



Project co-financed by the European
Regional Development Fund



Project Acronym: ODEON

Project title: Open Data for European Open iNnovation

Grant Agreement number: 3MED17_1.1_M2_061

D.3.2.1

Capitalisation of running Open Data platforms within ODEON platform

WP n°:	3
Task n°:	3.2 Capitalization of EU and past experiences to increase the availability of OD/LOD
Author(s):	GFOSS
Contributors:	Veneto Region, Sarga
Type:	<R = Report, Document>;
Dissemination level:	<PU = public>
Revision:	Final Version
Due Date:	Here the deadline indicated by the Application Form (C.1.5)
Date of submission:	Here the effective date of finalisation (with the quality approval)

ODEON – D.3.2.1

Deliverable History

This deliverable history should be removed from the document once it has been finalized. It can then be stored as a separate document on the server, next to the final version.

Version	Date	Status	What's new?
0.1	11/09/2018	draft	0.1_DRAFT v20180911 Methodology, ToCs, Template and request for contributions
0.2	05/08/2019	Draft	0.2_Version 1
0.3	26/08/2019	draft	0.3 _Version including Veneto Region's comments

Summary

This preliminary study is aiming at the fine-tuning of instruments, technologies and projects already available at MED level by consolidating EU level platforms, providing the necessary input for the development of ODEON platform.

It consists of the analysis and identification of possible synergies of 4 platforms (www.europeandataportal.eu, www.homer.eu, www.linda-project.eu and www.yourdatastories.eu). The availability and quality of open data will be processed and validated in order to be capitalized in the project.

Contents

1.	Introduction	6
1.1	Categories of Datasets of ODEON platform	6
2.	Analysis of OD/LOD available in MED area from selected platforms	8
2.1	OD/LOD in European Data Portal	8
2.1.1	Blue Growth Data available in European Union Open Data Portal.....	8
2.1.1.1	Quantity of Data	8
2.1.1.2	Quality of Data	8
2.1.2	Green Growth Data available in European Union Open Data Portal	9
2.1.2.1	Quantity of Data	9
2.1.2.2	Quality of Data.....	9
2.1.3	Creative and Cultural Industries Data available in European Union Open Data Portal	10
2.1.3.1	Quantity of Data	10
2.1.3.2	Quality of Data.....	10
2.2	YDS Platform – Linked Open Data	10
2.2.1	Quantity of Data	10
2.2.2	Quality of Data.....	11
2.3	Homer Federation Open Data	11
2.5.1	Blue Growth Data available in Homer Federation Open Data.....	11
2.5.1.1	Quantity of Data.....	11
2.5.1.2	Quality of Data.....	11
2.5.2	Green Growth Data available in Homer Federation Open Data	11
2.5.2.1	Quantity of Data.....	11
2.5.2.2	Quality of Data	11
2.5.3	Creative and Cultural Industries Data available in Homer Federation Open Data.....	12
2.5.3.1	Quantity of Data.....	12
2.5.3.2	Quality of Data	12
3.	Description of selected tools and platforms features - benchmarking.....	13
3.1	Your Data Stories Platform.....	13
3.1.1	Description of Your Data Stories Platform.....	13
3.1.2	Available tools in Your Data Stories Platform	13
3.2	Homer Federation.....	13
3.2.1	Description of Homer Federation	13

3.2.2 Available tools in Homer Federation	14
3.3 LinDA Workbench.....	14
3.3.1. Description of LinDA Workbench	14
3.3.2. Available tools in LinDA Workbench	15
4. Scoping of ODEON platform	16
4.1 ODEON platform Features	16
4.2 Users of the Platform	16
5. Findings/Identification of tools and technologies that can be utilized within ODEON platform...	17
5.1 Tools and Technologies for the ODEON Platform.....	17
1) Harvester tool importing Datasets from Open Data Platforms	17
2) Validity check tools.....	17
3) Metadata aggregator	17
3.1) Enrichment of Metadata	17
4) Linked Data tool.....	17
5) Visualization and Presentation of Data.....	18
6) Odeon Platform API.....	18
7) MarketPlace	19
8) User Forum.....	19
6. Conclusions – Technical Specifications for the ODEON Platform	20

1. Introduction

The four platforms that have been selected to be analysed within this study were implemented in different periods and were addressing different needs.

[HOMER Federation](#) which has been the result of the [HOMER project](#) was created to set up the basis for a transnational digital market in the Mediterranean area, setting up MED PSI Federation, thus promoting interoperable and multilingual solutions and the development of a Mediterranean community of stakeholders.

[LinDA Workbench](#) is a complete open-source package of Enterprise Linked Data tools to quickly map and publish data in the Linked Data Format, interlink them with other public or private data, analyse them and create visualizations.

Your [Data Stories \(YDS\) Platform](#) is a recent platform that aims to Explore & Analyse open data. YDS offers easy access to high volumes of data instantly with smart search. YDS contextualizes data enabling meaningful comparisons and transforms findings into attractive easy to read visualisations.

[The European Data Portal](#) is an initiative of the European Commission, and is part of the Digital Single Market. As a platform harvests the metadata of Public Sector Information available on public data portals across European countries. The ODEON platform will harvest datasets from European Data portal and will combine them in the three Categories of the ODEON platform

1.1 Categories of Datasets of ODEON platform

The Odeon Platform will provide quality open datasets in three sectors

- Blue Growth
- Green Growth and Creative
- Creative and Cultural Industries

BLUE GROWTH DATASETS

According to European Commission [Blue economy report](#) 2019 there are 6 main sectors of Blue Growth and Economy

- Extraction and commercialisation of marine living resources
- Marine extraction of minerals, oil and gas
- Maritime transport and Ports
- warehousing and construction of water projects
- Shipbuilding and repair
- Coastal tourism

The ODEON Platform will harvest and will provide datasets for the Blue Growth section

based in these 6 distinctive categories.

GREEN GROWTH DATASETS

Taking into consideration the United Nations publication “A guidebook to the Green Economy- Issue 1: Green Economy, Green Growth, and Low-Carbon Development – history, definitions and a guide to recent publications”

(<https://sustainabledevelopment.un.org/content/documents/GE%20Guidebook.pdf>) we identified 4 distinctive categories for the Green Growth section of the ODEON Platform:

- Climate
- Energy Production
- Energy Consumption
- nature conservation

The ODEON Platform will harvest and will provide datasets for the Green Growth section based in these 4 distinctive categories.

CREATIVE AND CULTURE INDUSTRIES

According to Unesco publication : [Cultural times. The first global map of cultural and creative industries](#) there are 11 sectors of the Creative and Culture Industries

- Advertising
- Music
- Radio
- Architecture
- Movie
- TV
- Books
- Newspapers magazines
- Visual arts Design
- Gaming
- Performing Arts

The ODEON Platform will harvest and will provide datasets for the Creative and Culture Industries section based in these 11 distinctive categories.

2. Analysis of OD/LOD available in MED area from selected platforms

2.1 OD/LOD in European Data Portal

The European Open Data Portal (<https://www.europeandataportal.eu/>) is one of the biggest sources of Open Data and Linked Data. European Data Portal provides access to more than 892,000¹ public datasets from 35 countries (European Union member states, the EEA, Switzerland and countries in the EU Neighbourhood Policy programme). The Data resources are indexed by the European Commission from national, regional, local and domain-specific public data providers. The interface is available in 24 languages (but through machine translation) but most metadata are currently available in a limited number of languages (English, French and German). Some of the metadata (e.g. names of the data providers and geographical coverage) are in 24 languages. and all the data are free to use and reuse for commercial or non-commercial purposes.

2.1.1 Blue Growth Data available in European Union Open Data Portal

2.1.1.2 Quantity of Data

After a preliminary analysis in the European Data Portal we found that there are more than 99000 available datasets related to Blue Growth.

We used as reference the six established sectors of Blue Economy according to the 2019 Blue Economy Report of European Commission² and were identified more than 99000 related datasets

2.1.1.3 Quality of Data

From the 99334 available datasets the 1/2 of the datasets are reusable in Odeon Platform since the 34% of the datasets are of unknown format and around 15% in a non-machine-readable format

1292 datasets are available in .csv format

7481 datasets in WMS format ³

3458 datasets in WFS format⁴

1 <https://www.europeandataportal.eu/data/#/datasets?locale=en>

2 European Commission [Blue economy report](#) 2019 (The 6 sectors: Extraction and commercialisation of marine living resources 1)Marine extraction of minerals, oil and gas 2) Maritime transport 3) Ports,4)warehousing and construction of water projects 5)Shipbuilding and repair 6) Coastal tourism)

3 WMS, short for *web map service*, is a popular way of publishing maps by professional GIS software . This format is similar to map tiles, but more generic and not so well optimized for use in web maps. A WMS image is defined by the coordinates of its corners)

4522 datasets in HTML format
485 datasets in JSON
284 datasets in xls format
444 datasets in xlsx format
865 datasets in PDF format
285 datasets in xml
312 datasets in zip format
157 datasets in Shp format
the rest are in various format

2.1.2 Green Growth Data available in European Union Open Data Portal

2.1.2.1 Quantity of Data

After a preliminary analysis in the European Data Portal we found that there are more than 137000 available datasets related to Green Growth. We used as reference the main keywords that are related with of Green Growth and Economy⁵ and were identified more than 137000 related datasets.

2.1.2.2 Quality of Data

From the 13700 available datasets the 1/2 of the datasets are reusable in Odeon Platform since the 35% of the datasets are of unknown format and around 15% in a non-machine-readable format.

5325 datasets are available in .csv format
8463 datasets in WMS format
5844 datasets in WFS format
4300 datasets in HTML format
2485 datasets in JSON
2483 datasets in xls format
1379 datasets in xlsx format
1531 datasets in PDF format
1236 datasets in xml format
1839 datasets in zip format

4 **Web Feature Service (WFS)** Interface Standard provides an interface allowing requests for geographical features across the web using platform-independent calls. One can think of geographical features as the "source code" behind a map, whereas the WMS interface or online tiled mapping portals like Google Maps return only an image, which end-users cannot edit or spatially analyze. The XML-based GML furnishes the default payload-encoding for transporting geographic features, but other formats like shapefiles can also serve for transport.)

5 United Nations "A guidebook to the Green Economy- Issue 1: Green Economy, Green Growth, and Low-Carbon Development – history, definitions and a guide to recent publications"
<https://sustainabledevelopment.un.org/content/documents/GE%20Guidebook.pdf>

465 datasets in Shp format
the rest are in various machine-readable formats

2.1.3 Creative and Cultural Industries Data available in European Union Open Data Portal

2.1.3.1 Quantity of Data

After a preliminary analysis in the European Data Portal we found that there are more than 95000 available datasets related to Creative and Cultural Industries.

We used as reference the 11 sectors of CCI as defined by UNESCO⁶ definition and were identified more than 137000 related datasets that belong in the Creative and Cultural Industries section of the ODEON Platform.

2.3.2 Quality of Data

From the 95786 available datasets the 1/2 of the datasets are reusable in Odeon Platform since around 35% of the datasets are of unknown format and around 15% are in a non-machine-readable format

7050 datasets are available in csv format
19080 datasets in WMS format
5204 datasets in WFS format
6975 datasets in HTML format
2005 datasets in JSON
2483 datasets in xls format
1187 datasets in xlsx format
2007 datasets in PDF format
1236 datasets in xml format
1048 datasets in zip format
1599 datasets in Shp format
the rest are in various machine-readable formats

2.2 YDS Platform – Linked Open Data

2.2.1 Quantity of Data

YDS platform collects and visualises financial open data from public projects, contracts, trade activities, and aid activities from Europe. YDS offers easy access to high volumes of data instantly with smart search. YDS contextualizes data enabling meaningful comparisons and transforms findings into attractive easy to read visualisations. Since YDS focuses on continuous data streams, it

automatically updates and integrates new information directly from the sources as it becomes available.

Currently there is an availability of data for more than 12000 public projects, around 6 million trade activities, more than 19000 aid activities and around 5,9 million contracts.

Through the use of YDS tools and APIs, ODEON platform can use the relevant financial data to Blue growth, Green Growth and Culture and Creative Industries sectors.

2.2.2 Quality of Data

Relevant data from YDS is machine readable and can be integrated to ODEON Platform

2.3 Homer Federation Open Data

2.5.1 Blue Growth Data available in Homer Federation Open Data

2.5.1.2 Quantity of Data

After a preliminary analysis in the Homer Federation we found that there are around 37 available datasets related to Blue Growth.

We used as reference the six established sectors of Blue Economy according to the 2019 Blue Economy Report of European Commission.

2.5.1.3 Quality of Data

All of the 37 available datasets are reusable in Odeon Platform, since they are in machine readable open format (json- xml). Before integrating them in the ODEON platform there is the need to check if there are any updated datasets in the primary source of the datasets.

2.5.2 Green Growth Data available in Homer Federation Open Data

2.5.2.1 Quantity of Data

After a preliminary analysis in the Homer Federation we found that there are 60 available datasets related to Green Growth.

We used as reference the main keywords that are related with Green Growth and Economy.

2.5.2.2 Quality of Data

All of the 60 available datasets are reusable in Odeon Platform, since they are in machine readable open format (json- xml). Before integrating them in the ODEON platform there is the need to check if there are any updated datasets in the primary source of the datasets.

2.5.3 Creative and Cultural Industries Data available in Homer Federation Open Data

2.5.3.1 Quantity of Data

After a preliminary analysis in the Homer Federation Open Data we found that are 22 available datasets related to Creative and Cultural Industries

2.5.3.2 Quality of Data

All of the 22 available datasets are reusable in Odeon Platform, since they are in machine readable open format. Before integrating them in the ODEON platform there is the need to check if there are any updated datasets in the primary source of the datasets.

3. Description of selected tools and platforms features - benchmarking

3.1 Your Data Stories Platform

3.1.1 Description of Your Data Stories Platform

Your Data Stories Platform, YDS in short, is a platform that helps make sense of open and social data. It provides data through specific queries by turning data into the substrate for stories that human beings can comprehend and act upon. YDS addresses professionals in government, public administration, business, and journalism, but it is also made for citizens.

It offers easy access to high volumes of data instantly with smart search. YDS contextualizes data enabling meaningful comparisons and transforms findings into attractive easy to read visualizations.

3.1.2 Available tools in Your Data Stories Platform

Your Data Stories Platform offers a number of opens source tools and APIs:

- YDS offers an API built based on the JSON API specification as well as a JSON-LD API.
- All tools used in the YDS platform are opensource and can be found under <https://github.com/YourDataStories>
- Visualization Components, a collection of Angular based components to visualize open data
- YourDataStories Ontology, the Linked Open Economy (LOE) ontology which is developed to describe financial open data
- Harvesters, a set of harvesting tools to acquire data from open data sources and the social Web implemented in Java
- YDS mobile application, the android app for viewing, rating and commenting on public projects.
- YDS Twitter Service, a java application that posts tweets using data from the YDS API.

3.2 Homer Federation

3.2.1 Description of Homer Federation

HOMER was a strategic MED project that focused on the theme of Open Data, a world-wide policy aiming at making available and exploitable Public Sector Information (PSI). One of the objectives of the project was to set up the basis for a transnational digital market in the Mediterranean area, setting up MED PSI Federation, thus promoting interoperable and multilingual solutions and the development of a Mediterranean community of stakeholders.

Federation means the virtual system composed by a software able to create a common catalogue and an index supported by a search engine that using the translator component, allows multi language search of metadata. The metadata catalogue is built using the metadata cards exposed by all Homer Open Data Portals' Partners.

Homer Federation is conceived as potentially open to future accession by other Partners -free to conduct their own Open Data initiative independently of the HOMER project- with the opportunity to aggregate their respective and complementary skills in order to maximize advantages and effects of a common federated index.

3.2.2 Available tools in Homer Federation

The main tools in the implementation of Homer Federation was the "Federated Semantic Search Engine".

The Federated Semantic Search Engine is an indexing and searching engine component.

The components behind the Federated Semantic Search Engine were

- **"Fed-Index Homer"** - a federated index file component - containing the complete list of metadata which describes all the data related to the 5 selected categories, and contained in all the Open Data Portals connected to the Federation; It ensures the uniqueness of the id cards that are brought by the portals participating at the federation;
- **"Fed-Translator"** - is the translator component - customized for search query, that allows combining all necessary vocabulary and thesaurus families using tags translated in the metadata card. Every tag in metadata card is generated and translated using the translator component, and all terms in the thesaurus in all languages are considered as labels of the same concept;
- **"Fed-Searcher"** – is the centralized semantic search engine component;
- **"Fed-Loader API"**- is the loader that catches the APIs or web services provided by each Open Data Portal Partners' that will be used by the federated engine to extract via harvesting the common metadata set involved into federation, and collect them into the federated Index;

3.3 LinDA Workbench

3.3.1. Description of LinDA Workbench

LinDA Workbench is a complete open-source package of Enterprise Linked Data tools to quickly map and publish your data in the Linked Data Format, interlink them with other public or private data, analyze them and create visualizations.

3.3.2. Available tools in LinDA Workbench

Linda Workbench consists of the following tools:

- **LinDA Transformation engine**, a lightweight transformation to linked data tool (i.e converts csv to rdf linked data)
- **LinDA Vocabulary repository**, a vocabulary for increasing the semantic interoperability for the data. The transition from traditional relational databases to semantically rich data representations, and RDF in particular, is to track down the appropriate vocabularies from which terms will be used to represent the various entities, subjects and operations of an organization.
- **LinDA RDF2Any**, a tool for converting RDF to conventional data structures in order to be used by legacy applications,
- **LinDA Query Builder and Query designer for** easier navigation and query the data
- **LinDA visualization** to perform smart visualizations on linked data out-of-the box
- **LinDA Analytics** package for running analytic processes against data . A library data analytic functionality is provided enabling SMEs to utilise and share analytic methods on Linked Data for the discovery and communication of meaningful new patterns that were unattainable or hidden in the previous isolated data structures.

4. Scoping of ODEON platform

4.1 ODEON platform Features

The main purpose of building the new ODEON platform Odeon will be to obtain the following:

- A new model of data offering through API (Application program interface) to facilitate developers and users to interact with the data (DATA AS A SERVICE). This will provide a real framework of public data, ready for application development, and described with high-value metadata.
- The platform must ensure a monitoring system that will be used to understand which and how many applications will leverage the data and then will offer the decision-maker an analysis tool that enables them to understand the value of data in order to make decisions.
- The platform must be designed and engineered at multiple levels:
 - Shared data catalog and common metadata profile.
 - API Gateway
 - Developers community
 - Market place APP and API

4.2 Users of the Platform

The ODEON platform is targeted at individuals, SMEs, public and private organizations and various groups of people from different scientific fields and entrepreneurial areas, aiming at exploiting open data according to their needs and intentions.

In addition to data publishers who upload, update and manage datasets, the platform will be targeted at researchers, business people, students, national or international organizations, start-ups, government agencies, websites, applications, and anyone interested in making use of open data sets for the development of added-value products and services

5. Findings/Identification of tools and technologies that can be utilized within ODEON platform

5.1 Tools and Technologies for the ODEON Platform

The scope of this chapter is to identify the tools and technologies that will be utilised and integrated within the ODEON Platform

1) Harvester tool importing Datasets from Open Data Platforms

The ODEON platform harvester that will rely on [ckanext-harvest extension](#), which allows a host CKAN instance to collect and import datasets and/or metadata from other CKAN catalogues and on YourDataStories harvesters <https://github.com/YourDataStories/harvesters>

2) Validity check tools

Data Quality varies widely across different portals, reflecting the publication processes and requirements of the hosting organizations. In general, it is difficult for users to assess the quality of the data and there is a lack of descriptors for the actual data fields. At the publisher level, while strong emphasis has been put in metadata standards and interoperability, publishers don't generally have the same help or guidance when dealing with data quality or description.

Data quality in Open Data portals can have a central place on both these fronts, user-centric and publisher-centric. ODEON platform will use technologies like [ckanext-validation](#) (a CKAN extension that provides a low level API and readily available features for data validation and reporting) and validation mechanisms from YourData Stories harvesters <https://github.com/YourDataStories/harvesters>

3) Metadata aggregator

Based on **Fed-Index Homer** the tool will contain the complete list of metadata with descriptions describes all the data related to the 3 Odeon Main categories, and thematic units such as environment, economy, agriculture, fisheries, culture, transport, energy etc.

3.1) Enrichment of Metadata

with the help of the European dictionary used by Homer Federation: <http://eurovoc.europa.eu/>, and the Metadata vocabulary of European Data Portal

4) Linked Data tool

In order to produce Linked Open Data (LOD) the ODEON platform will use a combination of tools such as:

- Linda Transformation Engine - RDF Conversion Tool
- YDS stories Linked Data Tools

5) Visualization and Presentation of Data

The Odeon Platform must offer all the datasets offering through API (Application program interface) to facilitate developers and users to interact with the data (DATA AS A SERVICE). This will provide a real framework of public data, ready for application development, and described with high-value metadata.

The tools that will be used for the visualization and presentation of Datasets will be a combination of:

- **Your Data Stories Visualization Components**
(<https://github.com/YourDataStories/components-visualisation>)

Your Data Stories Visualization Components is a collection of visualization components based on AngularJS and other third-party JS libraries that can provide to ODEON platform:

- Statistics
 - Project Info
 - Search
 - Results
 - Facets
 - Browse
 - Grid
 - Line Chart
 - Bar Chart
 - Pie Chart
 - Map
 - Heatmap
 - Geo Editing Tool
- **LinDA visualization tools** (<https://github.com/LinDA-tools/LindaWorkbench>)

For visualizations on linked data and the LinDA Analytics package for running analytic processes of the data.

6) Odeon Platform API

The term API is an acronym, and it stands for “Application Programming Interface.” In [computer programming](#), an **application programming interface (API)** is a set of subroutine definitions, [communication protocols](#), and tools for building software. In the Case of the ODEON platform the API will be a web based RESTful API since REST technology

is generally preferred to the more robust Simple Object Access Protocol ([SOAP](#)) technology because REST leverages less [bandwidth](#), making it more suitable for internet usage.

Odeon Platform will provide a REST API based on CKAN in order to be accessible from other portals, developers, SMEs and users. Under API there will be available all of the portal core functionalities.

7) MarketPlace

A place that all the applications and APIs that were built around the ODEON platform. YDS marketplace and Homer is an example.

8) User Forum

A forum for the community of the developers that are interested in the platform

6. Conclusions – Technical Specifications for the ODEON Platform

The ODEON platform is expected to host services and tools for publishing, developing, accessing, using and visualizing Open Data sets from Private and Public Organizations as well as well-known open data portals such as HOMER, YourDataStories, EU Open Data Portal, Linda and others. The link to the European Open Data Platforms and the use of previous projects with relevant quality protocols is expected to ensure the continuous flow of data to appropriate standards (JSON, XML, CSV, XLS), their categorization, indexing and access users of different scientific fields and subject areas. The purpose of the ODEON platform is to provide users with an easy-to-access environment and is expected to include tools that will ensure the quality and validity of the data.

By reusing and combining data, tools and services from three major Open Data Projects (Homer, Linda, Your Data Stories) the ODEON platform will provide a new framework of public data, ready for application development, and described with high-value metadata.

The suggested Tools and Technologies for the ODEON Platform as they are described in Chapter 5 of this report, will provide all the necessary infrastructure for achieving the goals of the project.

The development of the ODEON platform is closely linked with the local Data Innovation Hubs. Communication and marketing actions in the territories as well as experimentation and training on the platform in the local data hubs is essential for the development and the improvement of the ODEON platform.