



Meteo Italian Supercomputing poRtAL

Deliverable

D5.2 Open Data Catalogue

| | |
|----------------------|------------|
| Deliverable Lead: | Dedagroup |
| Deliverable due date | 30/06/2020 |
| Status | FINAL |
| Version | V1.0 |



this project has received funding from *European Commission*
Connecting Europe Facility 2014-2020
AGREEMENT No INEA/CEF/ICT/A2017/1567101

Document Control Page

| | |
|---------------------------|-----------------------------------------------------------------------------------|
| Title | Mistral_D5.2 Open Data Catalogue |
| Creator | DEDAGROUP |
| Publisher | Mistral Consortium |
| Contributors | Martina Forconi (DEDAGROUP), Andrea Finazzi (DEDAGROUP) |
| Type | Report |
| Language | en-GB |
| Rights | Copyright "Mistral Consortium" |
| Audience | <input checked="" type="checkbox"/> public <input type="checkbox"/> restricted |
| Requested deadline | |

Contents

| | |
|-----------------------------------------------------|----|
| Contents | 3 |
| Executive Summary | 4 |
| 1 DCAT-AP_IT Italian profile for data catalog | 5 |
| 2 Open data catalog architecture..... | 7 |
| 2.1 CKAN portal | 9 |
| 2.1.1 CKAN data model overview | 10 |
| 3 Installing CKAN with docker compose..... | 10 |
| 3.1.1 Ckan extensions..... | 12 |
| 3.1.2 Webserver | 13 |
| 3.1.3 Ckan configuration | 13 |
| 3.1.4 Volumes | 16 |
| 3.1.5 Organizations..... | 16 |
| 3.1.6 Groups/Categories..... | 17 |
| 3.1.7 Search through the CKAN front-end..... | 18 |
| 4 Catalog harvesting..... | 18 |
| Annex I Mistral Catalog and Portal Images | 21 |

Executive Summary

The goal of Mistral project is to facilitate and foster the re-use of datasets by the Italian meteorological community, as well as by its cross-area communities including private providers and citizens.

This deliverable aims to describe the architecture of the Mistral Open Data Catalog, that is compliant with DCAT-AP_IT. The DCAT-AP_IT is the Italian Extension of DCAT-AP (Data Catalog Vocabulary application profile) for describing public sector catalogs, aimed at ensuring semantic interoperability between European open data portals.

The Open Data catalog represents the very useful mechanism for the discovery and the fruition of Meteo data produced by the Mistral project, but also of the main datasets available in Italy.

The Mistral Open data catalog is a customized version of Comprehensive Knowledge Archive Network (CKAN;<http://ckan.org/>), the open-source data portal software developed and managed by the Open Knowledge Foundation.

The open-data presentation / distribution subsystem is the access point to the data accumulated and generated by the project Mistral. It is achieved through a series of different integrated components, among them the CKAN portal, through which information is mainly provided indexing on available data and actual data servers for which access is provided according to a type protocol REST. The storage is essentially based on two operational databases: DBAll.e as the shortterm repository and Arkimet as the long-term repository.

1 DCAT-AP_IT Italian profile for data catalog

The DCAT-AP_IT is the Italian Extension of DCAT-AP (Data Catalog Vocabulary application profile) for describing public sector catalogs, aimed at ensuring semantic interoperability between European open data portals. The Italian Extension has been defined AGID, whose first version was published in April 2017¹.

The **DCAT-AP_IT** standard is derived from the DCAT-AP standard, which in turn is derived from the W3C DCAT standard. **DCAT** is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. DCAT is a way to provide context or metadata for your datasets.

DCAT-AP, the European standard, makes some fields mandatory, recommended and optional compared to DCAT.

DCAT describes datasets on 3 different levels. The Catalog, the Dataset and the Distribution level. Each level contains information, meta-data, about itself, such as the title, description, and so on. Furthermore, the upper layers contains URI references to its children, which means it's easy to cross-reference all of the data.

dcatalog:Catalog: the Catalog is the upper level in the DCAT standard. It contains a description of itself, such as the title, description, publisher and licenses. It also contains references to the different Datasets which can be found within this particular Data Catalog.

dcatalog:Dataset: the Dataset contains information about the individual Dataset, such as the author, which category the dataset belongs to, geographical information, and so on. It also has references to all the different Distribution methods of this particular dataset.

dcatalog:Distribution: This is the deepest level of the DCAT standard. At this level we define the access URL for the data, type (CSV, JSON, ...), license, size, download URL, and so on.

The **DCAT-AP_IT** extends DCAT-AP by adding new properties for Italy described in "Table 1 - Changes to DCAT-AP v 1.1 in Italian profiles".

¹ <https://www.dati.gov.it/content/dcat-ap-it-v10-profilo-italiano-dcat-ap-0>

TABLE 1 - CHANGES TO DCAT-AP v 1.1 IN ITALIAN PROFILES

| Property | Class | Comment | URI | Range |
|------------------------|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|---------------------------------------|
| rightsHolder | dataset | added new role for agent, i.e., rightsholder | dct:rightsHolder | dcapit:Agent (subclass of foaf:Agent) |
| creator | dataset | added new role for agent, i.e., creator | dct:creator | dcapit:Agent (subclass of foaf:Agent) |
| sub-theme | dataset | added a recommended property for sub theme or controlled keywords (vocabulary Eurovoc) | dct:subject | skos:Concept |
| identifier | agent | agent identifier as a literal (introduced for implementation reasons for dati.gov.it) | dct:identifier | rdfs:Literal |
| name | licencedocument | Recommended licenses: CCs. However, if other or proprietary licenses are to be specified, name and version can be potentially provided | foaf:name | rdfs:Literal |
| version | licencedocument | the version of the license | owl:versionInfo | rdfs:Literal |
| identifier | standard | The class standard has been extended with a number of properties that can allow us to provide details on the standard to which the dataset is compliant. The class standard can be used to model legislation references as well. | dct:identifier | rdfs:Literal |
| title | standard | | dct:title | rdfs:Literal |
| description | standard | | dct:description | rdfs:Literal |
| referenceDocumentation | standard | | dcapit:referenceDocumentation | rdfs:Literal |
| name | organization | In the Italian profile the vcard:Kind abstract class is instantiated with the class vcard:Organization. The vcard:Organization has a set of properties defined as reported in https://joinup.ec.europa.eu/asset/dcat-ap_implementation_guidelines/asset_release/dcat-ap-how-describe-contact-information | vcard:fn | rdfs:Literal |
| email | organization | | vcard:hasEmail | vcard:Email |
| telephone | organization | | vcard:hasTelephone | vcard:TelephoneType |
| site | organization | | vcard:hasURL | owl:Thing |
| geographicalIdentifier | location | geonames.org or mdr/authority/place. The Italian profile follows the recommendation of the GeoDCAT-AP extension | dcapit:geographicalIdentifier | rdfs:Literal |

Dataset mandatory fields in DCAT-AP_IT²:

- Dataset identifier (dct:identifier)
- Title for the dataset (dct:title)
- Description for the dataset (dct:description)
- Update/modification date (dct:modified)
- ThemeTaxonomy (dcat:theme)

² <https://docs.italia.it/italia/daf/linee-guida-cataloghi-dati-dcat-ap-it/stabile/dataset.html>

- Rights Holder: dct:rightsHolder
- Dataset Update frequency dct:accrualPeriodicity
- Distribution for the dataset (dcat:distribution)

Dataset recommended fields in DCAT-AP_IT

- Sub-theme of the dataset (dct:subject)
- Contact Point for the dataset (dcat:contactPoint)
- Publisher for the dataset dct:publisher

The **DCAT-AP_IT** profile includes the following subjects:

- **dct:rightsHolder**: The rights holder represents an organization (or public administration) responsible for the overall management of the dataset by virtue of its institutional duties.
- **dct:publisher**: The publisher is an organization (or public administration) responsible for making the dataset available (publishing).
- **dct:creator**: The creator is an organization (or public administration) that has materially produced the dataset.
- **dcat:contactPoint**: The contact point it is advisable not to confuse this metadata with the organization that owns or publishes or creates the dataset. The contact point is typically used to specify the references of an organization office that can provide information on the dataset.

2 Open data catalog architecture

The open-data presentation / distribution subsystem is the access point to the data accumulated and generated by the project Mistral. It is achieved through a series of different integrated components, each with characteristics and tasks dedicated. Among them the CKAN portal, through which information is mainly provided indexing on available data and actual data servers for which access is provided according to a type protocol REST. The storage is essentially based on two data sources:

- **DB-All.e** (GitHub address <https://github.com/ARPASIMC/dballe/>): a fast on-disk database used exclusively to store observed data. It allows to manage large amounts of data using its simple API and provides tools to import and export in the standard BURF format. Its main use is to accommodate and make readily available the most current data while the older ones are periodically archived;
- **Arkimet** (GitHub address <https://github.com/ARPASIMC/arkimet/>): both forecast and observed data are collected and stored for a long time. It currently supports data in GRIB,

BUFR formats. Arkimet manages a set of datasets, each of which contains homogeneous data stored in segments. It exploits the commonalities between the data in a dataset to implement a fast, powerful and space-efficient indexing system. When data is ingested into Arkimet, it is scanned and annotated with metadata, such as reference time and product information, then it is dispatched to one of the datasets configured in the system

The ingestion and harmonization of observational data is coordinated in Mistral by Apache NiFi (<https://nifi.apache.org/>), an open-source application for

the design and management of data transformation flows, and relies on two different operational databases: DBAll.e (<https://github.com/ARPA-SIMC/dballe>) as the shortterm repository and Arkimet (<https://github.com/ARPASIMC/arkimet>) as the long-term repository. The main flows developed in Mistral (with Apache NIFI/python scripts) are:

- Forecast data ingestion to Arkimet
 - COSMO-2I: COSMO at 2.2km - Italy area
 - COSMO-5M: COSMO at 5km - Mediterranean Region
 - Italy Flash Flood: ecPoint rainfall blended with COSMO 2.2 ensemble post-processed Italy area
- Radar data ingestion to ArkiMet
- Ground station data to DBAll.e via AMQP
- Ground station data to DBAll.e via FTP
- Ground station data to DBAll.e via meteotrentino open data API
- Multimodel forecast to to DBAll.e via AMQP
- DBAll.e to ArkiMet data transfer

Figure 1 summarizes the CKAN components and their interactions with other key architectural elements involved in collections data.

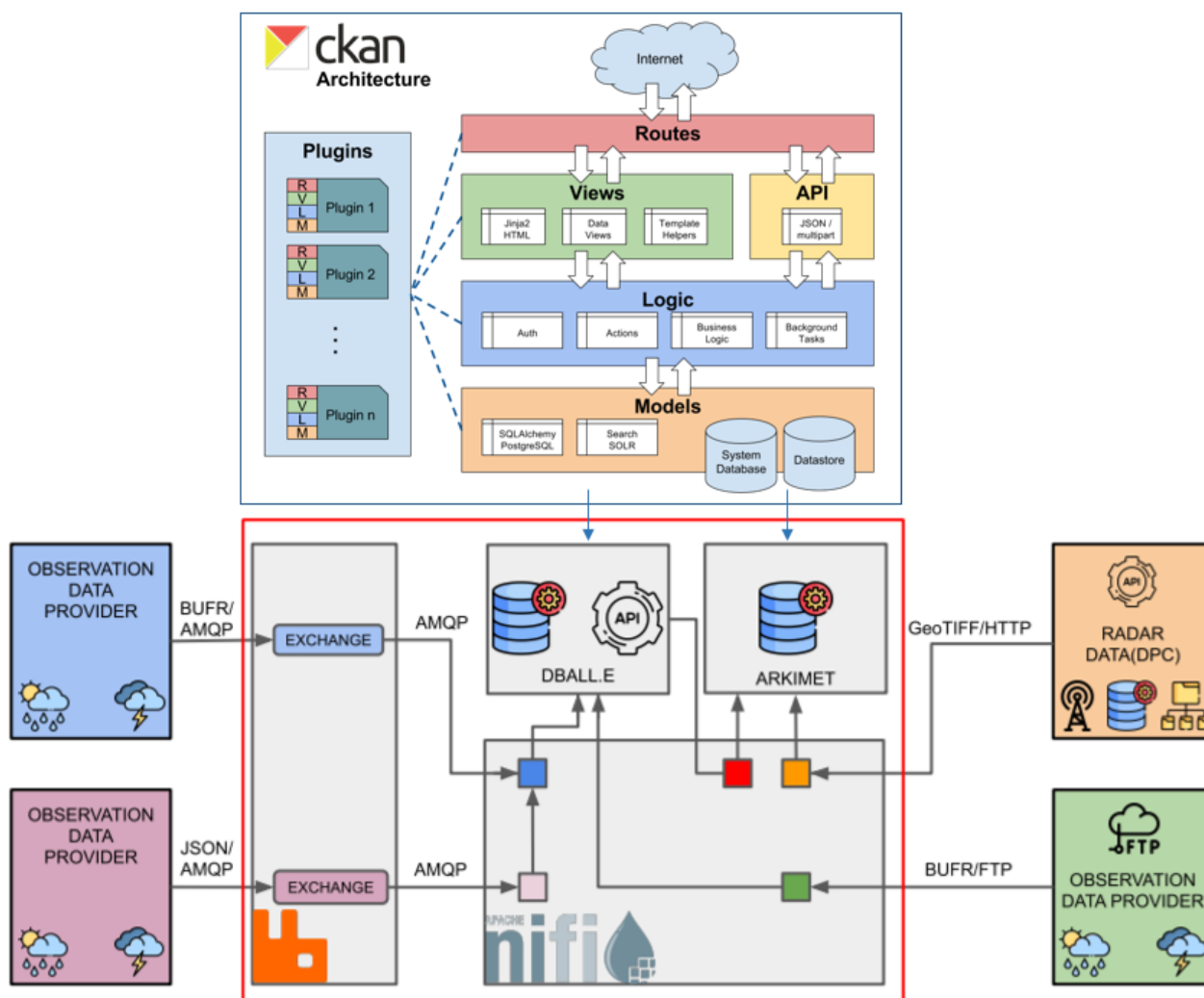


Figure 1 - Ckan and Mistral architecture

2.1 CKAN portal

The Mistral Open data catalog is a customized version of Comprehensive Knowledge Archive Network (CKAN;<http://ckan.org/>), the open-source data portal software developed and managed by the Open Knowledge Foundation.

CKAN is one of the reference software for the publication of Open Data both institutional, such as the datasets of Municipalities and Regions, both governmental and private, being widely used also by companies that provide their users with information relating to the functioning of the services provided and the work performed (transport, services to citizens, utilities in general).

The CKAN portal is structured in several integrated components:

1. **the backend (Python)**, which manages the graphics, the integration between the various components, authentication and API publication;
2. **the database (PostgreSQL)**, which contains the metadata, references and published datasets;
3. **the search engine (Apache Solr)**, which queries the database based on the datasets requested by the user, both through predefined identifiers (tags, groups, organizations) and through full-text searches.

The CKAN portal of the Mistral project is used **to index the data** collected and produced by the Mistral project. Great some of these do not reside on CKAN itself, being accessible through the REST portal specialized in providing large ones moles of data. CKAN therefore contains links to such data, grouped by dataset described by metadata useful for their recovery and use, and allows searches on the catalog of datasets from web interface or via API.

The CKAN metadata complies with the Data Catalog Vocabulary (DCAT; <https://www.w3.org/TR/vocab-dcat/>) standard; user inter-faces and download mechanisms provide a rich interface to explore data, directly leveraging its own API.

CKAN is a modular system, and the base functionality can be customized and extended with extensions.

The **ckanext-dcatapit extension** provides plugins that allow CKAN to expose and consume metadata from other catalogs using RDF documents serialized according to the Italian DCAT Application Profile. The Data Catalog Vocabulary (DCAT) is “a RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web”.

2.1.1 CKAN data model overview

Datasets are the central entity in CKAN, acting as ‘umbrella’ objects holding the metadata and one or more **resources** (a file, URL or other resource). When a resource is uploaded to CKAN, the file is processed and stored according to its data type. All uploaded files are added to the FileStore, a persistent local or cloud file storage location. FileStore objects can be discovered using their data set metadata and only downloaded in their entirety; CKAN is essentially agnostic about their contents. However, if one of these files contains structured information in a known format—for example, an Excel or CSV spreadsheet—a further processing step takes place and the data are extracted and imported into the DataStore. The DataStore provides a database for structured storage of data, in which individual spreadsheet rows can be accessed and queried via the web API.

3 Installing CKAN with docker compose

The stack is based on Docker containers deployed with docker-compose, as described in the following documentation:

<https://docs.ckan.org/en/2.8/maintaining/installing/install-from-docker-compose.html>

The following tools are required:

4. **Docker Engine:** the containerization runtime
5. **Docker Compose:** multi-container orchestration tool

The following container have been created:

| Service Name | image | Network | Volumes | notes |
|--------------|---------------------------------------------------------|---------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|
| webserver | nginx:alpine | 80:8080 443:8443 | webserver_config:/etc/nginx | Reverse proxy, remaps: /catalog -> ckan:5000 /datapusher -> datapusher:8800 |
| ckan | Built from source. v2.8.2 | Internally on :5000 | ckan_config: /etc/ckan ckan_home:/usr/lib/ckan ckan_storage:/var/lib/ckan | https://github.com/ckan/ckan/tree/ckan-2.8.2 |
| db | Built from source, starting from <i>mdillon/postgis</i> | Internally on :5432 | pg_data:/var/lib/postgresql/data | https://hub.docker.com/r/mdillon/postgis |
| solr | Built from source, starting from <i>solr:6.6.2</i> | Internally on :8983 | | https://hub.docker.com/layers/solr/library/solr/6.6.2 |
| redis | redis:latest | internally on :6379 | | |

| CONTAINER ID PORTS | IMAGE | COMMAND NAMES |
|--------------------------|-------------|----------------------------------|
| 9c5a904ea8c8 5000/tcp | docker_ckan | "/ckan-entrypoint.sh..." ckan |

| | | |
|---------------------------------------------------------------------|---------------------------|---------------------------------------|
| b30ff54f9106 80/tcp, 0.0.0.0:80->8080/tcp, 0.0.0.0:443->8443/tcp | nginx:alpine | "nginx -g 'daemon of..." webserver |
| c390866ab566 6379/tcp | redis:latest | "docker-entrypoint.s..." redis |
| 27e4074af172 datapusher | clementmouchet/datapusher | "python datapusher/m..." |
| 11eee4e9e787 8983/tcp | docker_solr | "docker-entrypoint.s..." solr |
| 2d560a0a08ae 0.0.0.0:5432->5432/tcp | docker_db | "docker-entrypoint.s..." db |

3.1.1 Ckan extensions

A CKAN extension is a Python package that modifies or extends CKAN. Each extension contains one or more plugins that must be added to your CKAN config file to activate the extension's features.

The **ckanext-mistral** extension has been created to customize the CKAN template.

The following extension have been installed.

| extension | Url | version | Plugin attivati |
|----------------------------|-----------------------------------------------------------------------------------------------|------------|----------------------------------------------------------------------|
| <i>Plugin in ckan core</i> | | | Stats Image_view Text_view Recline_view |
| Datastore | <i>Ckan core</i> | | datastore |
| Datapusher | https://github.com/ckan/datapusher | 2019-01-18 | datapusher |
| Ckanext-spatial | https://github.com/ckan/ckanext-spatial | 2019-03-15 | resource_proxy spatial_metadata spatial_query csw_harvester |

| | | | |
|-------------------|-------------------------------------------------------------------------------------------------------------------------|------------|---------------------------------------------------------------------------|
| Ckanext-dcat | https://github.com/ckan/ckanext-dcat | 2019-06-25 | dcat dcat_json_interface structured_data |
| Ckanext-dcatapit | https://github.com/geosolutions-it/ckanext-dcatapit | 2019-12-09 | dcatapit_pkg dcatapit_org dcatapit_config dcatapit_csw_harvester |
| Ckanext-mistral | Mistral extensions | 1.0.0 | mistral |
| Ckanext-geoview | https://github.com/ckan/ckanext-geoview/ | 2019-04-09 | geo_view |
| Ckanext-harvest | https://github.com/ckan/ckanext-harvest | 2019-07-01 | harvest ckan_harvester |
| Ckanext-multilang | https://github.com/geosolutions-it/ckanext-multilang | 2019-02-01 | multilang |

3.1.2 Webserver

The webserver, running on the nginx:latest image, works as a reverse proxy for the environment.

Two **upstreams** are configured:

1. Upstream **ckan** toward ckan:5000
Proxy_pass under location **/catalog** and **/catalog/**
2. Upstream **datapusher** toward datapusher:8800
Proxy_pass under location **/datapusher** and **/datapusher/**

3.1.3 Ckan configuration

The following **environment variables** are set by docker-compose and overrides the configuration file's settings. The variable propagation is described at

<https://docs.ckan.org/en/2.8/maintaining/installing/install-from-docker-compose.html#environment-variables>

| Variable name | Config equivalent | Set value |
|---------------------------|---------------------------------------|------------------------------------------------------------------------|
| CKAN_SQLALCHEMY_URL | <code>sqlalchemy.url</code> | postgresql://ckan:\${POSTGRES_PASSWORD}@db/ckan |
| CKAN_DATASTORE_WRITE_URL | <code>ckan.datastore.write_url</code> | postgresql://ckan:\${POSTGRES_PASSWORD}@db/datastore |
| 'CKAN_DATASTORE_READ_URL' | <code>ckan.datastore.read_url</code> | postgresql://datastore_ro:\${DATASTORE_READONLY_PASSWORD}@db/datastore |
| CKAN_REDIS_URL | <code>ckan.redis.url</code> | redis://redis:6379/1 |
| CKAN_SOLR_URL | <code>solr.url</code> | http://solr:8983/solr/ckan |
| CKAN_DATAPUSHER_URL | <code>ckan.datapusher.url</code> | http://datapusher:8800 |
| CKAN_SITE_URL | <code>ckan.site_url</code> | https://www.mistralportal.it |

The full list of effective environment variable is available at

<https://docs.ckan.org/en/2.8/maintaining/configuration.html#environment-variables>

The CKAN configuration file is located at **“/etc/ckan/production.ini”** in the CKAN container. The changes applied to the config setting are:

```
## Site Settings

ckan.site_url = https://mistralportal.it
ckan.root_path = /catalog

## Search Settings

ckan.site_id = default
solr_url = http://solr:8983/solr

## Redis Settings
```

```
# URL to your Redis instance, including the database to be used.
ckan.redis.url = redis://redis:6379/0

## CORS Settings

# If cors.origin_allow_all is true, all origins are allowed.
# If false, the cors.origin_whitelist is used.
ckan.cors.origin_allow_all = true
# cors.origin_whitelist is a space separated list of allowed domains.
# ckan.cors.origin_whitelist = http://example1.com http://example2.com

## Plugins Settings

ckan.plugins = stats text_view image_view recline_view datastore datapusher
resource_proxy spatial_metadata spatial_query geo_view harvest ckan_harvester
mistral dcat dcat_json_interface structured_data dcatapit_pkg dcatapit_org
dcatapit_config

## Dcatapit Extension settings
ckanext.dcat.rdf.profiles = euro_dcat_ap it_dcat_ap
ckanext.dcat.base_uri = https://www.mistralportal.it/catalog
ckanext.dcatapit.form_tabs = False

## Spatial Extension settings
ckanext.spatial.search_backend = solr-spatial-field
ckan.spatial.srid = 4326

## Front-End Settings

ckan.site_title = CKAN
ckan.site_logo = /images/logo-mistral-bianco-web-300x127.png
ckan.site_description =

## Internationalisation Settings
ckan.locale_default = it
```

```
ckan.locale_order = it en
ckan.locales_offered = en it
```

3.1.4 Volumes

| Name | Mount point(s) |
|------------------|-----------------------------------------|
| webserver_config | Webserver-> webserver_config:/etc/nginx |
| ckan_config | ckan -> ckan_config:/etc/ckan |
| ckan_home | ckan -> ckan_home:/usr/lib/ckan |
| ckan_storage | ckan -> ckan_storage:/var/lib/ckan |
| pg_data | db -> pg_data:/var/lib/postgresql/data |

3.1.5 Organizations

CKAN Organizations are used to create, manage and publish collections of datasets. Users can have different roles within an Organization, depending on their level of authorisation to create, edit and publish.

Organizations are controlling which users can add, update and publish which datasets. Every dataset in CKAN must belong to exactly one organization. If a user is an editor or admin of an organization, then she can create new datasets in that organization, and can edit and publish the datasets that belong to that organization.

In Mistral has been defined only one organization "Mistral".

In ckanext-dcatapit extensions holder_name and holder_identifier behaves differently in new DCAT-AP-IT version. When dataset is created locally (wasn't harvested), **rights holder** information is gathered directly from **organization** to which dataset belongs. Organization is the source of holder_name and holder_identifier fields (including multilang name). However, harvested datasets will preserve original holder information that is attached to dataset.

In Mistral the ckanext-dcatapit has been modified in order to make the rights holder field editable.

3.1.6 Groups/Categories

CKAN Groups are used to create and manage collections of datasets. This could be to catalogue datasets for a particular project or team, or on a particular theme, or as a very simple way to help people find and search your own published datasets.

Groups are meant to be used by the community of users of a site (the people consuming the data, not the people who're publishing the datasets on the site) to collect related datasets together into themes like "climate" etc.

Sometimes CKAN Groups are renamed in Categories like in the dati.gov.it.

The categories of dati.gov.it (Figure 2) are the dataset themes (dcat:theme). In the local portal it is possible to define different categories/groups and the dcat:theme could be associated to the dataset as an attribute.



FIGURE 2 - DATI.GOV.IT CATEGORIES

In Mistral has been created 3 categories (CKAN groups) as shown in Figure 3:

1. Observations
2. Forecasts
3. Radar

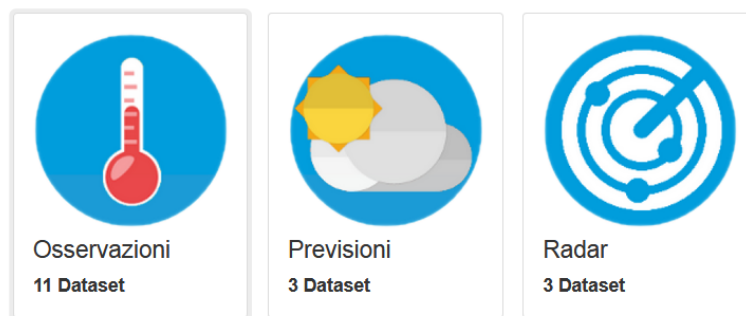


FIGURE 3 - MISTRAL CATEGORIES

3.1.7 Search through the CKAN front-end

Through the CKAN portal it is possible to access the available contents in two ways: with free navigation of datasets, possibly grouped by organization or group; via search, both full-text and tag-based. Research it can be refined by filtering through a collection of tags, desired formats, license types, groups and organizations.

The same functionality is exposed through a REST API. The main URLs are as follows:

1. https://www.mistralportal.it/catalog/api/3/action/package_list: dataset list; On early CKAN versions, datasets were called “packages” and this name has stuck in some places, specially internally and on API calls. Package has exactly the same meaning as “dataset”.
2. https://www.mistralportal.it/catalog/api/3/action/group_list groups lists
3. https://www.mistralportal.it/catalog/api/3/action/package_show?id=multimodel-forecasts Get a full JSON representation of a dataset, resource
4. https://www.mistralportal.it/catalog/api/3/action/resource_show?id=77bc16ff-2fa7-43f1-976d-d9a4eaaffa5e Get a full JSON representation of resource
5. https://www.mistralportal.it/catalog/api/3/action/package_search?q=Italy%20Flash%20Flood performs a full-text search and returns relevant datasets

4 Catalog harvesting

The ckanext-dcatapit extension provides plugins that allow CKAN to expose and consume metadata from other catalogs using RDF documents serialized according to the Italian DCAT Application Profile. The Data Catalog Vocabulary (DCAT) is “a RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web”.

The RDF DCAT Endpoints that expose the catalog's datasets in RDF serializations is:

<https://www.mistralportal.it/catalog/catalog.rdf>

The mistral catalog is harvested by national portal dati.gov.it; see <https://www.dati.gov.it/elenco-harvest-sources>

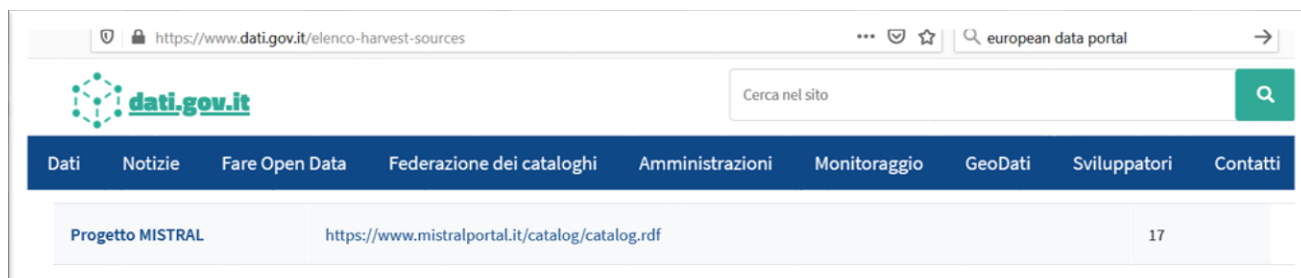


FIGURE 4 - MISTRAL IN DATI.GOV.IT HARVEST SOURCES

The mistral dataset in the dati.gov.it portal (filter by organization=progetto-mistral):

<https://dati.gov.it/view-dataset?organization=progetto-mistral>

The screenshot displays the 'dati.gov.it' website interface. At the top, the browser address bar shows the URL 'https://dati.gov.it/view-dataset?organization=progetto-mistral'. The header includes the 'Agenzia per l'Italia Digitale' logo and the 'dati.gov.it' logo with the tagline 'i dati aperti della pubblica amministrazione'. A search bar is present with the text 'Cerca nel sito'. Below the header, a navigation menu lists various categories: Dati, Notizie, Fare Open Data, Federazione dei cataloghi, Amministrazioni, Monitoraggio, GeoDati, Sviluppatori, and Contatti. The main content area shows a filter applied: 'Catalogo: Progetto MISTRAL'. The title '17 Dataset' is displayed. A search bar within the dataset list contains the text 'Cerca tra i Dati' and a 'Filtro' button. The first dataset listed is 'Rilevazione di pioggia intensa – HRD', published by 'Progetto MISTRAL' on '17-07-2019'. The description states: 'Prodotto che ha l'obiettivo di individuare delle aree in cui sono in corso precipitazioni...'. The dataset is available in 'html' and 'xml' formats. The category 'Ambiente (17)' is also visible in the left sidebar.

FIGURE 5 – DATASET EXAMPLE IN DATI.GOV.IT FILTER BY MISTRAL PROJECT

The dati.gov.it is then harvested by the European Data Portal.

<https://www.europeandataportal.eu/data/datasets/fe988f50-d0aa-44cf-aa35-2bac8383cfec?locale=en>

<https://www.europeandataportal.eu/data/datasets/669ceb26-5f0e-4e94-a169-b2a56fc60ab7?locale=en>

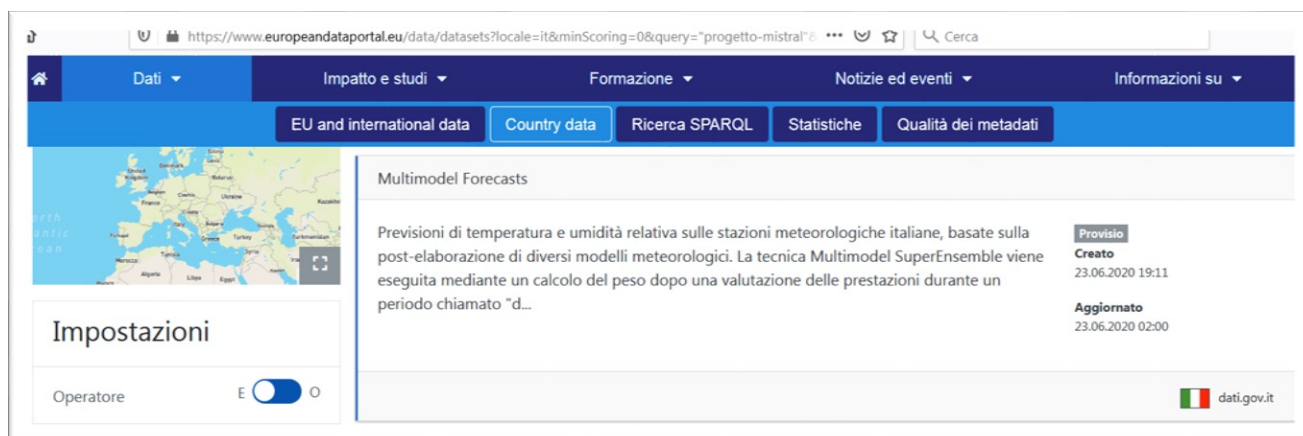



FIGURE 6 - MISTRAL DATASET EXAMPLE IN EDP

Annex I Mistral Catalog and Portal Images

The Mistral Catalog is available at the URL <https://www.mistralportal.it/catalog/>



MISTRAL
Meteo Italian Supercomputing Portal


[Accedi](#) [Iscriviti](#)

[Dataset](#) [Organizzazioni](#) [Gruppi](#) [Informazioni](#)


[Home](#) / **Gruppi**

? Cosa sono i Gruppi?


Puoi usare i gruppi di CKAN per creare e gestire collezioni di dataset, come un catalogo di dataset di un progetto o di un team, su un particolare argomento o semplicemente come un modo semplice per consentire di trovare e cercare i dataset che hai pubblicato.

Cerca gruppi... 


3 gruppi trovati Ordina per: Nome Crescente ▼



Osservazioni
11 Dataset




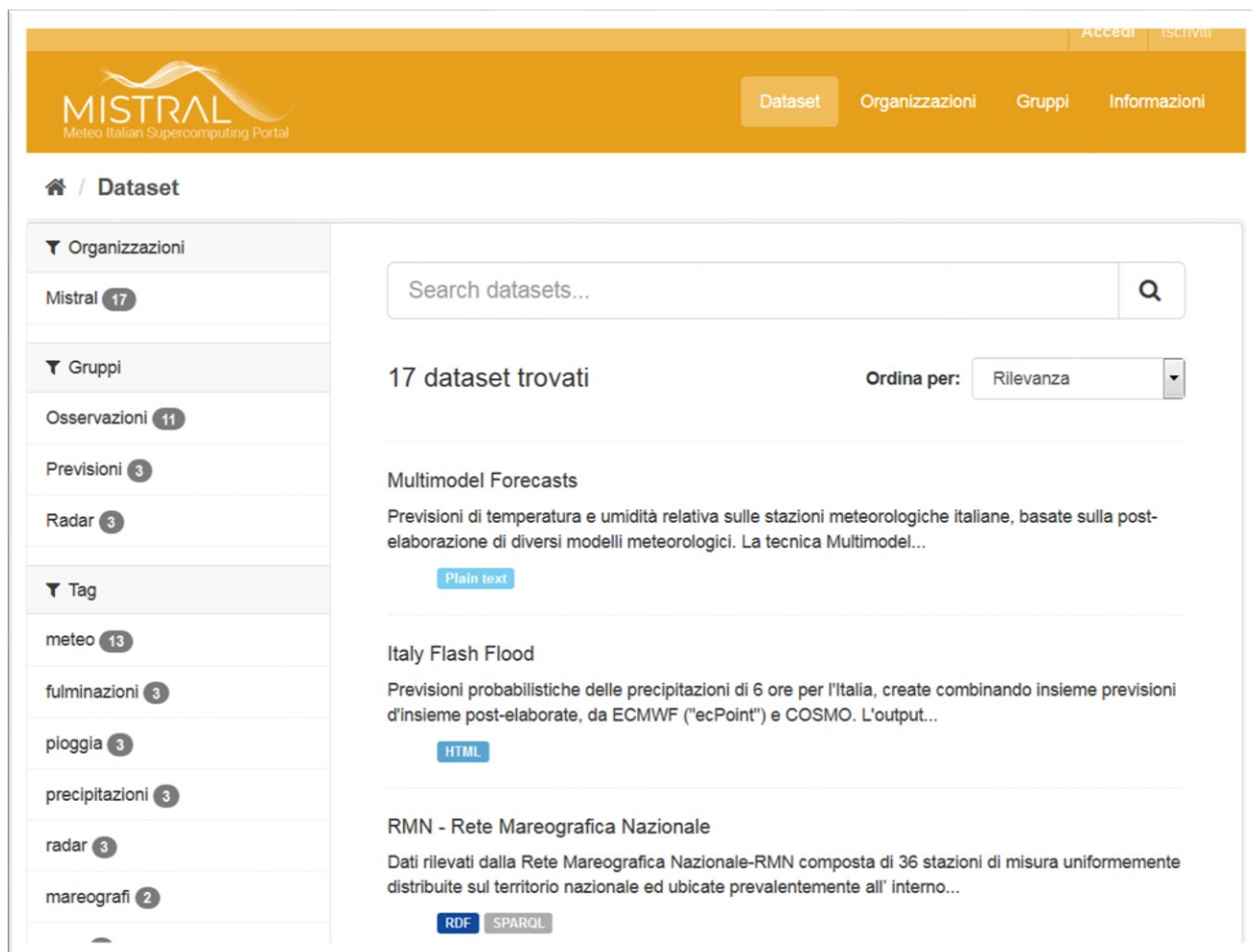
Previsioni
3 Dataset



Radar
3 Dataset

MISTRAL Project
CKAN API
[OPEN DATA](#)

 Co-financed by the Connecting Europe Facility of the European Union



The portal home page exposes all the services implemented in Mistral project
<https://www.mistralportal.it/opendata/>

The portal home page is composed of:

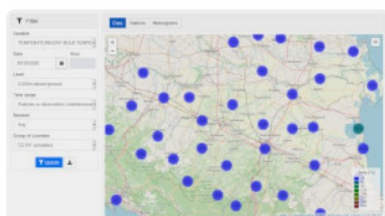
1. Header with menu
2. Mistral open services
 - a. Environmental Observation
 - b. Forecast Presentation
 - c. Weather Data DOWNLOADER
 - d. Italy Flash Flood
 - e. Multimodel Forecasts
 - f. MeteoTiles
3. Other open services
 - a. Radar & Realtime maps

4. FILTER BY DATASET
 - a. SEARCH BY FORECAST, OBSERVATION, REALTIME DATASET
 - b. Type to search weather data
5. Footer



MISTRAL OPEN SERVICES

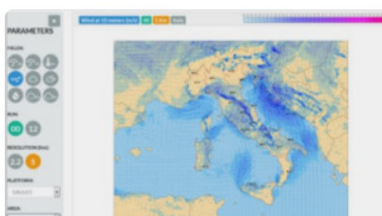
The MISTRAL partners provide free and professional ready-to-use services, you can find a constantly updated list below



ENVIRONMENTAL OBSERVATION

This Application provides observed meteorological data over Italian locations obtained from public weather station. Real time and historical data is provided by using a geospatial information system or displayed in interactive graph.

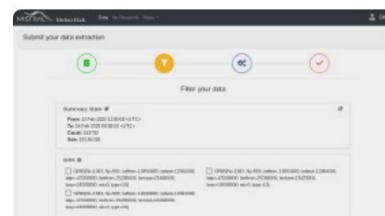
[Read more...](#)



FORECAST PRESENTATION

Application for forecast analysis, the user can choose parameters from a complete catalogue and move through the model time step. Creating a slideshow of information useful to make an accurate forecast overview on the Italian and Mediterranean area.

[Read more...](#)

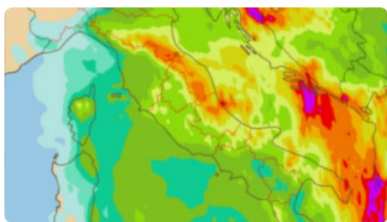


WEATHER DATA DOWNLOADER

Meteo-Hub Application allows you to create and download your personal weather data collection. Choose from different forecast models, weather stations, parameters and time validity range.

Download BUFR and GRIB format.

[Read more...](#)

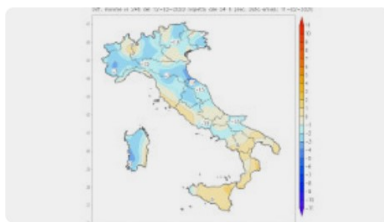


ITALY FLASH FLOOD

Application delivering probabilistic forecasts of 6-hour rainfall for Italy, created by blending together post-processed ensemble forecasts, from ECMWF ("ecPoint") and COSMO. Output shows pre-defined rainfall percentiles and probability thresholds for each COSMO grid point.

[Read more...](#)

OPEN

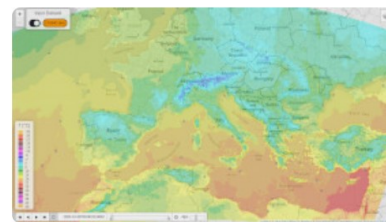


MULTIMODEL FORECASTS

Application based on the post-processing of different meteorological Models, for issuing temperature and relative humidity forecasts over Italian weather stations. The Multimodel SuperEnsemble technique is performed by a weight calculation after a performance evaluation during a training period

[Read more...](#)

OPEN



METEOTILES

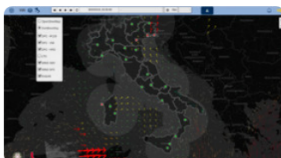
MeteoTiles application represent each meteorological variable as a web map side by side. This allows a dynamic visualization of the maps, allowing navigation and zooming, as well as the superimposition of multiple maps, each representing a different weather variable.

[Read more...](#)

OPEN

OTHER OPEN SERVICES

The goal of the MISTRAL portal is to facilitate and foster the re-user of the datasets by the weather community, as well as by its cross-area communities, to provide value services through the use of HPC resources, turning it into the level of new business opportunities.



RADAR & REALTIME MAPS

Realtime applications for nowcasting weather alert, multilayer information system, allow the user to create an interactive picture of the Italian and Mediterranean area filled with radar, wind animation and more

[Read more...](#)

FILTER BY DATASET

SEARCH BY FORECAST, OBSERVATION, REALTIME DATASET

The data will be available in terms of gridded fields, probabilistic products (as rainfall forecasts for flash flood prediction) or point time-series coming from the Italian operational forecast modelling chain and post-processing fields. All data are cataloged and freely downloadable using this area.



The project is funded under the Connecting Europe Facility (CEF) – Telecommunication Sector Programme of the European Union and started on 1 October 2018.

The goal of the MISTRAL portal is to facilitate and foster the re-use of the datasets by the weather community, as well as by its cross-area communities, to provide added value services through the use of HPC resources, turning it into the level of new business opportunities.



Co-financed by the Connecting Europe
Facility of the European Union



