



The Twelfth International Conference on Advanced
Geographic Information Systems, Applications and Services
GEOProcessing 2020
21-25 November 2020, Valencia (Spain)



Automatic Publication of Open Data from OGC Services: the Use Case of TRAFAIR Project

Javier Nogueras-Iso^(*), Héctor Ochoa-Ortiz, Manuel Ángel Jañez,
José R. R. Viqueira, Laura Po, Raquel Trillo-Lado

^(*) Aragon Institute of Engineering Research (I3A), Universidad de
Zaragoza (Spain)

jnog@unizar.es



Universidad
Zaragoza



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA



Javier Nogueras-Iso biography

- ❑ Javier Nogueras-Iso holds MS and PhD degrees in Computer Science from the University of Zaragoza.
- ❑ In 1998, he started his research at the Advanced Information Systems Laboratory of the University of Zaragoza. Currently, he is an Associate Professor of Computer Science at that University.
- ❑ Between 2011 and 2017 he was Director of Catedra Logisman on 'Technological Document Management', and between 2015 and 2019 he was Associate Director of the Aragon Institute of Engineering Research (I3A).
- ❑ His research interests are focused on Information Retrieval and Semantic Web technologies applied to different domains, although with a special emphasis on Geographic Information Infrastructures.



Outline

- ❑ 1. Introduction
- ❑ 2. A metadata profile for Open Spatial Data
- ❑ 3. Workflow for the publication of Open Spatial Data
- ❑ 4. Deployment of Open Data in Zaragoza, Santiago and Modena
- ❑ 5. Conclusion and future work



1. Introduction

- TRAFair: Understanding traffic flows to improve air quality
 - ❖ European project co-financed by the Connecting Europe Facility of the European Union
 - ❖ Project Nr. 2017-EU-IA-0167 (Nov 2018 – Oct 2020)
- 10 partners based in 6 cities from 2 European countries
 - ❖ 4 universities, 2 research centres
 - ❖ 3 public administrations
 - ❖ 1 regional in-house providing company



Project goals

- Design and develop the necessary infrastructure to **estimate** the pollution level on urban scale
 - ❖ Test the infrastructure in 6 European cities of different size
- Sub-goals:
 - ❖ Provide real-time **monitoring** of air pollution in the city on an urban scale
 - ❖ Develop an air quality **forecasting** service based on the weather forecasts and the urban traffic flow
 - ❖ Publish monitoring and forecasting air quality and traffic data as **open data**
 - ❖ Develop applications for end-users and public administrations

Focus of this work

❑ **Publication of Open Data**

❖ Monitoring data

- Low cost sensor data
- Coarse urban air pollution maps in semi-real time

❖ Forecast maps

- Dispersion maps of pollutants

❑ Given the spatial character of data, it was clear that Spatial Data Infrastructures (SDI) could provide the necessary technology for the management/publication of data

❑ However, publication through typical SDI services compliant with OGC is not enough for Open Data

❖ Data must be registered in Open Data portals

❖ TRAF AIR data should be visible through the European Data Portal (EDP)

Objectives of this work

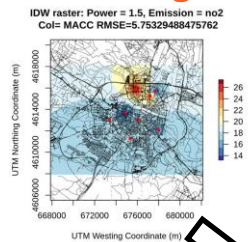
- ❑ Propose a workflow for the publication of Open Spatial Data
- ❑ Desirable features:
 - ❖ Automate as much as possible the generation of metadata without human interaction
 - ❖ Adopt a metadata model compatible with both the Open Data domain and the spatial domain
 - Use of GeoDCAT-AP
 - ❖ Provide a solution based on Open Source software packages
 - GeoServer: management of spatial data
 - CKAN: platform for deploying Open Data portals

Software components for the management and publication of data

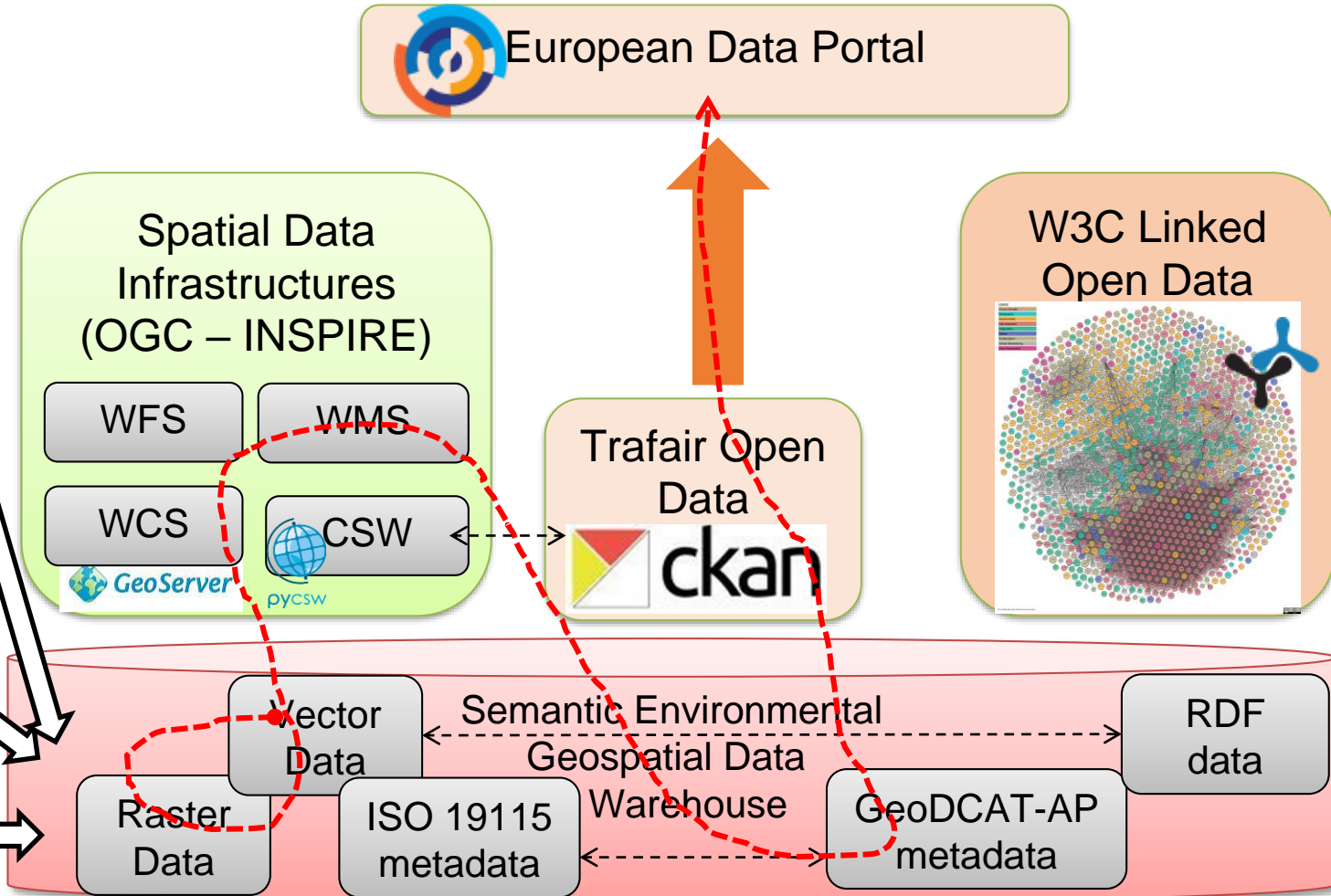
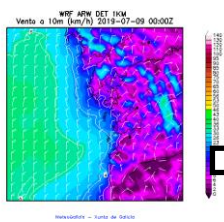
Low cost sensor
Data



Monitoring maps



Forecast maps

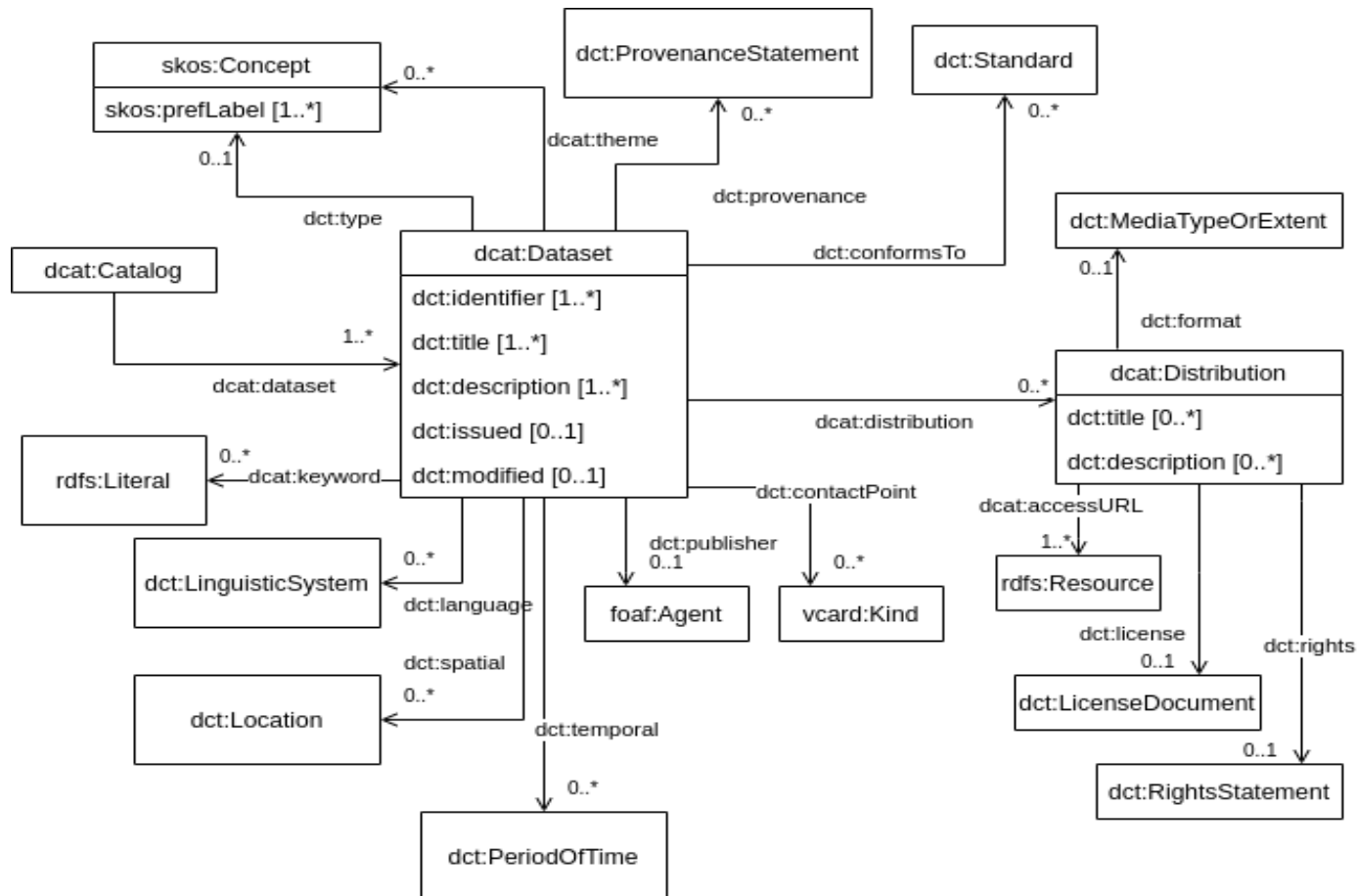


2. A metadata profile for Open Spatial Data

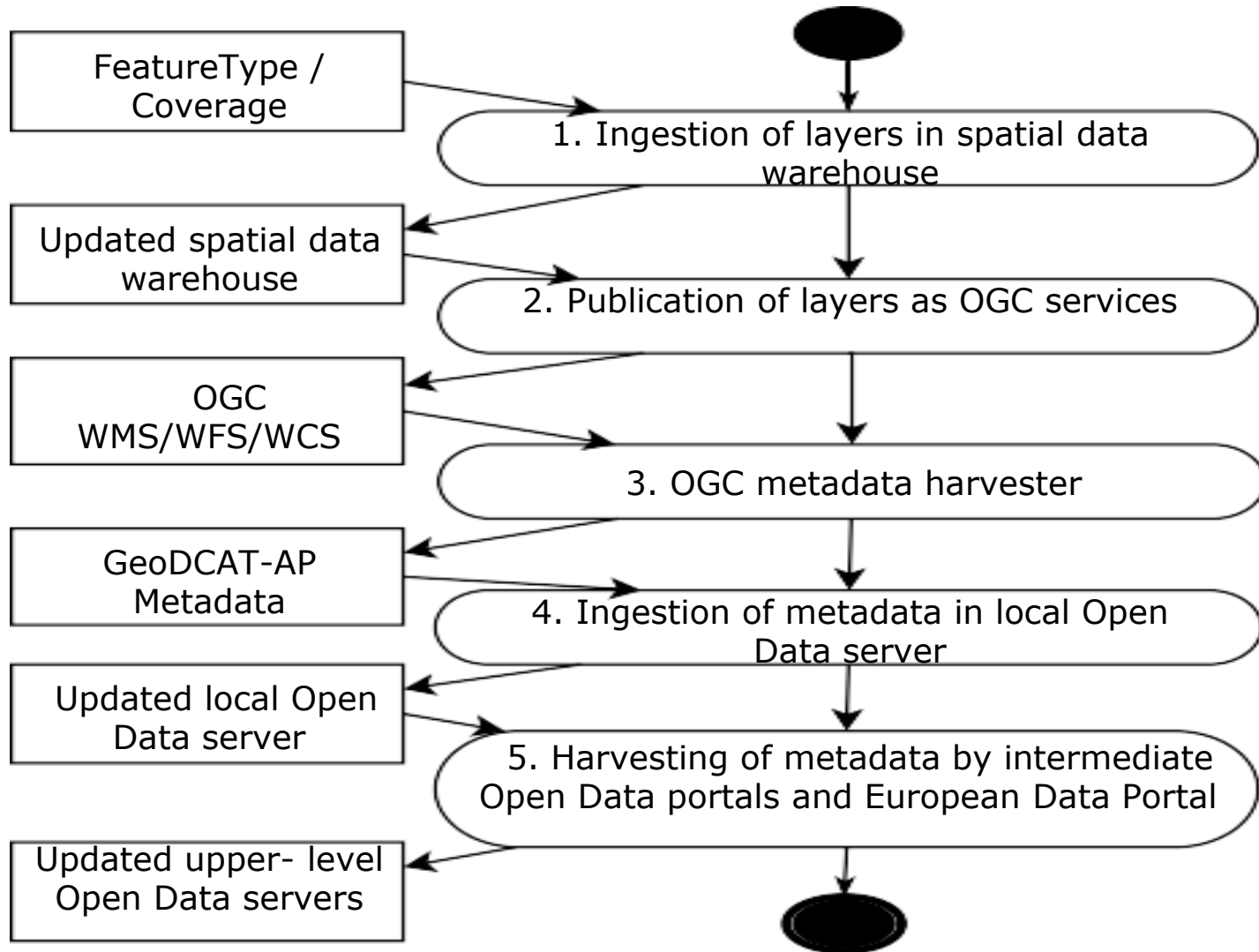
- ❑ Metadata model alternatives
 - ❖ ISO 19115: traditional metadata for geographic information datasets and services
 - large and complex model, whose automation is complicated
 - ❖ DCAT: Data Catalogue vocabulary proposed by W3C for the description of Open Data
 - ❖ DCAT-AP: European Application Profile of DCAT for public sector datasets
 - ❖ GeoDCAT-AP: extension of DCAT-AP for geographic information
 - Easily accepted in Open Data software
 - Designed to assure compliance with European INSPIRE directive for establishing a spatial information infrastructure in Europe (and ISO 19115)



- ❑ Light adoption of GeoDCAT-AP
 - ❖ Use mostly core properties (direct binding with ISO 19115-INSPIRE)
 - ❖ All properties compatible with DCAT-AP
 - ❖ All elements editable with CKAN software

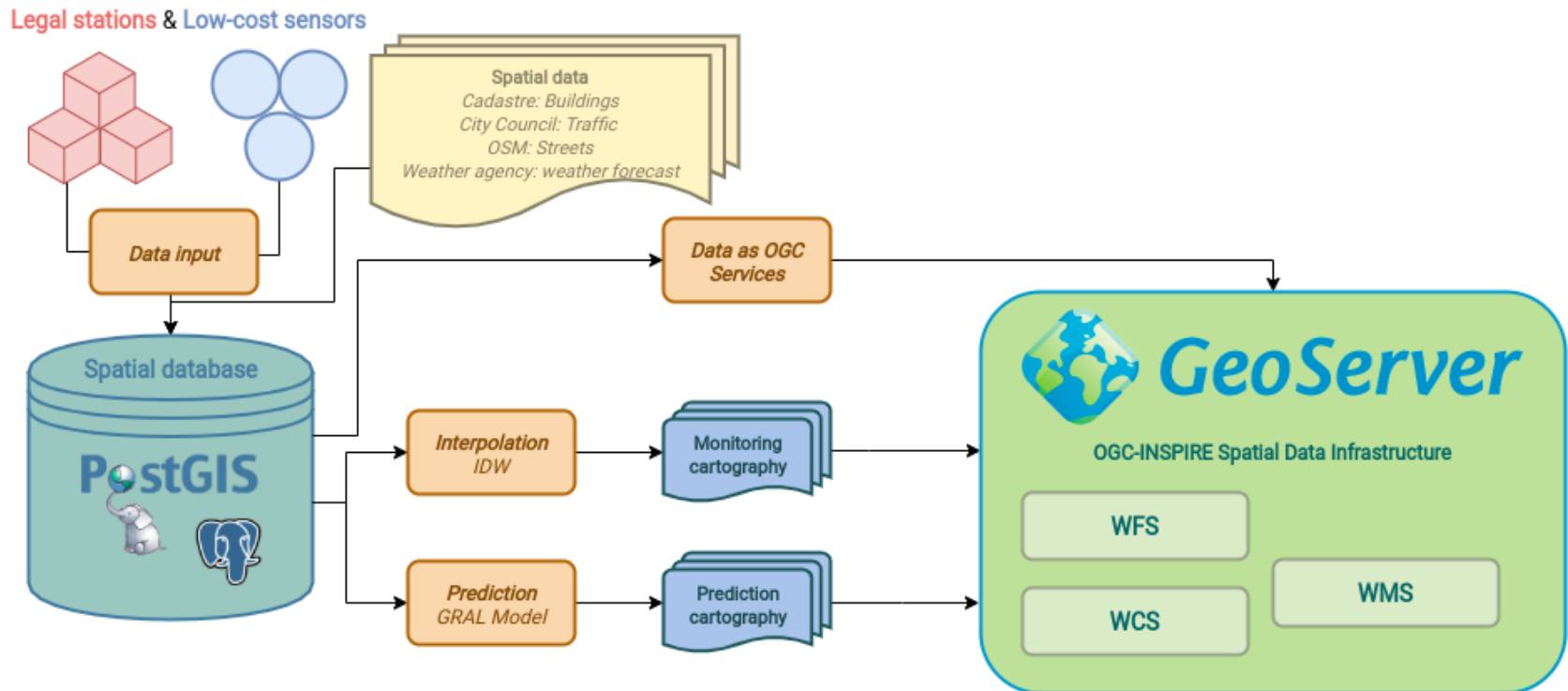


3. Workflow for the publication of Open Spatial Data



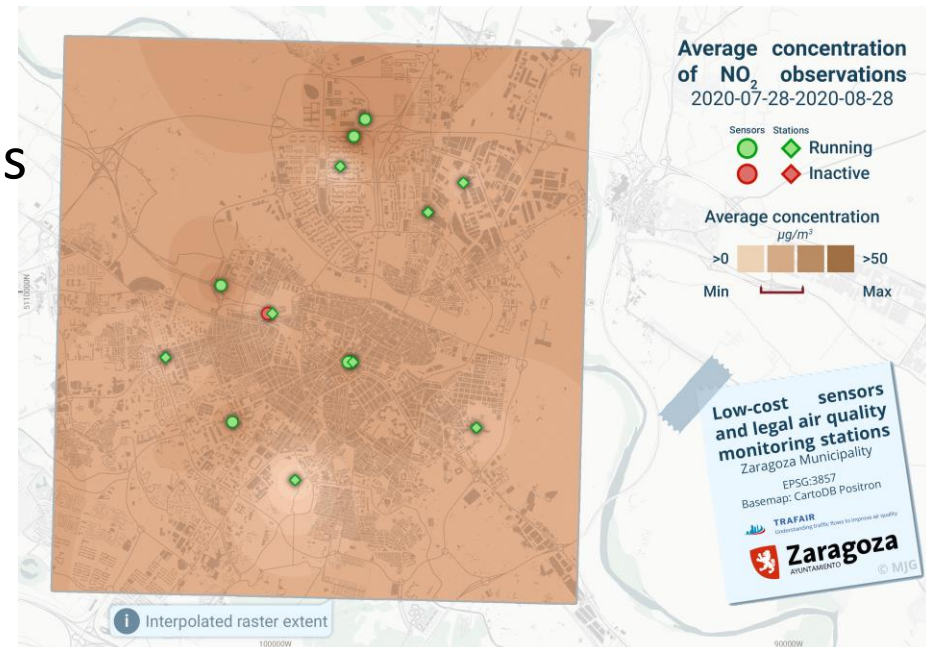
3.1. Ingestion of layers in spatial data warehouse

- ❑ We have developed specific software in Java and R languages to ingest layers in GeoServer
 - ❖ Feature types (monitoring observations)
 - ❖ Coverages (interpolation of observations, predictions based on a Lagrangian model for dispersion of pollutants)



3.2. Publication of layers as OGC services

- ❑ During the ingestion of layers, we can setup GeoServer to provide access to some layers through OGC services
- ❑ In the TRAFair case
 - ❖ Feature Types are downloaded through a Web Feature Service (WFS)
 - ❖ Coverages are downloaded through a Web Coverage Service (WCS)
 - ❖ In addition, some layers (coverages, traffic models , ...) are also available as maps through a Web Mapping Service (WMS)



3.3. OGC metadata harvester

- ❑ Metadata is harvested from OGC services through its *GetCapabilities* operation
- ❑ Python program using OWSLib package to retrieve *GetCapabilities*
- ❑ For each WCS/WFS layer we generate one Dataset and one Distribution
 - ❖ A second optional distribution if the layer is rendered through a WMS

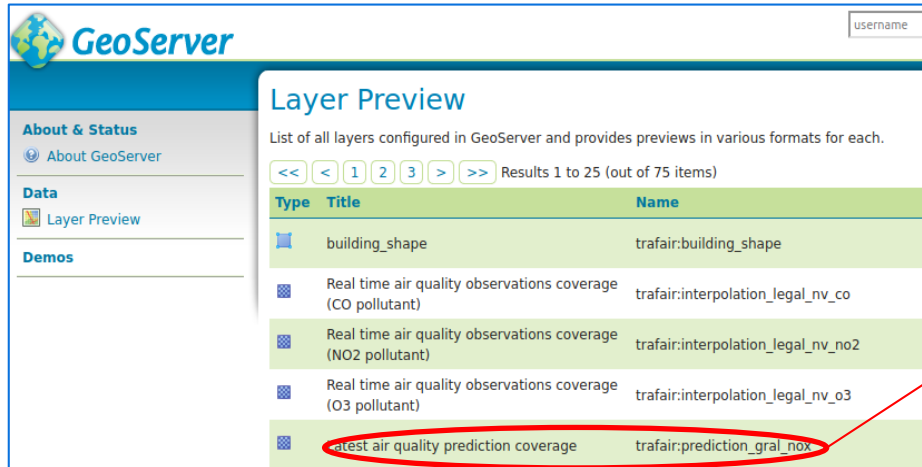
GeoServer	OWSLib	CKAN	GeoDCAT-AP
featureType/name, coverage/name	<i>layerName</i>	extra:identifier	Dataset/dct:identifier
featureType/title, coverage/title	contents[<i>layerName</i>].title	title	Dataset/dct:title
featureType/description, coverage/description <i>(software in step 1 introduces predefined descriptions according to name patterns)</i>	contents[<i>layerName</i>].abstract	notes	Dataset/dct:description
	<i>("series" for OGC services with temporal dimension, or "dataset" without temporal dimension)</i>	extra:dcats type	Dataset/dct:type
	<i>(default language proposed in step 3)</i>	extra:language	Dataset/dct:language
	<i>(default INSPIRE data themes and ISO 19115 topic categories proposed in step 3)</i>	extra:theme	Dataset/dcat:theme
<i>(some default keywords are automatically introduced by GeoServer)</i>	contents[<i>layerName</i>].keywords	tags	Dataset/dcat:keyword
<i>(computed automatically by GeoServer)</i>	contents[<i>layerName</i>].boundingBoxWGS84	extra:spatial	Dataset/dct:spatial
<i>(start date and end date are automatically updated by GeoServer)</i>	contents[<i>layerName</i>].timepositions	extra:temporal start + extra:temporal end	Dataset/dct:temporal
		extra:issued <i>(automatically inserted with first ingestion in CKAN)</i>	Dataset/dct:issued
		extra:modified <i>(automatically updated with every update of a dataset in CKAN)</i>	Dataset/dct:modified
	<i>(default provenance proposed in step 3)</i>	extra:provenance	Dataset/dct:provenance
	<i>(default INSPIRE for data distribution)</i>	extra:provenance	Dataset/dct:provenance

3.4. Ingestion of metadata in local Open Data server

- ❑ Our Python program transforms the information retrieved in the previous step into a dictionary with the required items to construct a dataset and its associated resources, which can be ingested through the REST API of CKAN
- ❑ We select the appropriate tags to assure that GeoDCAT-AP RDF can be generated with *ckanext-dcat* plugin

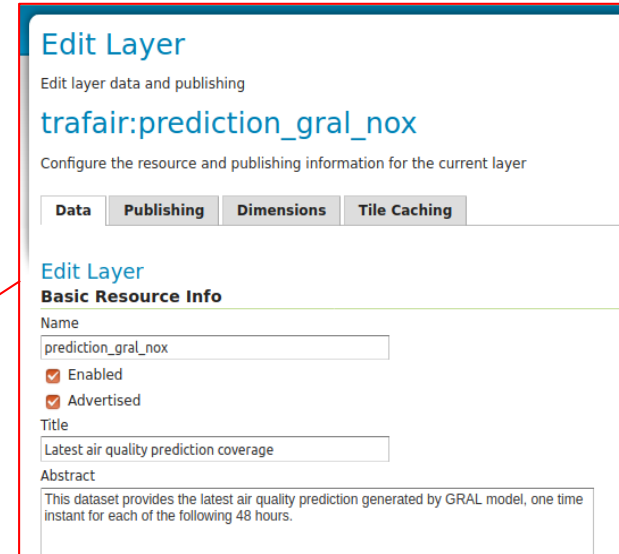
GeoServer	OWSLib	CKAN	GeoDCAT-AP
featureType/name, coverage/name	<i>layerName</i>	extra:identifier	Dataset/dct:identifier
featureType/title, coverage/title	contents[<i>layerName</i>].title	title	Dataset/dct:title
featureType/description, coverage/description <i>(software in step 1 introduces predefined descriptions according to name patterns)</i>	contents[<i>layerName</i>].abstract	notes	Dataset/dct:description
	<i>("series" for OGC services with temporal dimension, or "dataset" without temporal dimension)</i>	extra:dcat type	Dataset/dct:type
	<i>(default language proposed in step 3)</i>	extra:language	Dataset/dct:language
	<i>(default INSPIRE data themes and ISO 19115 topic categories proposed in step 3)</i>	extra:theme	Dataset/dcat:theme
<i>(some default keywords are automatically introduced by GeoServer)</i>	contents[<i>layerName</i>].keywords	tags	Dataset/dcat:keyword
<i>(computed automatically by GeoServer)</i>	contents[<i>layerName</i>].boundingBoxWGS84	extra:spatial	Dataset/dct:spatial
<i>(start date and end date are automatically updated by GeoServer)</i>	contents[<i>layerName</i>].timepositions	extra:temporal start + extra:temporal end	Dataset/dct:temporal
		extra:issued <i>(automatically inserted with first ingestion in CKAN)</i>	Dataset/dct:issued
		extra:modified <i>(automatically updated with every update of a dataset in CKAN)</i>	Dataset/dct:modified
	<i>(default provenance proposed in step 3)</i>	extra:provenance	Dataset/dct:provenance

Example of steps 1-4



GeoServer Layer Preview interface showing a list of layers. The layer 'latest air quality prediction coverage' (traffair:prediction_gral_nox) is highlighted with a red circle.

Type	Title	Name
building_shape		traffair:building_shape
Real time air quality observations coverage (CO pollutant)		traffair:interpolation_legal_nv_co
Real time air quality observations coverage (NO2 pollutant)		traffair:interpolation_legal_nv_no2
Real time air quality observations coverage (O3 pollutant)		traffair:interpolation_legal_nv_o3
Real time air quality prediction coverage		traffair:prediction_gral_nox



GeoServer Edit Layer interface for the layer 'traffair:prediction_gral_nox'. The 'Basic Resource Info' section is visible, showing the layer name, title, and abstract.

Edit Layer
traffair:prediction_gral_nox

Configure the resource and publishing information for the current layer

Basic Resource Info

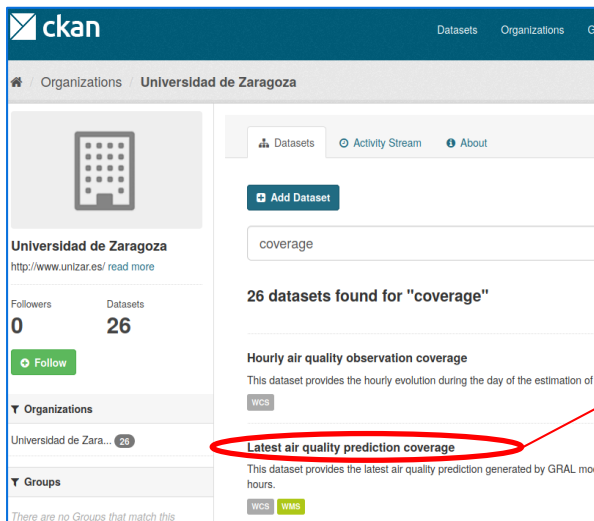
Name: prediction_gral_nox

Enabled

Advertised

Title: Latest air quality prediction coverage

Abstract: This dataset provides the latest air quality prediction generated by GRAL model, one time instant for each of the following 48 hours.



CKAN interface showing search results for 'coverage'. The dataset 'Latest air quality prediction coverage' is highlighted with a red circle.

ckan

Universidad de Zaragoza

26 datasets found for "coverage"

Hourly air quality observation coverage

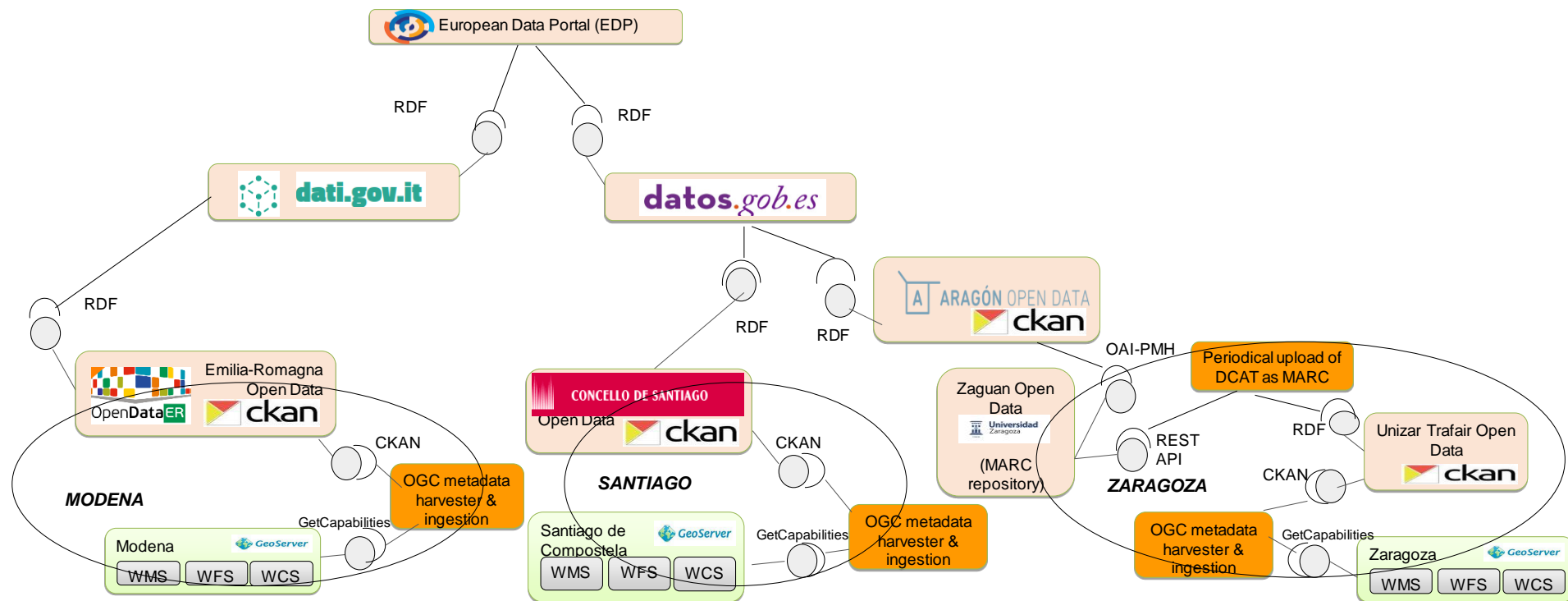
Latest air quality prediction coverage

```
<dc:Dataset rdf:about="http://atila.unizar.es:3394/dataset/dbb3799e-dc88-4e8f-995b-7db">
  <dc:title>Latest air quality prediction coverage</dc:title>
  <dc:description>This dataset provides the latest air quality prediction generated by
  <dc:type>http://inspire.ec.europa.eu/metadata-codelist/ResourceType/series</dc:type>
  <dc:theme rdf:resource="http://inspire.ec.europa.eu/metadata-codelist/TopicCategory/
  <dc:temporal>
    <dc:PeriodOfTime rdf:nodeID="N4229e255693543359f8bdad8d23747c4">
      <schema:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2019-07
      <schema:endDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2020-10-3
    </dc:PeriodOfTime>
  </dc:temporal>
  <dc:theme rdf:resource="http://inspire.ec.europa.eu/theme/ac"/>
  <dc:keyword>TRAFAIR</dc:keyword>
  <dc:keyword>prediction_gral_nox</dc:keyword>
</dc:Dataset>
```

3.5. Harvesting of metadata by intermediate Open Data portals and European Data Portal

- ❑ This step is beyond the scope of the TRAFair project.
- ❑ Our local CKAN servers include the installation of *ckanext-dcat* plugin
 - ❖ This allows the publication of datasets metadata as RDF in compliance with DCAT-AP vocabularies
- ❑ We assume that upper-level portals are based on CKAN technology (or have a similar mechanism for the harvesting of subscribed lower level catalogs)
 - ❖ The *ckanext-harvest* plugin of CKAN allow to harvest the contents of different types of catalog sources
 - ❖ Most regional/national portals based on DCAT interact with lower-level CKAN servers through its catalog endpoint to retrieve contents as RDF

4. Deployment of Open Data in Zaragoza, Santiago and Modena



An example of the harvesting process (I)



Organizations

Universidad de Zara... **109**

Groups

There are no Groups that match this search

Tags

TRAFAIR **109**

WCS **70**

ImageMosaic **65**

features **23**

GeoTIFF **5**

[+ Add Dataset](#)

109 datasets found

Order by:



Latest air quality prediction coverage

Followers

0

[Follow](#)

Organization



Universidad de Zaragoza

<http://www.unizar.es/> [read more](#)

Social

[Twitter](#)

[Facebook](#)

[License](#)

[Dataset](#)[Groups](#)[Activity Stream](#)[Manage](#)

Latest air quality prediction coverage

This dataset provides the latest air quality prediction generated by GRAL model, one time instant for each of the following 48 hours.

Data and Resources



open_data_latest_air_quality_prediction_coverage

Spatial resolution (distance): 4 m

[Explore](#)

open_data:latest_air_quality_prediction_coverage


[Explore](#)

[ImageMosaic](#) [TRAFAIR](#) [WCS](#) [prediction_gral_nox](#)

Additional Info

Field	Value
State	active
Last Updated	October 16, 2020, 6:03 AM (UTC+02:00)
Created	October 10, 2020, 9:03 AM (UTC+02:00)
Conforms to	https://inspire.ec.europa.eu/documents/commission-regulation-eu-no-13122014-10-december-2014-amending-regulation-eu-no-10892010-0 , https://inspire.ec.europa.eu/documents/inspire-metadata-regulation ,

An example of the harvesting process (II)



Universidad Zaragoza

ZAGUAN

Universidad de Zaragoza Repository

Submit Personalize Help EN / ES

Information References (0) Citations (0) Keywords Discussion (0) Usage statistics Files Plots Holdings Linkbacks

TRAFair-2020-027

Latest air quality prediction coverage

RESUMEN: This dataset provides the latest air quality prediction generated by GRAL model, one time instant for each of the following 48 hours.

DATA TYPE: <http://inspire.ec.europa.eu/metadata-codelist/ResourceType/series>


THEME: <http://inspire.ec.europa.eu/metadata-codelist/TopicCategory/environment>

KEYWORDS: TRAFair ; WCS ; ImageMosaic ; prediction_gral_nox

LANGUAGE: <http://publications.europa.eu/resource/authority/language/ENG>

PUBLISHED BY: Universidad de Zaragoza

PUBLICATION DATE: 10/10/2020

 You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. You may not use the material for commercial purposes.

Este registro pertenece a la(s) siguiente(s) colección(es): [Open data](#) > [Research Opendata](#) > [Environmental Research Opendata Opendata \(Other\)](#)

An example of the harvesting process (III)

A ARAGÓN OPEN DATA

Latest air quality prediction coverage



Información del conjunto de datos

Título

Latest air quality prediction coverage

Descripción

This dataset provides the latest air quality prediction generated by GRAL model, one time instant for each of the following 48 hours.

Categoría

Educación

Etiquetas y palabras clave

Imagemosaic

Prediction_gral_nox

Trafair

Wcs

Organización responsable de la publicación

[Universidad de Zaragoza](#)

Territorio en el que existen datos

Este conjunto de datos contiene datos en el territorio de Aragón
Esta es la [URI del recurso](#)

Periodo de tiempo en el que existen datos

2019-07-22 - 2020-10-31

Licencia

[Creative Commons Attribution-NonCommercial 4.0](#)

Descargas

Archivos con los datos

Latest air quality prediction coverage

URL

URL

URL

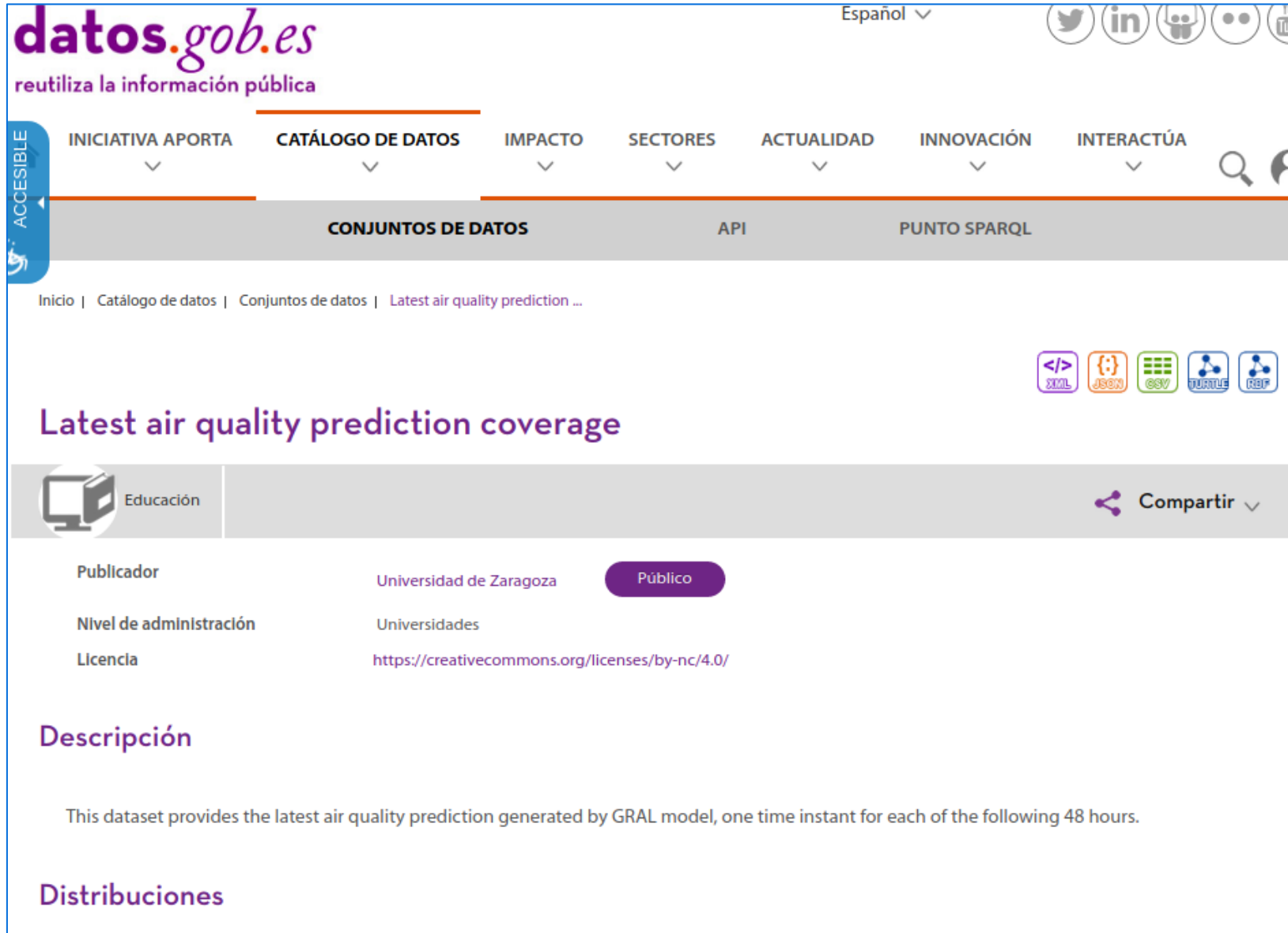
Metadatos de la colección

RDF

Valora estos datos



An example of the harvesting process (IV)



datos.gob.es
reutiliza la información pública

Español ▾

INICIATIVA APORTA | **CATÁLOGO DE DATOS** | IMPACTO | SECTORES | ACTUALIDAD | INNOVACIÓN | INTERACTÚA

CONJUNTOS DE DATOS | API | PUNTO SPARQL

Inicio | Catálogo de datos | Conjuntos de datos | Latest air quality prediction ...

XML | JSON | CSV | TURTLE | RDF

Latest air quality prediction coverage

Educación

Compartir ▾

Publicador Universidad de Zaragoza **Público**

Nivel de administración Universidades

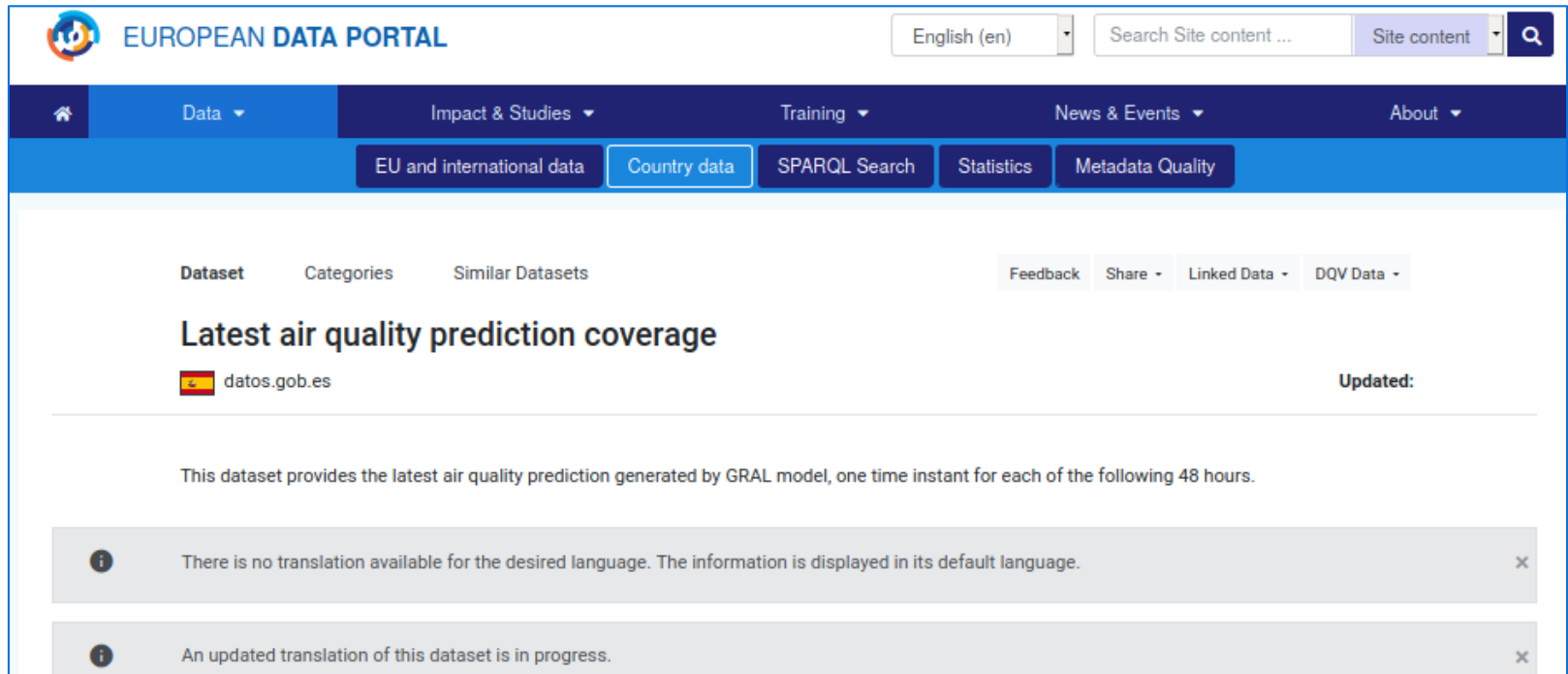
Licencia <https://creativecommons.org/licenses/by-nc/4.0/>

Descripción

This dataset provides the latest air quality prediction generated by GRAL model, one time instant for each of the following 48 hours.

Distribuciones

An example of the harvesting process (V)



EUROPEAN DATA PORTAL


English (en) Search Site content ... Site content

Data Impact & Studies Training News & Events About

EU and international data Country data SPARQL Search Statistics Metadata Quality

Dataset Categories Similar Datasets Feedback Share - Linked Data - DQV Data -

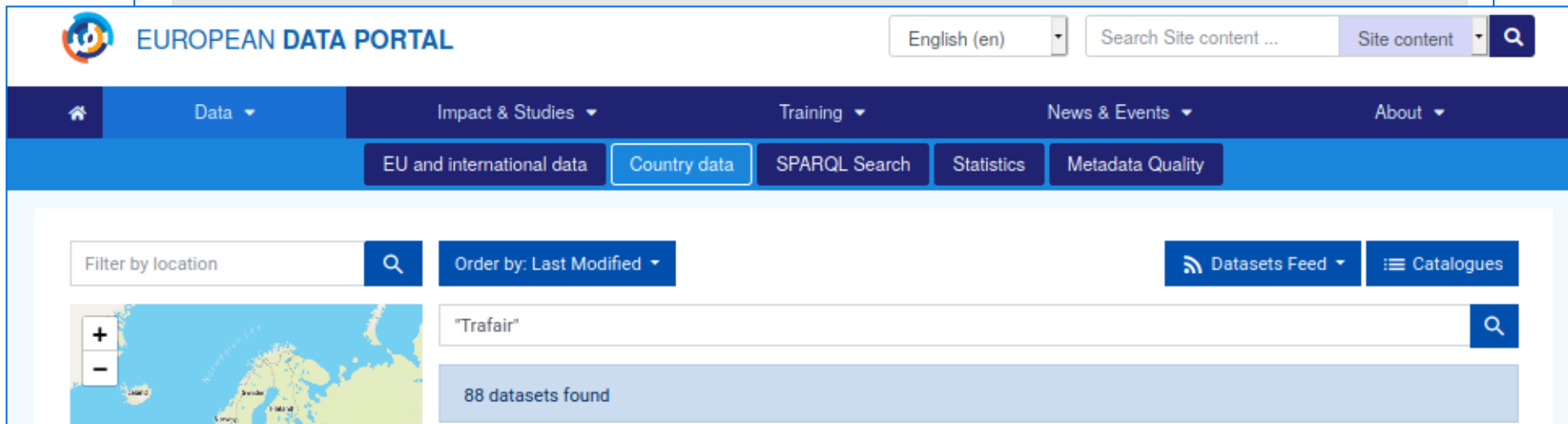
Latest air quality prediction coverage

 datos.gob.es Updated:

This dataset provides the latest air quality prediction generated by GRAL model, one time instant for each of the following 48 hours.

There is no translation available for the desired language. The information is displayed in its default language.

An updated translation of this dataset is in progress.







EUROPEAN DATA PORTAL

English (en) Search Site content ... Site content

Data Impact & Studies Training News & Events About



EU and international data Country data SPARQL Search Statistics Metadata Quality

Filter by location  Order by: Last Modified 

"Trafair"

88 datasets found

 Datasets Feed  Catalogues

5. Conclusion and future work (I)

- ❑ We have proposed the workflow for the publication of Open Spatial Data that solves jointly 3 challenges:
 - ❖ the automatic generation of metadata from the *GetCapabilities* responses of OGC services
 - ❖ the generation of DCAT-based metadata
 - ❖ the synchronized publication of data and metadata
- ❑ We have demonstrated how GeoDCAT-AP metadata can be applied in a real use case to describe more specifically spatial data than other more general metadata vocabularies based on DCAT
- ❑ This workflow can be customized to other projects dealing with spatial data that must be publicly accessible

5. Conclusion and future work (II)

□ Lessons learned

- ❖ Not all GeoDCAT-AP guidelines to fill metadata elements could be followed by local Open Data portals because they must comply with constrained profiles of DCAT-AP imposed by national governments
 - Instead of INSPIRE data themes (i.e. “atmosphere” and “environment facilities”), local portals must choose a theme from a national controlled vocabulary
 - Some elements are not allowed in national profiles (e.g. *dct:provenance*)

□ Future work

- ❖ Integrate software as a new plugin of CKAN, or as an extension of *ckanext-spatial*
- ❖ Evaluate quality of generated metadata through different methods (e.g. EDP Metadata Quality Assessment methodology)

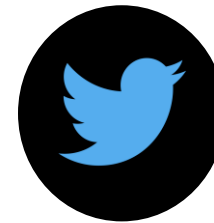


TRAFair

Understanding traffic flows to improve air quality



<http://trafair.eu/>



@trafairproject



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA



UNIVERSITÀ
DEGLI STUDI
FIRENZE

DINFO
DIPARTIMENTO DI
INGEGNERIA
DELL'INFORMAZIONE

DISIT
DISTRIBUTED SYSTEMS
AND INTERNET
TECHNOLOGIES LAB



Universidad
Zaragoza



Consiglio Nazionale
delle Ricerche
Istituto per la BioEconomia



CONCELLO DE
SANTIAGO



REGIONE
TOSCANA



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