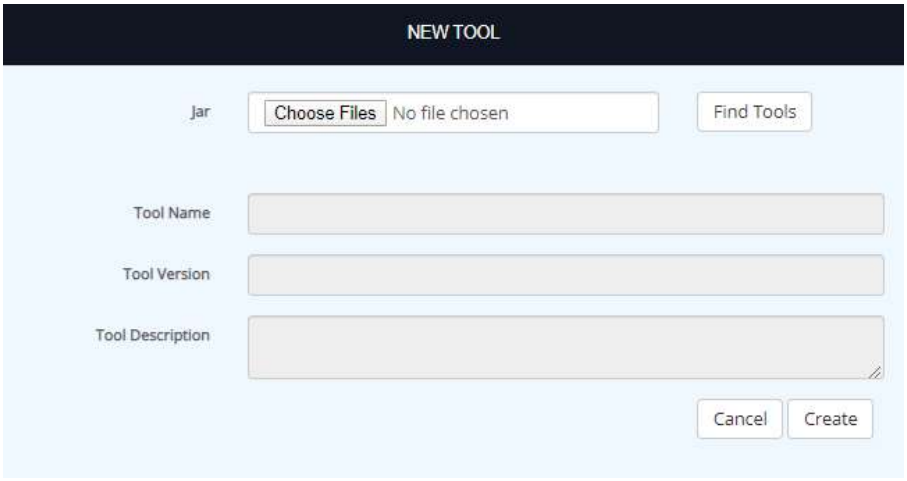


Figure 44: Upload a new tool dialog.

Through this dialog, the user can upload the java tool (.jar file) and provide some additional details, such as the tool name, version and description.

5.2 Creating Workflows

Through the Build page, new workflows can be created. A new, untitled view is always visible on the right part of the page. There, the user can drop any number of tools, selected from the Tools list, and interconnect them in order to create a new workflow.

When each tool is dropped on the workflow it is labelled by its name. The tool's inputs are indicated as endpoints on the left side of the rectangle that represents the tool, while its outputs are indicated on the right side. In case of an error, any tool can be removed by selecting the "x" symbol on the top right of the tool.

The interconnection between two tools that have been dropped on the main workflow view, can take place by clicking on any output endpoint and dragging the mouse to any tool's input endpoint. A typical example of a workflow consisting of two tools is presented in Figure 42.

Any workflow can be saved by selecting the "Save As" button on the top right of the graph. A new name for the workflow must be defined in the dialog in which popups and the Save button must be selected. Both the current tab's title and the workflow title in the Workflows list will be automatically updated.

Copyright © FirEURisk Consortium. All rights reserved.

5.3 Executing Workflows

By selecting the Execute button, the Execution page loads. From this page, the user can configure and trigger a workflow execution. In the Workflows view, all the available workflows are listed. To open a workflow the user has to select it from the list and press the (open workflow icon) button . A form will be displayed on the right part of the page. Here all the parameters and inputs must be configured.
Once all input fields are filled, the user must select the Execute button to submit the workflow for execution. A popup window should indicate the successful workflow execution.

Copyright © FirEUrisk Consortium. All rights reserved.

6 Conclusions

Within the FirEURisk project, several models, tools and data will be used. The management and sharing of this information will be supported internally by the cooperative platform. This platform has been deployed on infrastructure, making it available to users/partners. Deliverable D4.3 accompanies the deployment of the internal cooperative platform, presenting details about its design and provided functionalities. The information sharing and data management mechanism follows the DMP that has been defined in deliverable D4.2 – "Data Management Plan". As the DMP is a living document, any change/update that will be proposed during the upcoming period, will be further analysed and transferred to the cooperative platform. Additionally, this platform is planned to be the basis for the open platform of models, data and knowledge that will be developed within Task A4.3.2.

Users can use the current deliverable, as training material, in order to be informed about the provided functionalities and the way they could use these. Any update or refinement of the provided functionalities will be documented on an updated version of the deliverable.

Copyright © FirEURisk Consortium. All rights reserved.

7 References

[1] https://www.mongodb.com/
[2] https://www.docker.com/
[3] http://geoserver.org/
[4] https://www.keycloak.org/
[5] https://clowderframework.org/
[6] https://www.javascript.com/
[7] https://reactjs.org/
[8] https://www.scala-lang.org/
[9] https://en.wikipedia.org/wiki/Representational_state_transfer
[10] https://www.oracle.com/pt/cloud-native/container-registry/what-is-docker/
[11] https://gitea.io/en-us/
[12] https://git-scm.com/
[13] https://inspire.ec.europa.eu/
[14] https://github.com/

Copyright © FirEURisk Consortium. All rights reserved.