D3.9 – TransformingTransport Open Data Portal

<table>
<thead>
<tr>
<th>Project Acronym</th>
<th>TT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Title</td>
<td>Transforming Transport</td>
</tr>
<tr>
<td>Grant Agreement number</td>
<td>731932</td>
</tr>
<tr>
<td>Call and topic identifier</td>
<td>ICT-15-2016</td>
</tr>
<tr>
<td>Funding Scheme</td>
<td>Innovation Action (IA)</td>
</tr>
<tr>
<td>Project duration</td>
<td>30 Months [1 January 2017 – 30 June 2019]</td>
</tr>
<tr>
<td>Coordinator</td>
<td>Mr Rodrigo Castiñeira (INDRA)</td>
</tr>
<tr>
<td>Website</td>
<td><a href="http://www.transformingtransport.eu">www.transformingtransport.eu</a></td>
</tr>
</tbody>
</table>

Funded by the European Union’s H2020 GA - 731932
### Document fiche

<table>
<thead>
<tr>
<th>Author:</th>
<th>Francisco Yedro [UPM] – editor; Oscar Corcho [UPM] – contributor; Víctor Rodríguez Doncel [UPM] – contributor;</th>
</tr>
</thead>
</table>
| Internal reviewers:         | Philipp Bohn [UDE]
                            | Pablo Fernández Vivanco [CI3] |
| Work Package:               | WP3                                                                 |
| Task:                       | T3.3                                                                |
| Nature:                     | ORDP                                                               |
| Dissemination:              | PU                                                                  |

### Document History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Contributor(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>16/05/2017</td>
<td>UPM</td>
<td>Document Creation</td>
</tr>
<tr>
<td>0.2</td>
<td>09/06/2017</td>
<td>UPM</td>
<td>Internal Review</td>
</tr>
<tr>
<td>0.3</td>
<td>14/06/2017</td>
<td>UPM</td>
<td>Little changes</td>
</tr>
<tr>
<td>0.4</td>
<td>15/06/2017</td>
<td>UPM</td>
<td>Final document for submission to internal peer review process</td>
</tr>
<tr>
<td>1.0</td>
<td>30/06/2017</td>
<td>UPM</td>
<td>Document reviewed, ready to submission</td>
</tr>
</tbody>
</table>
### Keywords:
Open, Data, Portal, CKAN, Data Management Plan, Dataset, Metadata, Domain, Pilot, Organization, Group, User.

### Abstract:
This deliverable describes the Open Data Portal of the project.

---

**DISCLAIMER**

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content. This document may contain material, which is the copyright of certain TT consortium parties, and may not be reproduced or copied without permission. All TT consortium parties have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the TT consortium as a whole, nor a certain party of the TT consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and does not accept any liability for loss or damage suffered by any person using this information.

---

**ACKNOWLEDGEMENT**

This document is a deliverable of TT project. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement Nº 731932
# Table of Contents

- **DEFINITIONS, ACRONYMS AND ABBREVIATIONS** ................................................................. 5
- **EXECUTIVE SUMMARY** ........................................................................................................ 6
- **1 INTRODUCTION** ................................................................................................................ 7
- **2 TRANSFORMINGTRANSPORT DATA PORTAL** ................................................................. 9
  - **2.1 INTRODUCTION** ........................................................................................................ 9
  - **2.2 WHY AN OPEN DATA PORTAL?** ............................................................................... 9
  - **2.3 CKAN** ...................................................................................................................... 9
- **3 DATA PORTAL GUIDE** ......................................................................................................... 14
  - **3.1 NEW USER** ............................................................................................................ 14
  - **3.2 USER’S DASHBOARD** .............................................................................................. 15
    - **3.2.1 News feed** ......................................................................................................... 15
    - **3.2.2 My Datasets** .................................................................................................... 15
    - **3.2.3 My Organizations** ............................................................................................ 16
    - **3.2.4 My Groups** ..................................................................................................... 17
  - **3.3 DATA PORTAL MENU BAR** ....................................................................................... 18
    - **3.3.1 Datasets** .......................................................................................................... 18
    - **3.3.2 Organizations** .................................................................................................. 19
    - **3.3.3 Groups** ............................................................................................................ 20
    - **3.3.4 About** ............................................................................................................... 21
  - **3.4 CREATING A DATASET** ............................................................................................ 21
  - **3.5 FINDING DATA** ........................................................................................................ 24
- **CONCLUSIONS AND NEXT STEPS** .................................................................................... 25
Definitions, Acronyms and Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>CKAN</td>
<td>Comprehensive Knowledge Archive Network</td>
</tr>
<tr>
<td>DMP</td>
<td>Data Management Plan</td>
</tr>
<tr>
<td>ODP</td>
<td>Open Data Portal</td>
</tr>
<tr>
<td>TT</td>
<td>Transforming Transport</td>
</tr>
<tr>
<td>WP</td>
<td>Work Package</td>
</tr>
</tbody>
</table>

Table of Figures

Figure 1. Dataset Access Pie Chart ................................................................. 7
Figure 2. Dataset Personal Data Pie Chart ....................................................... 8
Figure 3. Landing page of the TT data portal .................................................... 10
Figure 4. Register for an Account page ............................................................. 14
Figure 5. User options ....................................................................................... 15
Figure 6. Dashboard – News Feed ....................................................................... 15
Figure 7. Dashboard – My Datasets ..................................................................... 16
Figure 8. Dashboard – My Organizations ............................................................... 16
Figure 9. Dashboard – My Groups ....................................................................... 17
Figure 10. Menu Bar ......................................................................................... 18
Figure 11. Datasets page ................................................................................... 18
Figure 12. Organizations page ........................................................................... 19
Figure 13. Groups page ..................................................................................... 19
Figure 14. About page ....................................................................................... 21
Figure 15. Create a dataset - metadata ............................................................... 22
Figure 16. Create a dataset – add data ............................................................... 23
Figure 17. Editing dataset ................................................................................ 23
Figure 18. Search dataset ................................................................................ 24
Executive Summary

This document belongs to the framework of WP3 (Impact) of the TransformingTransport (TT) project and describes the Open Data Portal (ODP) which belongs to task T3.3.

This deliverable is developed to be an easy guide of the operation of the data portal and its objective is to describe the concept of data portal as well as to show each of the tool sections, organizations, groups and possible roles.

Firstly, a brief information about the number and type of datasets to be included is given. After, the concept of ODP is introduced and a justification of using an ODP and the choice of CKAN as ODP platform is given. Finally, a brief user’s guide is included with the basics functionalities of the ODP such as creating a new user, managing the dashboard, menu bar options, creating (and editing) a dataset and finding data. The document closes with some brief conclusions and the next steps to follow.

The data portal is available since the end of June 2017, and probably evolve throughout the life of the project to adapt the new needs that may arise.
1 Introduction

The purpose of this deliverable is to give an overview of what a data portal is and the functionalities that it offers. This document is the second part of the Task 3.3 which continues the first overall objective, the development of the project Data Management Plan (DMP). As a reminder, this DMP, released on deliverable 3.2, is a formal short plan that outlines what data will be generated or collected, how data will be managed (access, storage, backup...), the standards in use, the workflow to make the data accessible for use, reuse and verification and which plans for data sharing and preservation exist ensuring that data are well-managed.

A data portal is a web-based system that makes a collection of datasets available. In the case of open data portals, everyone can access to the data present in the portal. It is not necessary to log in in order to consult data, although it is compulsory to publish data.

The Transforming Transport ODP, will gather all the datasets that are being used in the context of the Transforming Transport pilots. For all datasets, their corresponding metadata is recorded, and in the case that such datasets can be made available openly for further reuse (e.g., because they come from other open data portals from municipalities or other public institutions), they will be also stored in the ODP. In the case of data available through APIs, URLs of such data access will be made available as part of the metadata, with the data stored in other data storage and access systems. Finally, for data that cannot be disclosed, only basic metadata will be made available.

So far today, a list of 142 datasets has been collected of which 41 have been identified as open data and 89 need approval (12 unknown).
Only 9 datasets have been confirmed containing personal data whilst 124 don’t contain personal data (9 unknown).

![Dataset Personal Data Pie Chart]

**Figure 2. Dataset Personal Data Pie Chart**

Regarding the type of the technology in which the datasets are made available, there is no homogeneity in the formats that are being used. The following formats and technologies are available within the portal: Web Services, REST APIs, Java, XML, XLS, HTML, XML, CSV and even PDF, depending on the pilot or the data provider. During the course of the project we will aim at achieving a larger degree of homogenization in this respect.

In Section 2 the configuration and deployment of a data portal based on CKAN will be described whilst in Section 3 a brief user guide will be included. Besides, as an example, an initial collection of datasets will be detailed.
2 TransformingTransport Data Portal

2.1 Introduction
An ODP is a list of datasets in a standard format that is readable by virtually all computer systems. Its purpose is to increase access to public data or the data used in the project, and to make it accessible to users in a conventional method. A data portal is not only a data repository. It also allows users to discovering, analysing and visualizing. Some of the most common features provided by data portals are:

- Create, modify and delete datasets.
- Search for datasets.
- Addition of metadata.
- Integration with other data repository.
- Notification of the inclusion of new datasets.
- Generation of APIs to allow more sophisticated queries.
- Integration with tools for visualizing and manipulating data.

2.2 Why an Open Data Portal?
The current situation shows us a dispersion of the data between all the pilots, one of the aims of this ODP is to get a single point of access of the datasets. Also, there are various ways to describe data but with the OPD all the datasets will be under common rules for describing data, promoting the use of open formats.

The open data could generate economic benefit because it enables the creation of new business models, allowing pilots to reuse this data to create new products and services.

There are some technical solutions to implement open data portals such as CKAN\(^1\), Socrata\(^2\) and OpenData Soft\(^3\). We will now describe CKAN, the solution that we have adopted for implementing de TT data portal.

2.3 CKAN
We have chosen CKAN as a technical solution since it is a complete open source software solution that makes data accessible and usable (by providing tools to streamline publishing, sharing, finding and using data, including storage of data and provision of robust data APIs),

---

\(^1\) CKAN Open Source data portal platform - https://ckan.org/
\(^2\) Socrata Platform - https://socrata.com/
\(^3\) Open Data Soft platform - https://www.opendatasoft.com/
that you can use without any license fees, and you retain all rights to the data and metadata you enter. It has additional advantages like:

- It can be extended with more than 300 open source extensions.
- It can plot geographic data in interactive maps, what may be relevant in the context of TT.
- It has an active developer community.
- It can work as data harvester, pulling data from other sources, in our case, initiatives.
- It has an API to interact with the resources of the portal (organizations, datasets, groups, etc.)

The CKAN style has been adapted to TT corporate image and it is accessible through the following address:

**Data portal address:** data.transformingtransport.eu

The following image shows the landing page of TT data portal:

---

**Figure 3. Landing page of the TT data portal**
The next step is to organize the structure of the portal. To do so, it is necessary to define some CKAN entities like organizations, groups and datasets that will be integrated in the portal. Note that datasets in CKAN only can be created within an organization (and must belong to exactly one organization). According to CKAN documentation⁴ we have the following structure:

- **Organizations** in CKAN are used to create, manage and publish collections of datasets. Users can have different roles within an Organization, depending on their level of authorization to create, edit and publish. Organizations are controlling which users can add, update and publish which datasets. When a user joins an organization, an organization admin gives them one of three roles: **member**, **editor** or **admin** with the following level of management:

  An **admin** can:
  
  - View the organization’s private datasets
  - Add new datasets to the organization
  - Edit or delete any of the organization’s datasets
  - Make datasets public or private
  - Add users to the organization, and choose whether to make the new user a member, editor or admin
  - Change the role of any user in the organization, including other admin users
  - Remove members, editors or other admins from the organization
  - Edit the organization itself (for example: change the organization’s title, description or image)
  - Delete the organization
  - Create new organizations

  An **editor** can:
  
  - View the organization’s private datasets
  - Add new datasets to the organization
  - Edit or delete any of the organization’s datasets

  A **member** can view the organization’s private datasets.

When a user creates a new organization, that user becomes the first admin of that organization.

Regarding datasets, these can be marked as public or private where public datasets are visible to everyone and private datasets can only been seen by logged-in users who are

---

⁴ CKAN documentation - http://docs.ckan.org
members of the dataset’s organization. Private datasets are not shown in datasets searches within the organization.

- **Groups:** CKAN groups allow creating and managing collections of datasets. This is useful to catalogue datasets for a particular domain, or on a particular theme, or as a very simple way to help people find and search your own published datasets.

Since organizations and groups can create and manage collections of datasets, what’s the main difference between them?

1. **Authorization:** An organization admin (or editor) can create new datasets and edit existing datasets in the organization but a group admin (or editor) only can take existing datasets that are already on the portal and add them to the group.
2. **Privacy:** Organizations can contain private datasets that are only visible to the members of the organization, groups can’t.

As commented in the introduction, one of the objectives of the TransformingTransport (TT) ODP is to integrate all data produced from pilots in all project domains. In this context, organization and pilot are the same concept. Therefore, each pilot, present and future, will have a CKAN organization in the data portal. The pilot coordinator will manage this organization, publishing or modifying datasets.

For the time being, thirteen organizations have been created to manage the data generated by the corresponding pilots:

- Airports Smart Passenger Flow (T8.2)
- Airports Smart Turnaround, ETA Prediction and Passenger Flow (T8.3)
- Dynamic Supply Networks Shared Logistics for E-Commerce (T10.2)
- Rail Predictive Asset Management (T6.2)
- Rail Predictive High Speed Network Maintenance (T6.3)
- Ports Valencia Sea (T7.2)
- Ports Duisport Inland (T7.3)
- Integrated Urban Mobility Tampere (T9.2)
- Integrated Urban Mobility Valladolid (T9.3)
- Smart Highways Ausol (T4.2)
- Smart Highways Norte Litoral (T4.3)
- Sustainable Connected Vehicles Cars (T5.2)
- Sustainable Connected Vehicles Trucks (T5.3)
Furthermore, seven groups have been created to classify these data (one for each domain):

- **Smart Highways**: Datasets sent by AUSOL Load Balancing Pilot and NORTE LITORAL Load Balancing Pilot.
- **Sustainable Connected Vehicles**: Datasets sent by Sustainable Connected Cars Pilot and Sustainable Connected Trucks Pilot.
- **Rail**: Datasets sent by Proactive Rail Infrastructure (Thales) and Predictive High Speed Network Maintenance Pilot (INDRA).
- **Ports**: Datasets sent by Valencia Sea Port Pilot (ITI) and Duisport Inland Port Pilot (UDE).
- **Airports**: Datasets sent by Smart Passenger Flow (INDRA) and Smart Turnaround, ETA Prediction and Passenger Flow Pilot (JEPP).
- **Integrated Urban Mobility**: Datasets sent by Tampere Integrated Urban Mobility and Logistics Pilot (VTT) and Valladolid Integrated Urban Mobility and Freight Pilot (CARTIF).
- **Dynamic Supply Networks**: Datasets sent by Dynamic Supply Networks (AUEB).
3 Data Portal Guide

This part of the document is not an exhaustive manual but a brief user guide for getting started on the usage of the TT data portal.

3.1 New User

Users can be created by clicking on the Register link at the top right of the screen, fill in the fields required and press the “Create Account” button.

![Register for an Account page](image)

**Figure 4. Register for an Account page**

After registering as a simple user, you can only add datasets or view an existing one. You have to contact with Francisco Yedro ([fyedro@fi.upm.es](mailto:fyedro@fi.upm.es)) indicating your position within a TT pilot, so that we can assign your organization, group and role.
3.2 User’s Dashboard

At the top of any page, select the dashboard symbol (next to your user name) to display the dashboard. We can choose between four options; News feed, My Datasets, My Organizations and My Groups:

![Figure 5. User options](image)

### 3.2.1 News feed

“News feed” shows changes to datasets you follow and any modification or new datasets in organizations that you follow. The number next to the dashboard icon indicates the number of new notifications since last time you entered the News feed zone. It is also possible to follow other users.

![Figure 6. Dashboard – News Feed](image)

### 3.2.2 My Datasets

“My Datasets” shows a list of all datasets created by the user (with name, description and formats). From this section, it is also possible to add new datasets as it is shown in the following screenshot:
3.2.3  My Organizations

“My Organizations” shows the list of organizations which the user belongs to. If the user has the necessary rights, he can also create new organizations.
3.2.4 My Groups

“My Groups” shows the list of groups which the user belongs to. If the user has the necessary rights, he can also create new groups.

Figure 9. Dashboard – My Groups

In each of these four options, users have the possibility to change their settings by clicking on “Edit Settings” button.
3.3 Data Portal Menu Bar

Below the previous bar we have the menu bar, where we can choose between four options: Datasets, Organizations, Groups and About:

![Menu Bar](image)

**Figure 10. Menu Bar**

3.3.1 Datasets

“Datasets” shows a complete list of all datasets visible by the user:

![Datasets Page](image)

**Figure 11. Datasets page**
In the column on the left side, users can find the datasets filtered by Organizations, Groups, Tags, Formats or Licenses and, on the right hand, a list of the visible datasets (name, description and formats) is shown.

3.3.2 Organizations

A list with all the organizations of the project is shown. Users can click on each one and see information, list of datasets or activities of the chosen organization:

![Organizations page](image)

**Figure 12. Organizations page**
3.3.3 Groups

A list with all the groups of the project is shown. User can click on each one and see information, list of datasets or activity of the chosen group:

![Groups page](image-url)

**Figure 13. Groups page**
3.3.4 About

“About” shows a brief purpose of the portal and description of the project.

Figure 14. About page

3.4 Creating a Dataset

In this part, we describe the four steps to be followed for the creation of new datasets:

**Step 1:**

Users can create a dataset in two different manners:

1- Select the “Datasets” link at the top of any page. From this, above the search box, select the “Add Dataset” button.

2- Alternatively, select the “organizations” link at the top of a page. Now select the page for the organization that should own your new dataset. Provided that you are a member of this organization, you can now select the “Add Dataset” button above the search box.

**Step 2:**

CKAN will ask the user for information about the data. This information is based on “Data Assets and API ID card” developed in Task 2.3:
NOTE: By default, the only required field on this page is the title. However, it is good practice to include, at the minimum, a short description and, if possible, the license information. You should ensure that you choose the correct organization for the dataset, since at present, this cannot be changed later. You can edit or add the other fields later.

**Step 3:**

When you have filled in the information on this page, select “Next: Add Data” button. (Alternatively select “Cancel” to discard the information filled in)
Step 4:

CKAN will display the “Add data” screen:

![Add data screen](image)

**Figure 16. Create a dataset – add data**

This is where you will add one or more “resources” which contain the data for this dataset. Choose a file or link for your data resource and select the appropriate choice at the top of the screen:

- If you are giving CKAN a link to the data, like http://example.com/mydata.csv, then select “Link to a file” or “Link to an API”.
- If the data to be added to CKAN is in a file on your computer, select “Upload a file”. CKAN will provide you a file browser to select it.

In the same way that a dataset is created, a dataset can be edited by selecting the dataset and clicking on the Manage button (only visible if the user has permission).

![Editing dataset](image)

**Figure 17. Editing dataset**
3.5 Finding Data

There are two ways of finding data into the ODP:

- **Searching the site**: To find datasets in CKAN, type any combination of search words (e.g., “pilot”, “transport”, etc.) in the search box on any page. CKAN displays the first page of results for your search. You can:
  - View more pages of results
  - Repeat the search, altering some terms
  - Restrict the search to datasets with particular tags, data formats, etc. using the filters in the left-hand column

  If there are a large number of results, the filters can be very helpful, since you can combine filters, selectively adding and removing them, and modify and repeat the search with existing filters still in place.

  ![Search dataset](image)

  **Figure 18. Search dataset**

- **Searching within an organization**: If you want to look for data owned by a particular organization, you can search within that organization from its home page in CKAN.
  - Select the “Organizations” link at the top of any page.
  - Select the organization you are interested in. CKAN will display your organization’s home page.
  - Type your search query in the main search box on the page.

  CKAN will return search results as normal, but restricted to datasets from the organization.

  If the organization is of interest, you can opt to be notified of changes to it (such as new datasets and modifications to datasets) by using the “Follow” button on the organization page.
Conclusions and Next Steps

In this deliverable, the OPD for the TT project has been presented. It provides an overall view of what an OPD is and why an OPD is important and finally, a brief user’s guide is provided.

The objective of the TT ODP is to provide the community working on transport data across the different transport domains identified for TT (Smart Highways, Sustainable Connected Vehicles, Rail, Ports, Airports, Integrated Urban Mobility and Dynamic Supply Networks) with open datasets that they can reuse for their own purposes, as well as links and metadata to existing datasets that cannot be published under an open data license but where ad-hoc agreements may be established between the data producers (identified as part of such metadata) and the potential data reusers.

As of the dissemination of this document, users can start using the portal and we will delighted if we receive some feedback by way of new features or problems detected.

Just as the DMP, this is a living document and, as time goes by, new features will appear and some existing ones may not be relevant any more, what will require this document to be updated.

The next steps will be focused on polishing several aspects of the data portal, such as filling in all the metadata and data of the datasets that have been received and, afterwards, a community of Zenodo⁵ will be created for the archival and preservation of the metadata and datasets (as said before, only for those datasets that have been identified as open data). Zenodo is a research data open access repository created by OpenAIRE⁶ and CERN⁷ where researchers can deposit datasets, and where such datasets become citable.

Finally, Pilots are expected to create their users for managing everything about their datasets, groups, etc. so that they will be sent an email with the steps to follow (see page 13). This information will be added into the ODP “Register for an account” page.

---

⁵ Zenodo repository - https://zenodo.org/
⁶ OpenAIRE project - https://www.openaire.eu/
⁷ Cern project - https://home.cern/