

Datagate User Manual

<http://datagate.disit.org>

Sommario

1	Introduction	2
2	Datagate overview.....	3
2.1	Organizations, Datasets, Resources	3
3	User role CKAN.....	3
3.1	Guest user (user not registered)	3
3.2	User registered on the portal.....	4
3.3	Roles in Organizations	5
4	Dataset Creation / Management details.....	12
5	Creation / Management of Resources to be uploaded also on Km4City (<i>EnhanceData</i> module)	13
6	<i>EnhanceData</i> view detailed description	13
7	EnhanceData Resource Format	17
	References.....	18

1 Introduction

The Datagate service (<http://datagate.disit.org>) has been design with the aim to offer an environment able to establish a direct connections between the data upload processes on the datagate portal e the Km4City knowledge base [Km4City]. On the Datagate service it is possible to upload static and punctual data (POI, Points of Interest), execute a number of data quality improvements operations directly on the datagate portal (many of them are performed in a transparent manner for the user, or guided via a procedure on the portal), and publish the information on the semantic knowledge base Km4City.



Figure 1 – <http://datagate.disit.org> homepage

This allows to aggregate data in a semantic way, due the use of the Km4City ontology [Km4CityOnto]. Datasets will not remain isolated anymore and end in itself but, they will be: connected with other data, reconciliated, qualified, etc. Km4City is the semantic engine that uses a multi-domain smart city ontology so that, thanks to datagate service, the data will pass from a tabular format (csv, xls, etc.) to RDF triples.

Datagate implements a module/plugin (*EnhanceData*) that can be installed also in other CKAN portals to allow to any other yet used services, to pass from “dataset aggregator” to “semantic aggregated datasets”. *EnhanceData* is a plug-in developed with the aim to execute a Quality Improvement, the Validation of

a tabular file and the following publication in the Km4City Knowledge Base. In the portal <http://datagate.disit.org> (Fig. 1) the *EnhanceData* extension has been installed.

In any CKAN portal, with the *EnhanceData* module installed, will be possible not only to expose data as usual, but also, via some simple steps and operations that will be described in the following, perform the upload of these data in the Km4City Knowledge Base.

2 Datagate overview

In the following sections, the general functionalities of the Datagate portal (based on CKAN), will be shown. Then the detailed characteristics of the *EnhanceData* module will be examined and all the actions that the operators have to follow to upload their datasets also on the Km4City KB.

CKAN is a tool that helps to manage and publicize on the web large amounts of data. It is typically used by local and national government, research institute (like the Datagate version managed by the DISIT Lab), and other organizations collecting data for private purposes or public use in the form of *opendata*.

2.1 Organizations, Datasets, Resources

The CKAN administrator can create a number of Organizations (group of users) and add in any Organization, a number of members, by selecting them from the list of yet registered users on the portal. These users will be authorized to public/visualize/modify the data of its Organizations. The data are published as a collection called Dataset. A Dataset is a collection of Resources and every resource is a file. *EnhanceData* elaborates resources created in a tabular file format in the CKAN (like csv, xls, xlsx).

Every dataset created in the CKAN, can contains:

- Metadata, that can be:
 - Mandatory: title
 - Optional: description, author, data, format, license, etc.
- A number of Resources containing the data. A Resource can be:
 - An XML file, a cvs file, a PDF document, an image, data connected in RDF format, etc.

CKAN can internally memorize the Resource and/or simply refer if as a link, since the resource could be hosted in another website.

3 User role CKAN

CKAN users can register and log in with their credentials. Each user can have a different role within different Organizations. Only registered users can upload datasets to the portal.

The possible user roles, which will be described in detail in the following, are:

- Guest
- Registered user to the portal
- User registered in an organization:
 - Member of an Organization
 - Curator of an Organization
 - Administrator of an Organization
- Sysadmin, system administrator

3.1 Guest user (user not registered)

Users who are not registered to the portal can perform the following actions:

- Display of basic portal information
- View the list of users registered on the portal, at the page: <http://datagate.disit.org/user>
- Register for the portal:
 - From the home page:
 - Fig. 2) Click the “Register” button at the top right



Figure 2 - How to subscribe to the portal dated from the home.

Fig. 3) complete the registration form.


Figure 3 - Form to subscribe to the datagate portal.

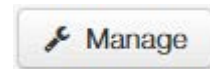
- (Alternative) Access via the direct link, the form to be completed for registration:
<http://datagate.disit.org/user/register>:
 - Enter the required information (user name, full name, e-mail, password) and click on the lower right button “Create Account” (Fig.3)
- Perform the following actions on public data sets:
 - See the list of data sets and related metadata (ex: user who uploaded them, title, description, user license, etc.)
 - Dataset search through search filters
 - Download the Datasets, paying attention to any license associated with them

3.2 User registered on the portal

Every registered user can:

- Perform every actions as the user guest
- Log-in/log-out from the portal
- Access/Modify its personal information in the profile;
 - Once logged in:

- Click the username (on the top right of the homepage) and click the Manage button
 - (Alternative) click the button  "Edit settings" (that is on the top of the page).
 - (Alternative) directly click the link (change the <username> accordingly):
<http://datagate.disit.org/user/edit/<username>>
- Can be associated to one or more Organizations (See Section 3.3)



3.3 Roles in Organizations

Users associated to Organization can perform all the actions of a registered user. There are three possible roles within an organization: member; curator and administrator.

3.3.1 Member (member)

This role is intended for external collaborators who can only view the current status of data, therefore the permissions are limited. Within an Organization, a Member can (once logged in):

- view the list of present Datasets (both public and private):
 - Home> Tab 'My Organizations'> <OrganizationName> (ex: 'DISIT-Unifi', direct link:
<http://datagate.disit.org/en/organization/disit-unifi>)

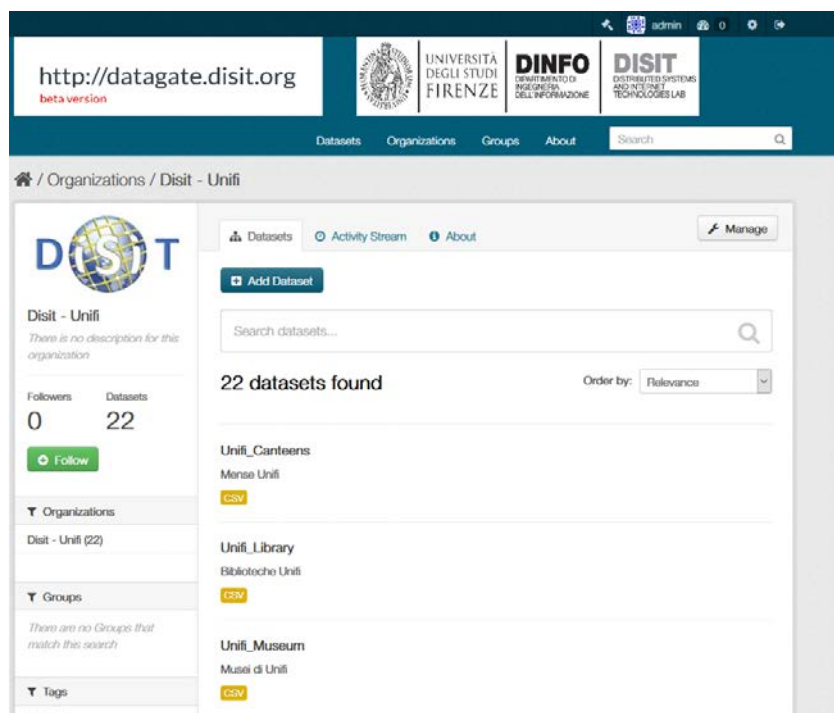


Figure 4 - List of datasets published by an organization

- Visualize the list of Resources available in every dataset.
 - Click the dataset and enter in the single resource selected. Each dataset can contains more than one resource (different files), as shown in Fig. 5.

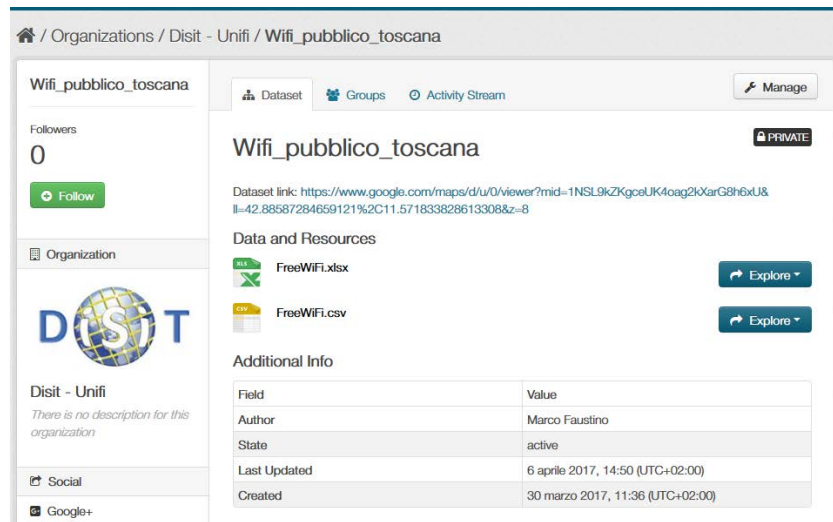


Figure 5 - Visualize the dataset files.

- Download the files of a dataset (click the file name, Fig. 5).
- Visualize resource data (edit is not allowed).

3.3.2 Curator

This role is for users that actively collaborate by inserting and treating data for an Organization. Curator users can:

- Add a new Dataset:
 - Home> Tab 'My Organizations'> <OrganizationName> (i.e.: 'DISIT-Unifi') > click the “Add Dataset” button (see Fig. 4).
- Modify the dataset metadata (Fig. 6):
 - Home> Tab 'My Organizations'> <OrganizationName> (i.e.: 'DISIT-Unifi') > click the dataset name (i.e. Wifi_pubblico_toscana) > click the “Manage” button

Add Dataset

Manage

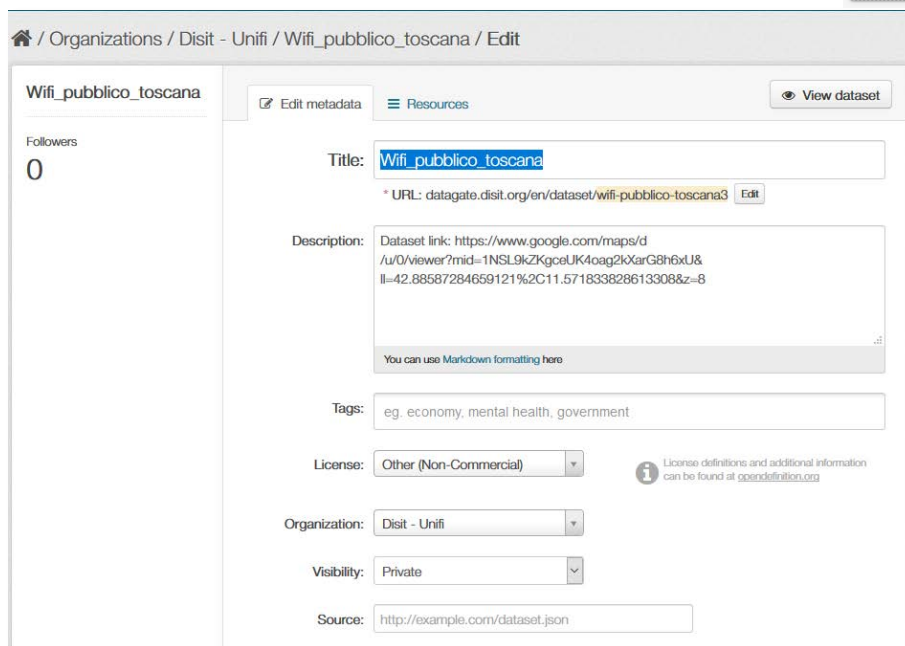


Figure 6 - Modify the dataset metadata

- Add resources (files) in a Dataset (Fig. 7):
 - Home> Tab 'My Organizations'> <OrganizationName> (i.e.: 'DISIT-Unifi') > click the dataset name (i.e. Wifi_pubblico_toscana) > click the “Manage” button, > Tab “Resources” > click the “Add new Resource” button.



Figure 7 - Add a resource to a Dataset

- Modify the metadata of a resource (Fig. 8):
 - Home> Tab 'My Organizations'> <OrganizationName> (i.e.: 'DISIT-Unifi') > click the dataset name (i.e. Wifi_pubblico_toscana) > click the “Manage” button, > Tab “Resources” > click the file name (“FreeWiFi.csv”).

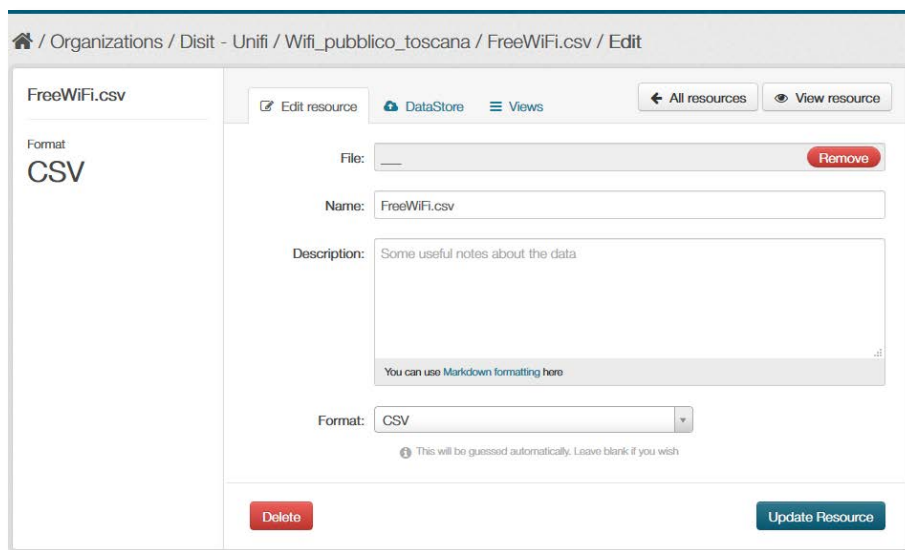


Figure 8 - Edit resource metadata

- Modify and update data of a resource using the *EnhanceData* view, available only for tabular resources (csv, xlsx, xls), compliant with the datagate template (see here: <http://datagate.disit.org/about/>). This modality has been added to the *EnhanceData* module, and is described below. It allows the upload of data in the Km4City Knowledge Base:
 - Home page > Tab “My Organization” > <OrganizationName> (i.e. Disit-Unifi) > click the dataset name (i.e. Wifi_pubblico_toscana) > click the “View resource” button.



Organizations / Disit - Unifi / Wifi_pubblico_toscana / FreeWiFi.csv

FreeWiFi.csv

URL: <http://datagate.disit.org/en/dataset/7860da18-46e5-420a-be49-808ac6f008e4/resource/91069577-7d7a-45ee-bb26-5e73159df7ec/dow...>

From the dataset abstract

Dataset link: <https://www.google.com/maps/d/u/0/viewer?mid=1NSL9kZKgcUK4oag2kXarG8h6xU&ll=42.88587284659121%2C11.571833828613308&z=8>

Source: [Wifi_pubblico_toscana](#)

EnhanceData View

Selezione Categoria: Selezione sottocategoria:

Id	otherCategoryITA	otherCategoryENG	nameITA *	nar
1			ACCESS POINT	
2			ACCESS POINT	
3			ACCESS POINT	
4			ACCESS POINT	

Figure 9 - View data via the EnhanceData module

3.3.3 Administrator of an Organization (admin)

This role is for users that have to completely manage an Organization. In addition to the features available for Members and Curators described above, the administrator can:

- Modify the metadata related to an Organization: name, description, image, etc.
 - Home > My Organizations > <OrganizationName> (es: Disit – Unifi) > click the “Manage” button

Organizations / Disit - Unifi / Manage

Disit - Unifi

There is no description for this organization

Edit Datasets Members View

Name: * URL: datagate.disit.org/en/organization/disit-unifi Edit

Description: You can use Markdown formatting here

Image URL: Remove

Custom Field: Key: Value:

Custom Field: Key: Value:

Custom Field: Key: Value:

* Required field

Delete Update Organization

Figure 10 - Organization management

- Manage all datasets of the Organization (Fig. 11):
 - Home > My Organization > <Organization Name> (i.e. Disit – Unifi) > click the “Manage” button > click the “Dataset” tab

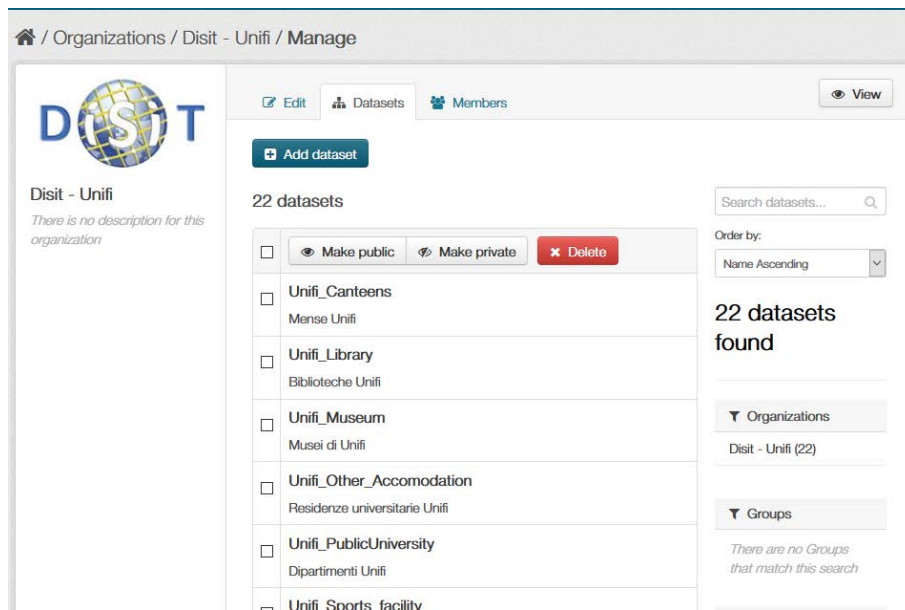


Figure 11 - Dataset management of an Organization

- Add users to an Organization or modify users' roles (Fig. 12)
 - Home > My Organization > <Organization Name> (i.e. Disit – Unifi) > click the “Manage” button > click the “Members” tab.

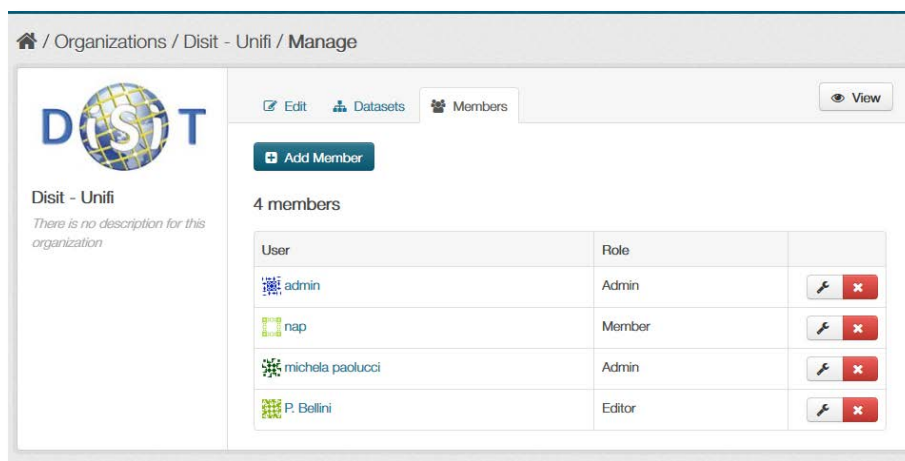


Figure 12 - Users of an Organization management

3.4 Sysadmin

The Sysadmin is the administrator of the Datagate/CKAN portal. The sysadmin is the user that can globally manage the whole portal, and in particular he can:

- Do all actions described above and referred to the other users
- Create an Organization:
 - Home > click “Organizations” (Fig. 13) > click the “Add Organization” button and fill in all the requested information (Fig. 14) and click the “Create Organization” button to confirm and save.

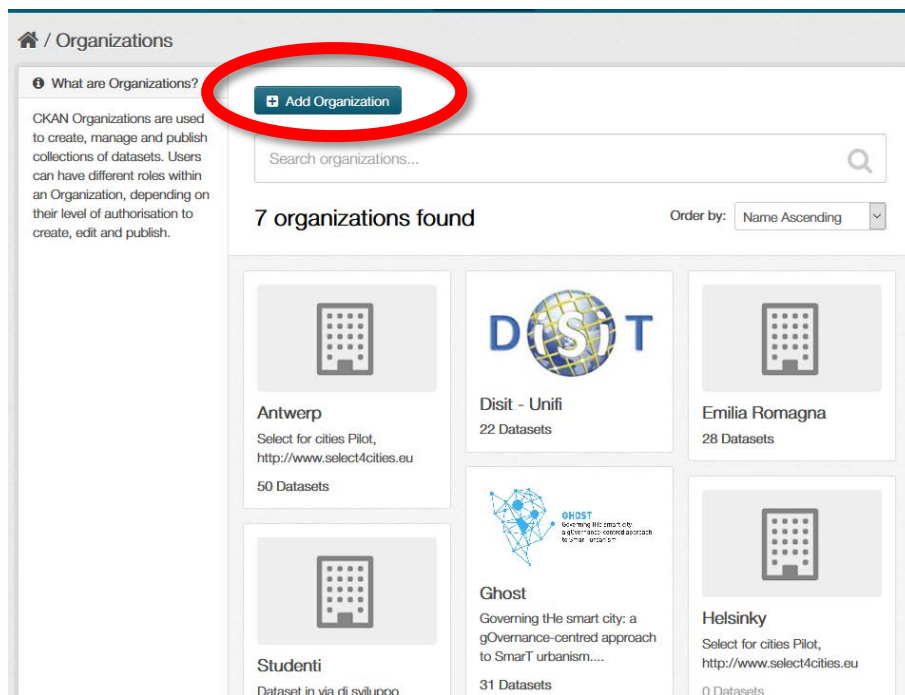


Figure 13 - Create a new Organization

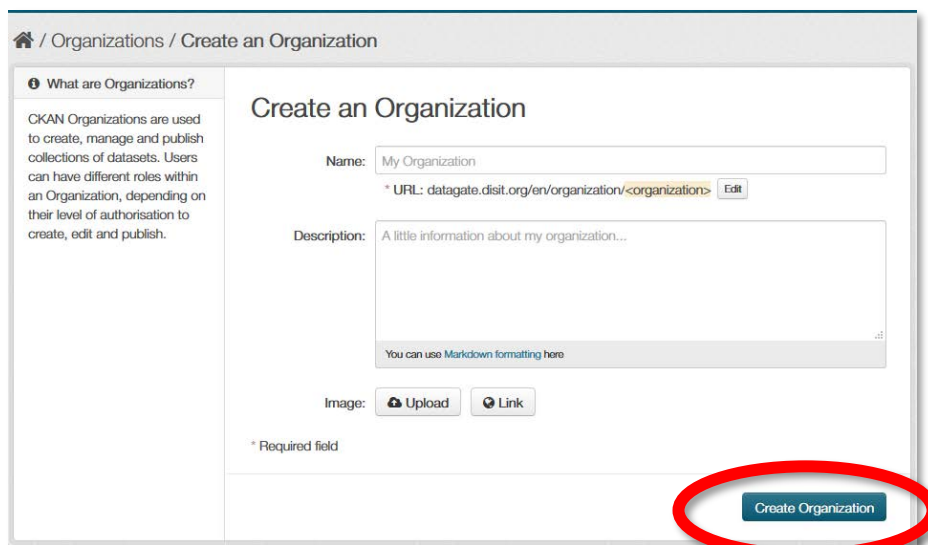




Figure 14 - Fill in the Organization requested information

- Manage the portal settings:
 - Click the  button on the top of the page
- Create/Modify the information of the main webpages:
 - Click the  button on the top of the page > click the Config Tab (Fig. 15)

CKAN config options

Sysadmins **Config** **Trash**

Site Title: This is the title of this CKAN instance. It appears in various places throughout CKAN.

Style: Choose from a list of simple variations of the main colour scheme to get a very quick custom theme working.

Site Tag Logo: This is the logo that appears in the header of all the CKAN instance templates.

About: This text will appear on this CKAN instances [about](#) page.

Intro Text: This text will appear on this CKAN instances [home](#) page as a welcome to visitors.

Custom CSS: This is a block of CSS that appears in <head> tag of every page. If you wish to customize the templates more fully we recommend [reading the documentation](#).

Homepage: This is for

Site Title: CKAN

Style: Default

Site Tag Line:

Site Tag Logo: /base/images/datagatebeta-disit.png

About: [DISIT lab](http://www.disit.org/drupal/?q=node/5499 "DISIT lab") and research group is active since 1994.

This portal and service is the main entrance for ready to use data sets for the Km4City infrastructure at DISIT lab.

You can use [Markdown formatting here](#)

Intro Text: [Km4city](http://www.km4city.org "Km4city") is a knowledge base and a research line of DISIT lab mainly developed before the start of Si-Mobility, RESOLUTE, REPLICATE projects.

You can use [Markdown formatting here](#)

Figure 15 -Portal Web page settings

- Modify the user metadata (name, username, email, description, password, Fig. 16):
 - Direct link: <http://datagate.disit.org/user> > click the name of the user to be edited > click the “Manage” button.

Users / michela paolucci / Manage

Account Info

Your profile lets other CKAN users know about who you are and what you do.

Change details

* Username: paolucci

Full name: michela paolucci

* Email: paolucci.michela@unifi.it

About: A little information about yourself

You can use [Markdown formatting here](#)

Change password

Figure 16 - Modify user details

- View the sysadmin list:
 - Direct link: <http://datagate.disit.org/ckan-admin>



4 Dataset Creation / Management details

A Dataset is a collection of resources uniquely linked to an Organization. Datasets allow to group a series of files (as Resources) associated to the same data source and edited by one or more responsible (Administrators and Curators) inside the Organization itself (Please refer to Fig. 6).

The resources within a Dataset share the metadata related to:

- Visibility: public or private
- Type of license: one of the possible available (ex: Creative Commons, GNU, etc.)
- Tags for data classification
- Author
- Maintenance

The visibility of a Dataset implies that all the resources contained inherit the same properties. Consequently, if a Dataset is defined as *private*, then it will be visible only to users who are part of its organization. Otherwise if a Dataset is public then this is visible to all users who visit the portal.

The License can be selected from those proposed by the site (the *Sysadmin* can add or remove a license during configuration).

Tags identify the keywords and the topics related to the Dataset. They can be used as filters in the internal search of the site, and in order to index and place on the search engines the public Datasets.

Author and administrator are the main responsible for Dataset; CKAN optionally asks to insert the email of both users to provide a point of reference to who will see and modify the data.

The metadata can be edited in the Edit panel (Fig.6) by the Administrators and Curators of the Organization.

Administrators and Curators can add Resources within the Dataset (Fig.7) by uploading a file from the browser or via url.

5 Creation / Management of Resources to be uploaded also on Km4City (*EnhanceData* module)

To add a new resource, a datagate user has to:

1. Be registered on the Datagate portal and logged in;
2. Be Curator or Administrator of an Organization
3. Download the template file from: <http://datagate.disit.org/about/>
4. Create a file (CSV or XLS) with the data according to the structure of the template file
5. Click the “Manage” button in the dataset page
6. Click the Resource Tab
7. Click the button: “Add a new resource” (see Fig. 7)
8. Add the Resource metadata. They are:
 - o Name
 - o Description
 - o Format (not mandatory. If omitted, it will be automatically recognized)

6 *EnhanceData* view detailed description

Once the resource has been uploaded, if it is in tabular format (CSV, XLS or XLSX and respects the template structure), then a view of *EnhanceData* (called *EnhanceData* View) is created by CKAN. You can access this view by clicking on the file name. The *EnhanceData* view allows you to view / edit individual data, and performs some quality checks on them.

The incoming data must contain the columns listed in <http://datagate.disit.org/about/> webpage, which also contains the link to the *EnhanceData* template.

This view requires an asynchronous process to be executed for file upload and quality improvement, so it may only be available after a few seconds (or minutes) depending on the size of the uploaded dataset. It is therefore necessary to wait and reload the page. The view that will be automatically shown as soon as ready.

The *EnhanceData* view (Fig.17) consists of a table (each page shows 15 lines), in which all the fields are displayed. The mandatory fields are highlighted with an asterisk ('*').

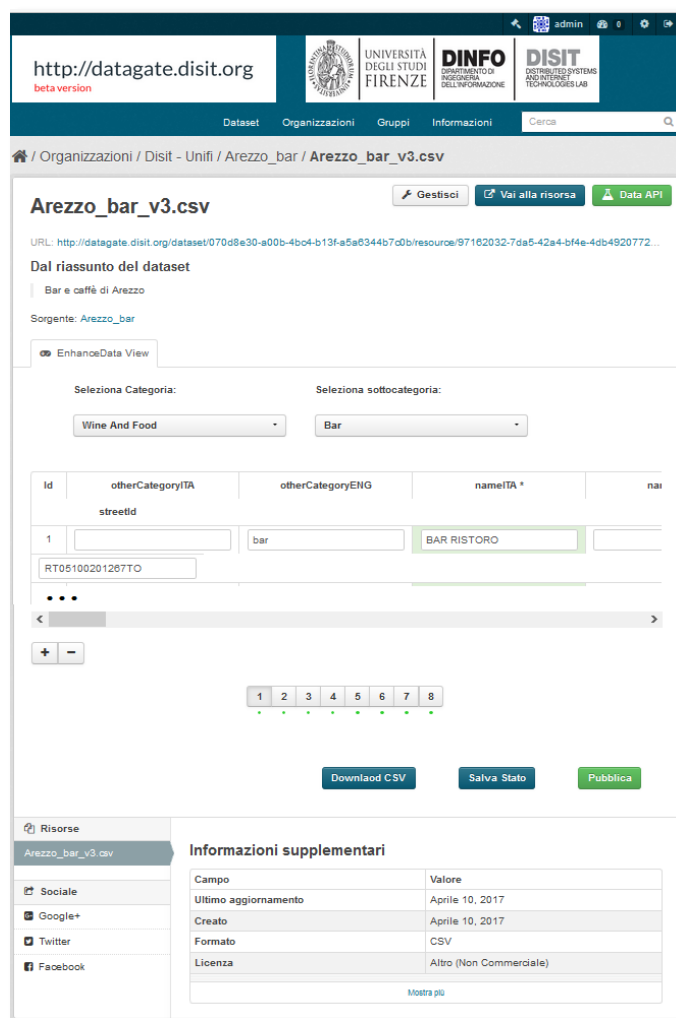
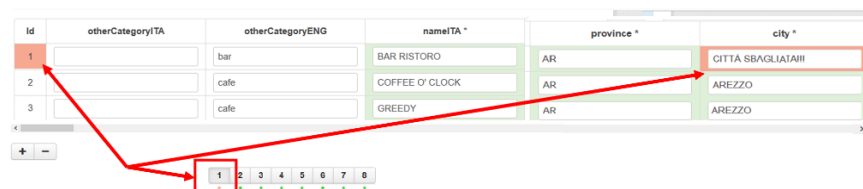


Figure 17 - View of the EnhancedData extension

A series of operations for the Quality Improvement (QI) of the data are carried out in a transparent way for the user (the streets are all written in capital letters, the telephone numbers are correctly written, the e-mails are verified in their form and substance, province and city compared so that they are consistent, etc.). These QI operations are performed automatically by the system, marking the green cells if the validation has been carried out correctly, otherwise the cells are marked in red if the validation has reported one or more errors (Fig.18a, 18b)). The fields in red must be corrected manually by the user. Every page that presents errors is highlighted with a red dot (at the bottom of the pagination bar), otherwise there is a green dot. The "Id" field of a line is marked in red if there are errors (Fig. 18 a)). Furthermore, text messages / comments are visible to help the resource managers to correct any errors (and / or inconsistencies) present in the resource (see Fig.18 c), green and blue blocks).



Id	otherCategoryITA	otherCategoryENG	nameITA *	province *	city *
1		bar	BAR RISTORO	AR	CITTÀ SBAGLIATA!!!
2		cafe	COFFEE O' CLOCK	AR	AREZZO
3		cafe	GREEDY	AR	AREZZO

Figure 18a - Errors warning in the module EnhanceData view

photo	latitude *	longitude *	streetId
	43.876502	11.058454	
	43.875658	11.033248	
	43.876239	11.03258	
		11.104317	
	43.886024	11.103764	
	43.898803		
	43.899589	11.111125	
	43.87965	11.112969	
	43879467	11.113182	
	43.874035	11.118629	RT10000501004TO
	43.901144	11.10847	
	43.8712	11.104516	RT10000501169TO
	43.87059	103807	RT10000500771TO
	43900117	11.101299	
	43.900609	11.101652	

+ -

1 2 3 4 5 6 7 8 9 10 11 12 13

Figure 18b - Errors warning in the EnhanceData module view

Arezzo_bar_v3.csv
Gestisci
Vai alla risorsa
Data API

URL: <http://datagate.disit.org/dataset/070d8e30-a00b-4bc4-b13f-a5a6344b7c0b/resource/97162032-7da5-42a4-bf4e-4db49207727b/download>

Dal riassunto del dataset

Bar e caffè di Arezzo

Sorgente: Arezzo_bar

EnhanceData View

Incorpora

Info Caricamento dati della Risorsa, si consiglia di attendere la fine del caricamento.

Nota La selezione di Categoria e Sottocategoria è obbligatoria.

Nota I campi * sono obbligatori. Gli errori sono segnalati in rosso, assicurati di correggerli prima della Pubblicazione.

Selezione Categoria:
Wine And Food

Selezione sottocategoria:
Bar

Figure 19 - Errors warning in the EnhanceData module view

For a Resource to be published on Km4City, the Category and Sub-Category related to the data must be selected (Fig.19). The list of Categories and Subcategories in the Knowledge Base is loaded automatically, and referred to the KM4City Ontology [KM4CityOnto]. It is also possible to see the list of categories on the [ServiceMap] tool, which also offers the possibility to see some examples of data already present in the Km4City ontology (this is useful if it is not clear which category and / or subcategory has to be associated with the data).

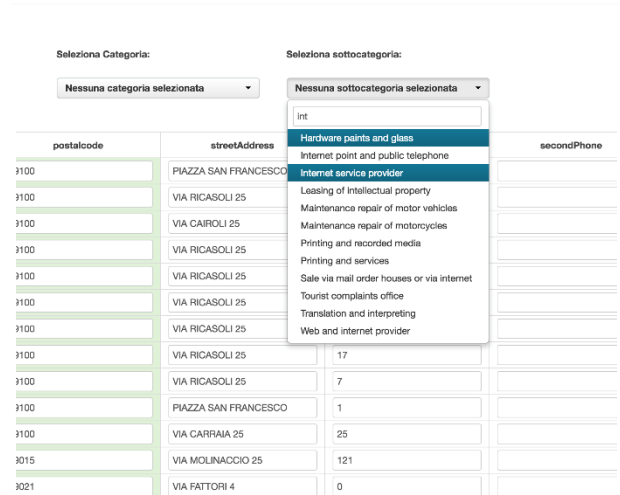


Figure 19 - Category and Subcategories selection in the Km4City ontology

To take advantage of the *EnhanceData* module view, it is advisable to wait for all pages has been loaded when the Resource is opened. The loaded pages are indicated with a dot (red or green) below the page number.

It is possible to delete or add one or more lines from the Resource by clicking on the '-' or '+' button on the bottom left. In the same way a row can be added to the table by clicking on '+'. This line will be added to the last page (Fig. 18 a)). It is not possible to add a new line if the last one is yet completely empty.

Once the changes to the Resource have been completed, it is possible to:

- Exit without saving changes
- Save the file (Resource on the portal) and download it by clicking on "Download CSV"
- Save the current status of the changes by clicking on "Save Status"
- Publish the Resource on KM4City, active ONLY IF there are no errors in the data, by clicking on "Publish".

In case a incomplete Resource that does not have all the necessary fields for the *EnhanceData* elaboration is inserted, an error page is displayed (Fig.20). In this page the missing and / or wrong fields are indicated and the possibility to download a template as example is provided.

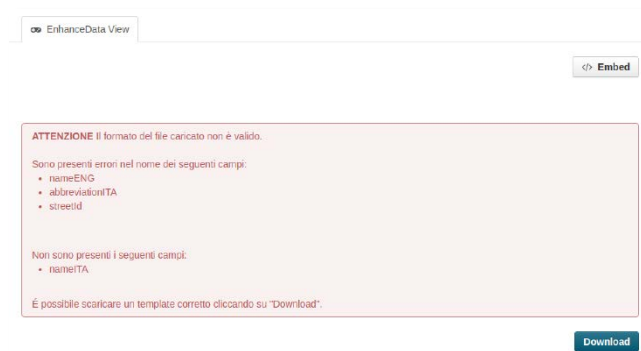


Figure 20 - EnhanceData view in case of error

7 EnhanceData Resource Format

The format of the Resource to be loaded has to contain the following fields (it can be downloaded here: <http://www.disit.org/7040>):

- otherCategoryITA: optional (es.: "SEDE CENTRALE UNIVERSITARIA")
- otherCategoryENG: optional (es.: "UNIVERSITY HEAD OFFICE")
- nameITA: **mandatory** (es.: "UNIVERSITÀ")
- nameENG: optional (ex.: "UNIVERSITY")
- abbreviationITA: optional (es.: "UNI")
- abbreviationENG: optional (es.: "UNI")
- descriptionShortITA: optional (es.: "SEDE CENTRALE UNIVERSITARIA")
- descriptionShortENG: optional (es.: "UNIVERSITY HEAD OFFICE")
- descriptionLongITA: optional (es.: "SEDE CENTRALE UNIVERSITARIA DELLA CITTÀ DI FIRENZE")
- descriptionLongENG: optional (es.: "UNIVERSITY HEAD OFFICE OF FLORENCE")
- phone: optional (es.: "055 275 8951")
- fax: optional (es.: "055 2756741")
- url: optional (es.: "http://www.unifi.it/")
- email: optional (es.: "informa.studenti@unifi.it")
- RefPerson: optional (es.: "Mario Rossi")
- province: **mandatory** (es.: "FI")
- city: **mandatory** (es.: "FIRENZE")
- postalcode: optional (es.: "50100")
- streetAddress: optional (es.: "via di Santa Marta")
- civicNumber: optional (es.: "3")
- secondPhone: optional (es.: "055 275 8932")
- secondFax: optional (es.: "055 2756743")
- secondEmail: optional (es.: "informa.studenti2@unifi.it")
- secondStreetAddress: optional (es.: "viale Morgagni")
- secondCivicNumber: optional (es.: "42")
- notes: optional (es.: "si prega di prendere appuntamento per via telefonica")
- timetable: optional (es.: "l'ufficio è aperto dalle ore 9 alle ore 17")
- photo: optional (es.: "http://mdthemes.unifi.it/azimuth/images/head.png")
- latitude: **mandatory** (make sure to not enter the ',' character) (es.: "43.71552303")
- longitude: **mandatory** (make sure to not enter the ',' character) (es.: "10.40238523")
- streetId: LEAVE IT EMPTY (ex.: ""), it is automatically filled by the system.



In the case where a district (frazione) is present in a data, it must be inserted in the "notes" field with the format: Frazione: 'districtname'.

References

- [KM4City] <http://www.km4city.org>
- [Km4CityOnto] Km4City ontology, <http://www.disit.org/6507>
- [CKAN] CKAN, The open source data portal software, <https://ckan.org>
- [DISIT] Disit Lab, DINFO, University of Florence, <http://www.disit.org>
- [CKAN_doc] CKAN user guide, <http://docs.ckan.org/en/latest/user-guide.html>
- [EnhanceDataTemplate] Template *EnhanceData* on Datagate, <http://www.disit.org/7040>
- [ServiceMap] Service Map Tool, <http://servicemap.disit.org>