



## Raising Open and User-friendly Transparency- Enabling Technologies for Public Administrations



Project number 645860  
H2020-INSO-2014

### D4.3 Beta version of SPOD

(Final, version 1.0, 1/31/2017)



WISE & MUNRO



**Document produced by**

Organisation: Università di Salerno

Author / email: Vittorio Scarano, vitsca@dia.unisa.it

Subject: Beta version of SPOD

Due date: 31 January 2017

Main contributions by:

- **Ancitel** team: John Forrester, Giuseppe Clementino, Emanuele Barucca, Paolo Molinari, Leonardo Rinaldi;
- **Ortelio** team: Ilias Trochidis, Vangelis Banos;
- **Warsaw School of Economics** team: Przemyslaw Szufel, Michał Wasiluk;
- **UNISA team**: Delfina Malandrino, Gennaro Cordasco, Giuseppina Palmieri, Carmine Spagnuolo, Luca Vicidomini, Donato Pirozzi, Renato De Donato, Luigi Serra, Andrea Petta

Dissemination level: Public [PU]

Reviewed and approved by

Date	Name	Organization
20-01-2017	Elena Palmisano	Prato Pilot
26-01-2017	Adegboyega Ojo	Insight Centre for Data Analytics National University of Ireland, Galway (NUIG)

**Revision History**

Version	Date	Authors	Status	Description of Changes
0.1	17-10-2016	G.Palmieri	Draft	A first organization
0.2	3-01-2017	V.Scarano	Draft	Organisation of the material
0.3	5-1-2017	A.Petta, R. De Donato, L. Vicidomini, G. Palmieri	Draft	Added material on roles, controllet, interactions with TET, releases table, feedback table.
0.4	6-1-2017	V.Scarano	Draft	Reorganization of the agile methodology cycle section
0.5	6-1-2017	L. Serra, V.Scarano	Draft	Included the cocreation part that was left behind in 0.3
0.6	7-1-2017	G.Palmieri	Draft	Added completed table of all Feedback, Event section
0.7	8-01-2017	A.Petta, L.Serra, R. De Donato, G. Palmieri.	Draft	Added material on (5.1) Base Social Platform: Why Oxwall and not Elgg; (6.1): the organization of cocreation (7.5.8.1): Decision Trees; User Interface and advance support; Reuse rather than reinvent: the collaboration platforms etherdocs and ethersheets; section of the Widgets in What's new

				page and table to explain feedback template. Add a new section for “Extended Search”
0.8	12-01-2017	A.Petta, L.Serra, R. De Donato, G. Palmieri; Ancitel Team	Draft	Added material on (6.5.8.4) Decision Tree Integration in SPOD; (6.5.12) Manager View, Email Notification; content in Fig.39 and Fig.40; 6. System Design of SPOD 10 Relevant publication; RRI-ICT section; Adding Maps from geo-server (Architecture and Technology)
0.9	13-01-2017	D. Malandrino, Balachander Krishnamurthy	Draft	Added material on Privacy and Appendix
0.10	17-01-2017	G.Palmieri, Ortelio team, V.Scarano	Draft	Add material on (6.5.11) Extend Search; From Alpha to Beta: a roadmap of major changes; updated Relevant publications, added table of feedback for map (6.2); Executive Summary
0.11	18-01-2017	G.Palmieri	Draft	Cosmetics
0.12	20-01-2017	E.Palmisano, G.Palmieri	Draft	Integration of internal review; added tables 2 for highlight “Main criteria to choose Oxwall instead of Elggs”; added new section “Planning for third year of SPOD”: new section 2.2.6 “Exploitation and implementation: container technology (All-in-one machine)
0.13	26-01-2017	Adegboyega Ojo, G.Palmieri, Luca Vicidomini, Renato De Donato	Draft	Integration of internal review; integration of Section 2.2.6 about “All-in-One” and Sec. 6.4.15: Wordpress plug-in to embed datalets in Wordpress pages or posts; A Table for the review of state of art for the design of the Controllet
1.0	31-01-2017	V.Scarano	Final	Release

# TABLE OF CONTENTS

1	Introduction.....	8
1.1	Addressing Consolidated Expert Review Report issues.....	8
1.2	From Alpha to Beta: a roadmap of the major changes .....	9
2	The Agile Development Process .....	12
2.1	The Overall Cycle .....	12
2.1.1	Release notes.....	13
2.1.2	Feedback received From Pilots .....	16
2.2	The Development process: Design, Implementation and Release .....	17
2.2.1	Design: Issue Tracking.....	18
2.2.2	Design/Implementation: Monitoring the development .....	19
2.2.3	Source Code repositories and metrics .....	22
2.2.4	Managing the Releases .....	23
2.2.5	Technical Documentation: the Wiki page .....	24
2.2.6	Exploitation and implementation: container technology (All-in-one machine) .....	25
3	Requirements .....	27
3.1	Functional Requirements .....	27
3.2	Non-functional Requirements.....	28
3.2.1	Performance .....	28
3.2.2	Privacy .....	28
4	Design principles and technology.....	28
4.1	Base Social Platform: why Oxwall and not Elgg.....	28
4.2	Reuse rather than reinvent: the collaboration platforms EtherDocs and EtherSheets .....	30
4.2.1	Overview.....	30
4.2.2	Possible existing solutions .....	31
4.2.3	SPOD collaboration platform .....	32
5	Architecture Design of SPOD .....	37
5.1	The organization of cocreation servers .....	37
5.2	Notification components.....	38
6	System Design of SPOD.....	39
6.1	Collaborative space .....	39
6.1.1	Knowledge room.....	41
6.1.2	Data Cocreation room.....	44
6.1.3	Blogs .....	48
6.2	Adding Maps from geo-servers .....	49
6.2.1	Architecture .....	50
6.2.2	Technology.....	50
6.2.3	Maps Integration in SPOD.....	50
6.3	Events.....	51
6.4	Other features .....	53
6.4.1	Attach a document/image .....	53
6.4.2	Changing the Welcome Page (Open Wall) .....	54
6.4.3	Delete a Public Room.....	55
6.4.4	Recommended datasets .....	55
6.4.5	Roles .....	58
6.4.6	Customizing the "What's new" page with widgets .....	58
6.4.7	Decision trees .....	59
6.4.8	Interactions with TET .....	62
6.4.9	Joint authentication .....	66
6.4.10	Extended search .....	67
6.4.11	Manager view .....	69
6.4.12	Email notification strategy .....	72
6.4.13	Agora settings .....	74
6.4.14	Privacy plugin.....	76
6.4.15	Datalets in Wordpress plug-in.....	77



7	User Interface and advanced support .....	81
7.1	Data visualization: a review of state of the art.....	82
7.1.1	Methodology for choosing the right graph.....	82
7.1.2	Taxonomy of interactive dynamics .....	83
7.1.3	Asynchronous collaborative Visualization systems.....	84
7.1.4	Other Data Visualization Tools.....	88
7.1.5	Which chart or graph is right for you? .....	93
7.1.6	Best Practices with data visualizations .....	94
7.1.7	Select the right graph for your message .....	96
7.2	A new design of user interface for the Controllet .....	97
7.3	The datalet wizard Controllet.....	104
7.3.1	Step 1: select dataset.....	106
7.3.2	Step 2: select data .....	108
7.3.3	Step 3: select visualization.....	110
8	Cross-cutting Design Concerns .....	111
8.1	Privacy .....	111
8.1.1	Privacy in the Agora .....	112
8.1.2	Privacy in the Data Cocreation.....	112
8.2	RRI-ICT .....	113
9	Planning for the third year of SPOD .....	113
	Relevant publications .....	116
10	Conclusions.....	117
11	Bibliography.....	118
12	Appendix - A: Feedback from the Privacy Advisor Dr. Balachander Krishnamurthy.....	119
12.1	Collecting personal and sensitive information .....	119
12.2	Awareness about discussions of sensitive nature .....	119
12.3	Authentication .....	120
13	Appendix – B: Metrics .....	121
13.1	SPOD Beta metrics .....	121
14	Appendix – C: Non-Functional Requirements from D2.1 .....	122

## Executive Summary

In this deliverable, the design and implementation of the Social Platform for Open Data Beta version 2.0 are described and carried out during the second year of the ROUTE-TO-PA project in accordance with all partners' feedback and team project meetings, within "Technological Development and Integration" Workpackage (WP4). The task in this year is T4.1 "Social Platform for Open Data" (M6-24). The work done in this task encompasses mainly the integration of new tools developed (i.e., Cocreation knowledge and data rooms, Events, Blogs), the enhancement of the User Interface for the data visualization (the Controllet), new features developed in collaboration with some team partners (Insight, Ancitel, Ortelio and WSE).

The web site of SPOD platform Beta version (v.2.0) with all features implemented and realized in these two years of the ROUTE-TO-PA project is reachable at the following address: <http://spod.routetopa.eu>.

Accordingly, with the same structure of the Deliverable 4.1, available as public deliverable on the ROUTE-TO-PA platform at the link: <http://routetopa.eu/public-deliverables/>, this document describes work carried out during the second year of the ROUTE-TO-PA project as follows:

- Section 2: The support of the process of communication with the pilots
  - The Agile Development Process" describes the planning of the pilot partners' received reports (i.e., Den Haag, Prato, Dublin, Groningen and Issy) taken into account for the design and implementation of main features and a table (see Table 2) summarizes all Pilots issues and enhancements to plan during the development of the SPOD platform.
- Sections 4, 5, 6: During the activity of the second year, team project meetings (Den Haag meeting in February and Warsaw meeting in May 2016 ) discussed mainly the Cocreation room (detailed in Section 6.1):
  - Cocreation Knowledge room (Sec. 6.1.1)
  - Cocreation Data room (Sec 6.1.2)

These sections identify a few novel features of the SPOD Beta, based on the feedback from users, the researcher's meetings, the Description of Work, the Annex 2 of the Grant Agreement (DoW) and the functional and non-functional requirements specified in this document (Section 3) and in the Deliverable D4.1 "Alpha version Of SPOD" (Section 6).

List of new features in collaboration with team partners:

- Map realized in collaboration with Ancitel team (Section 6.2.3)
- Decision Trees in collaboration with Warsaw School of Economics team (Section 6.4.7)
- Extended search in the Controllet in collaboration with Ortelio team (Sec. 6.4.10)
- Interaction with TET in collaboration with Insight team (Section 6.4.8)
- Section 7: Within this second year, the design and enhancement of the wizard for data visualizations (i.e., Controllet) has been one of the topics addressed during the team project meeting (Warsaw meeting in May 2016 ) and taking into account feedback generated by activities of Pilots to face the met difficulties in the visualization of data. A very important improvement of SPOD Beta platform is the User Interface, where first a review of the state of art of data visualization has been done and afterwards the design and implementation of the Controllet with the aim to display data in a simple way, once a user chose the dataset.
- Section 8: Privacy has been addressed in two different plug-ins, that is, Agora and Data Cocreation. All taken decisions have been included in Data Policy document (i.e., <http://spod.routetopa.eu/Data-policy>).

In summary, this document ends with the planning of the activities for the final and third year (detailed in Section 9) of the ROUTE-TO-PA project. The picture below depicts all the work that has already been done in the first two years of the project, what was planned for the third year for the final release of the SPOD platform. In detail, in the first year of the project, the design and realization of the alpha version of SPOD (WP4) were achieved considering as input the analysis of "User and system Requirement"s Workpackage (WP2) through the D2.4, the

“Models and Methods” Workpackage (WP3) through their deliverable D3.1, the requirements within this document and DoW. In the second year of the project, the realization of the Beta version of SPOD (WP4) was achieved considering as input the analysing requirements of the Deliverable D.4.1 "Alpha Version of SPOD", Pilot Feedback, Den Haag Meeting in February and Warsaw meeting in May 2016. Moreover, in the third year, we will be taking into account all feedback from the Pilots, the inputs from project meetings (such as the one who took place in Salerno in January and the following ones) as main suggestions and enhancements of the SPOD (summarized in the table of Section 9) that will be taken into account in order to increase performance of the platform and engage citizens.

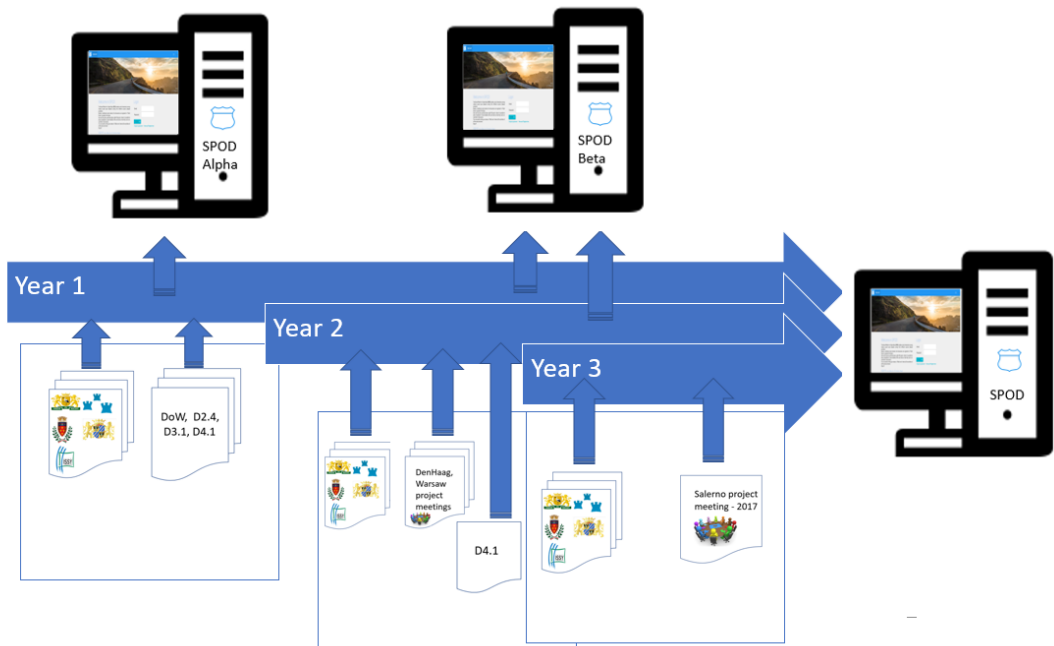


Table 1: Development of SPOD platform towards the third year of the ROUTE-TO-PA project

Finally, to make easier exploitation and dissemination, the introduction of container-based virtualization simplifies the deployment of the whole application suggesting a unique machine, i.e. “All-in-one” appliance, that allows easy deployment of the two platforms (SPOD and TET platforms) with a single address (URL) for the Public Administrations of the municipalities (Section 2.2.6).

# 1 INTRODUCTION

---

In this introduction we place this deliverable in the right context, as a followup of the activities that led to the Deliverable 4.1, produced on month 12 of the project.

In fact, in this deliverable, we need to address the Expert review reports suggestions and explain the process followed, and the outcome, for the production of the Beta version of the Social Platform for Open Data.

## 1.1 ADDRESSING CONSOLIDATED EXPERT REVIEW REPORT ISSUES

We address here, the general comments from the Consolidated Expert Review Report over the activities of the first year, but also the comments on D4.1, as this deliverable is the second year update on the D4.1.

### Recommendations over the future work:

#	Comment	Addressed in ..
R1	The knowledge co-creation process should be better designed, captured and embedded in the different activities and the implications/benefits for the desired results should be made more explicitly evident.	The process of how the Pilots feedback are embedded into the actual design of SPOD is addressed in Section 2
R2	It is recommended to pay more attention to the design, implementation and the reporting of the project's activities, as well as to the use of the results in the corresponding tasks and objectives	Section 2
R3	Demonstrate what is real novelty about SPOD and consider how it can boost engagement that will be able to address societal challenges and generate solutions that will create real value and lasting impacts.	Relevant publications in Section 10
R4	... Perhaps the role of leadership can be explored for the purpose.	Roles in Section 6.4.5
R5	... Introduce functions, which enable co-curation activities.	Co-creation, blogs in Section 6.1
R5	The value of involving ethical and privacy advisors should be demonstrated more explicitly in the research and innovation processes as well as in the project's products.	Privacy advisors feedback about privacy issues in Section 8.1 and Appendix –“A Feedback from privacy advisor from DR. Balachander Krishnamurthy”

### Objectives and Workplan: Is the progress reported in line with objectives and work plan as specified in the DoA? (About D4.1 Alpha version of SPOD)

#	Comment	Addressed in ..
O1	The deliverable is good but it does not make clear how were the non-functional requirements embedded from 2.1 in SPOD	Appendix – C “Non-Functional Requirement from D2.1”
O2	Also, how was the heterogeneity of the context dealt with in the pilots regarding the design and the development of SPOD. It is reported that between 2015 and March 2016, 10 versions of SPOD and TET were produced but there is no evidence of the success of the process.	Section 2

O3	The implementation of the agile methodology needs to be explained. Section 7 deals with Design concerns but also other concerns such as related to “usability” are not considered despite issues raised about effectiveness and efficiency of the platform.	Section 2
O4	Table 11 reports on feedbacks from partners but the report does not clarify how the data was generated as well as issues regarding to its validity and reliability.	Section 2
O5	The list of Acronyms includes terms and definitions, but the latter are not actual definitions. The project’s specific definitions need to be defined as to set up its own terms of reference.	The definitions are now explicit in the text and, therefore, we eliminated the list of Acronyms
O6	Finally, the implications of change of base social platform from Elgg to Oxwall should be better elaborated.	Section 4.1

### The implementation

#	Comment	Addressed in ..
I1	More attention should be done to envisage that ethical and privacy issues and recommendations are reflected in the research and innovation actions.	Section 8

## 1.2 FROM ALPHA TO BETA: A ROADMAP OF THE MAJOR CHANGES

The realization of the major features tackled in the Beta version of SPOD (WP4) was achieved considering as input the analysing requirements of the Deliverable D.4.1 "Alpha Version of SPOD", Pilot Feedback , Den Haag Meeting in February and Warsaw meetings in May 2016 (minutes, general assembly and other documents of the meetings are available at <http://service.routetopa.eu:8000/d/481c8b5f9c/>).

### In particular, the major topics addressed have been:

- **Collaborative space** (see Sec. 6.1) facilitates discussions around the meaning of data and two kinds of rooms have been identified *Cocreation Knowledge room* (See Sec. 6.1.1) and *Cocreation Data room* (see Sec. 6.1.2 ); The system design has been achieved considering as input the analysing requirements of the previous Deliverable D.4.1 "Alpha Version of SPOD", Den Haag Meeting on February 9-10 and Warsaw meetings on May 25th 2016 . *Cocreation Knowledge room* is the SPOD tool where users can remotely meet and digest a particular topic together; instead the *Data Co-Creation rooms* are virtual places where small groups of participants meet together to collaboratively create new datasets. The new feature is supported from SPOD v.1.3 onwards.
- **User Interface and advanced support** (see Sec. 7) is a section addressed to main and fundamental component in SPOD platform to visualize data and then for starting a discussion around the visual representation with charts, interactive maps. First, metrologies for choosing the right chart and the taxonomy of interactive dynamics, then two famous asynchronous collaborative visualization systems and other two minor tools have been analyzed. Finally, a list of the best practices with data visualization has been compared. The new design of the User Interface with new features of the Controllet has been realized from SPOD v. 1.7 onwards and considering as input the analysing requirements of the previous

Deliverable D.4.1 "Alpha Version of SPOD", feedback received from Pilots and Warsaw meetings on May 25<sup>th</sup>, 2016.

- **Maps from geo-servers** (see Sec. 6.2.3) is a new feature to allow creating a cartographic project composed of multiple layers of public datasets that contain geographic territorial data, tied together by some common theme. This new feature has been realized by Ancitel team and has been introduced from SPOD v. 1.11 onwards, considering as input the analysing feedback received from Pilots.

**The minor topics addressed have been:**

- **Blogs** (see Sec. 6.1.3) is a new component allowing to add post quick blog entries, with datalets embedded, and is public to all (even to unauthenticated users). The aim of introducing the Blog is to show the Knowledge as public to all Internet users as result of interactions within the SPOD platform, co-creations and discussions. For an example, on the Prato platform, the Blogs has been activated (see <http://prato.routetopa.eu/blogs>). This feature supported on SPOD v. 1.10 onwards.
- **Events** (see Sec. 6.3) allows managing an event with discussions and relevant datasets. The introduction of the events in SPOD was solicited by of the use cases highlighted in the DL 2.4 and DL 3.1 deliverables and it is available from SPOD v. 1.11 onwards.
- **Attach a document/image** (see Sec. 6.4.1) allows posting a file in the collaborative discussion chat of the Cocreation data room. This feature has been achieved considering as input the analysing requirements of the previous Deliverable D.4.1 "Alpha Version of SPOD" and available it is available from SPOD v. 1.12 onwards.
- **Changing the Welcome Page (Open Wall)** (see Sec. 6.4.2) provides the welcome information (such as contact email asking for information, web addresses of the project to watch activities on Facebook and Tweet) also to unauthenticated users, with embedded video to give a look at all features of the SPOD platform; furthermore, manual and videos are available to all unauthenticated users. This feature has been achieved considering as input the analysing requirements of the previous Deliverable D.4.1 "Alpha Version of SPOD", feedback received from Pilots and it is available from SPOD v. 1.7 onwards.
- **Delete a Public Room** (see Sec. 6.4.3) is a feature of the Administrator and has been achieved considering as input the analysing requirements of the previous Deliverable D.4.1 "Alpha Version of SPOD", feedback received from Pilots and it is available from SPOD v. 1.3 onwards.
- **Recommended datasets:** this feature (see Sec. 6.4.4) allows to stimulate discussion and simplify the use of Open Data all the users, that are authorized to create Agora room and it has been achieved considering as input the analysing requirements of the previous Deliverable D.4.1 "Alpha Version of SPOD".
- **Roles** (see Sec. 6.4.5 ) allow reserving special action for a specific role. In particular, actions for Blog, Event and Agora plug-ins have been implemented. The platform administrator can enable a specific role to create an event, a blog post or an Agora room. This feature has been achieved considering as input the analysing requirements of the previous Deliverable D.4.1 "Alpha Version of SPOD" and it is available from SPOD v. 1.7 onwards.
- **Customizing the "What's New" page with widgets** (see Sec. 6.4.6) allows to enrich the "What's new" page with custom and configurable widgets (e.g., Tweets, Agora Activities, SPOD Help RSS feeds, Static content, etc.) to provide extra information about what's happening in SPOD platform. This feature has been achieved considering as input the analysing requirements of the previous Deliverable D.4.1 "Alpha Version of SPOD", feedback received from Pilots and it is available from SPOD v. 1.7 onwards.

- **Decision trees** (see Sec. 6.4.7) allow to create and analyse Decision Tree and a mechanism has been implemented to integrate Decision Tree inside SPOD's My Space. The new feature has been implemented in collaboration with SGH Warsaw School of Economics team and has been introduced from SPOD v. 1.11 onwards.
- **Interaction with TET** (see Sec. 6.4.8) has been implemented through the DataFlow from SPOD to TET allowing participating to all discussion rooms created on the SPOD platform; all users have the ability to create Pivot tables on datasets in TET. This feature has been achieved considering as input the analysing requirements of the previous Deliverable D.4.1 "Alpha Version of SPOD"; it is available from SPOD v. 1.11 onwards.
- **Join authentication** (see Sec. 6.4.9) introduces an improvement of the OPENID Authentication Server to accommodate requests from partners and the pilots (e.g., new user Interface, password reset, new welcome page and user registration). It is available from SPOD v. 1.8 onwards.
- **Extended search** (see Sec. 6.4.10) allows discovering data sources, using an additional utility, UltraClarity, a search engine indexing the full text of datasets stored in CKAN repositories. It has been developed in collaboration with Ortelio team and has been introduced in SPOD platform v.1.5.
- **Manager view** (see Sec. 6.4.11) explores the back end of the SPOD platform, to describe the additional and advanced features list introduced by UNISA team (i.e., User roles, management Privacy, Manage rooms in the Agora, Customize the "What's new" page etc.)
- **Email Notification Strategy** (see Sec. 6.4.12) allows the users to know what activities are happening on the platform. At this time, a way to ensure this service has been addressed and only a working proof of concept related to the activities in the cocreation room has been implemented. This feature has been achieved considering as input the analysing requirements of feedback received from Pilots and it is available from SPOD v. 1.12 onwards.
- **Privacy** (see Sec. 8.1) has been addressed in two different plug-ins, that is, Agora and Data Cocreation. All decisions taken have been included in our Data Policy (i.e., <http://spod.routetopa.eu/Data-policy>). By default, rooms in the Agora are visible only to authenticated users. It is now possible to configure SPOD so that Agora rooms are visible (in read-only mode) by unauthenticated users. Indeed, in the Data Co-Creation rooms, citizens become active players because they can contribute to creation of the open data and an issue related to this functionality is the violation of privacy whether personal and sensitive information are unconsciously added in the dataset and it is addressed to make the creator of the dataset aware of this possibility. This feature is available from SPOD v. 1.10 onwards.
- **All-in-one virtual appliance** (see 2.2.6) provides the introduction of container-based virtualization to simplify the deployment of the whole application suggesting a unique machine (Virtual Machine). The container-based virtualization for ROUTE-TO-PA project includes SPOD platform, TET-enabled CKAN and Wordpress.

## 2 THE AGILE DEVELOPMENT PROCESS

### 2.1 THE OVERALL CYCLE

The SPOD project follows an AGILE methodology with continuous iteration of development and testing in the software development life cycle of the project.

The Cycle consists of 4 macro-steps:

1. Design: provided the Pilot Feedback from previous versions the feedback is catalogued and inserted as issues into the GitHub issue tracker system. A “*Reaction to feedback*” document is issued for each pilot, and constantly updated (on the project file repository) during the all the phases development, as issues are tackled and resolved (also in successive iterations).
2. Implementation: with internal testing each issue is placed in projects and dealt with by developers. More details in the following subsection.
3. Release: on the 5 platforms of the 5 pilots, so that they can participate in the next phase. The output of this phase is a document of “*Release Notes*”, where all the new features, issues, bugs dealt with and solved in the release are shown.
4. Supporting Pilots’ Testing: as testing is performed on pilots’ sides, according with their policies of engagement with the involved community, there is the need to support the process of communication with the pilots. This is realized through semi-structured interactions via exchange of documents. In this phase each Pilot (for each test that took place) produces a document, called “*Pilot Feedback*” where the results of the testing and all the comments are reported to the design team. In the next phase Pilots’ feedback is catalogued and inserted into the GitHub repository.

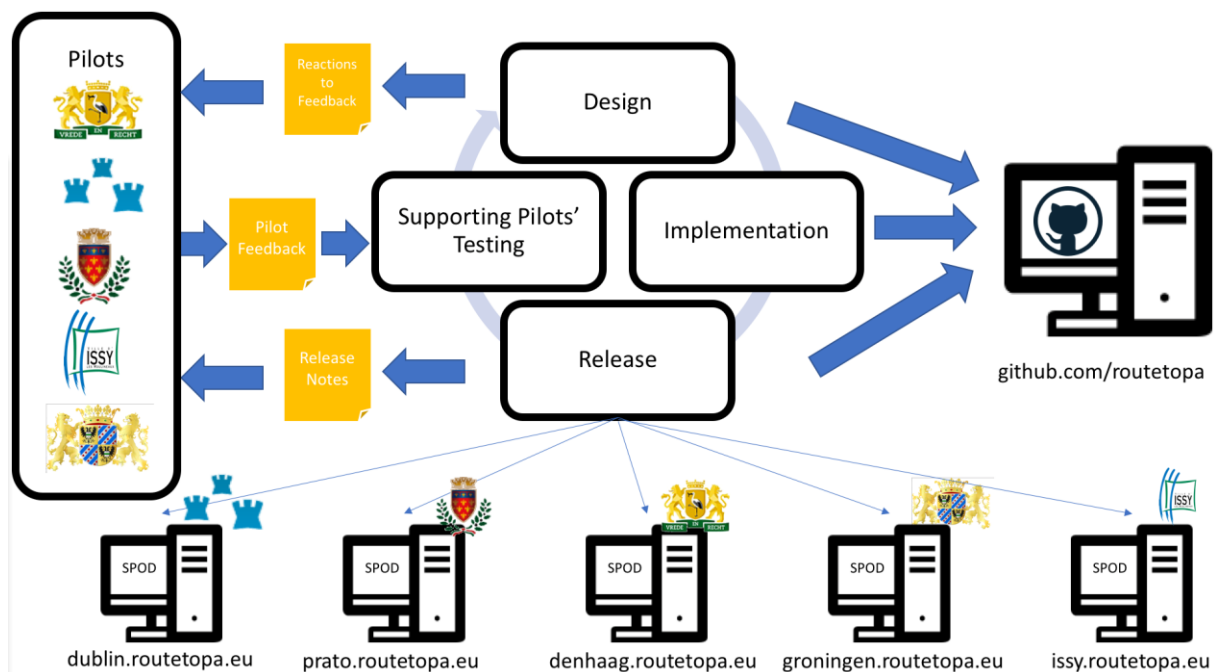





Figure 1: The Agile Development Cycle as implemented for SPOD




## 2.1.1 RELEASE NOTES

List all release notes from February 2016 until January 2017.

Name of the release	Major features
SPOD "Furud" v.1.1 (February 5 <sup>th</sup> , 2016)	Fixed bugs
SPOD "Giedi" v.1.2 (March 2 <sup>nd</sup> , 2016)	Controllet: (1) Filter criteria: Controllet makes possible to filter datasets by text (e.g., Start with, End with ) or to filter by number (e.g., =, !=, >, <, <=); (2) Aggregation: Controllet allows aggregation to combine rows together and to perform some operation on their combined values; common aggregation are count, sum, min , max avg, first and last.
SPOD "Heze" v.1.3 (March 23 <sup>rd</sup> , 2016)	Co-creation room: a new collaborative space to guide user to constructive discussions within a small, private (by invitation) group. It includes tools specifically designed for small groups such as cocreating a template document for each of the three phases explore, ideas, outcome and annotation via post-it of the datalets to be discussed (with synchronous notification for real-time discussions). <ul style="list-style-type: none"> <li>Browser compatibility: Co-creation works in these browsers such as Chrome, Safari (IOS);</li> <li>Known NOT to work on Firefox, Edge, Safari on Windows (popups appear repeatedly in the foreground in the SPOD platform);</li> <li>Notification server can send update coming from rooms that don't belong to you;</li> <li>First version of "Co-creation" room is available only in English.</li> </ul> Interaction between SPOD/TET: the following services are provided, although visible only on TET side: <ul style="list-style-type: none"> <li>"Save to my Space" feature: A datalet created with pivot table in TET will be imported in "My space", a personal space of SPOD;</li> <li>"Get public rooms where a dataset is used in discussion": list of rooms where a dataset imported from TET is used and showing the following information for each room: title, summary, number of view, number of comments, number of open data , creation and timestamp.</li> </ul> Agora: the administrator can delete a room
SPOD "Izar" v.1.4 (April, 8 <sup>th</sup> 2016)	Cocreation room: The following issues highlighted from Wise & Munro, Utrecht, Prato teams , have been resolved: <ul style="list-style-type: none"> <li>Create a room with empty field: It is possible to create a room leaving the "members" field empty to fill in later;</li> <li>Google docs;</li> <li>A new light User Interface without "log in" button and not chat enclosed available; the "enlarge" button in the upper left to enlarge the document.</li> <li>Revision history for a shared document: it is possible to see the changes made to the shared document, only if the user is already logged into his/her Google account;</li> <li>Resolved screens duplicate in smaller versions within SPOD;</li> <li>Datalets and notes work in the synchronous mode.</li> </ul> Exporting data of Agora discussion to different formats (i.e., Json, xls file types): in the administrator panel, for research purposes the administrator can see and save what happens in the room after a discussion.
SPOD "Jabbah" v.1.5 (May 19 <sup>th</sup> , 2016)	Towards Beta version of SPOD platform: <ul style="list-style-type: none"> <li>Co-creation room: knowledge workspace with new features</li> <li>Agora: delete and rename a room (only user with administrator role)</li> </ul> Cocreation room: a new shared document with new features ( e.g., a new collaborative editing with buttons to enlarge ( document or datalet ) or split on the top right side (  ), to add a note, a datalet or "Import from My space" on the right bottom side (  ); collaboration in real time with other writers remotely; recording changes and revisions, i.e a playing head slider to drag to the left to move between versions (  ); multiple authors with text in different color to allow to keep track of who has written what; import a created datalet in My Space ; Export datalet to png, HTML and RTF formats Controllet <ul style="list-style-type: none"> <li>A new interface in Step 1) to allow to show all metadata of a selected dataset;</li> <li>Search in Treemap and in List view;</li> </ul>

	<ul style="list-style-type: none"> <li>○ User interface improvements: each color of a datasets identifies the correspondent provider (i.e, same color as the treemap);</li> <li>○ Choice of datasets by Open Data provider (i.e.: CKAN, Issy-les-Moulineaux);</li> <li>○ Search also the content of datasets, through the UltraClarity search portal for CKAN, by using, n Step 1) a new tab, named “Extended search”;</li> <li>○ Datalets: added stacked column, bar and area charts.</li> </ul> <p>Agora</p> <ul style="list-style-type: none"> <li>○ Notification activities: to limit the number of notification (i.e., posting and likes in Agora) in “What's new” page, notifications are now grouped together in-group of 10 posts per room (this value is configurable for the moment only by code) and notification appears only once;</li> <li>○ Display full messages/posting: avoid extra click in enlarging the message for long messages/post;</li> <li>○ Changes the icons from the gray in blue in the Agora-room;</li> <li>○ Administration panel: To analyse data of each discussion room of the Agora it is possible to export excel (or JSON) file with the discussion of each rooms of the Agora;</li> <li>○ Security problems</li> </ul>
SPOD “Kuma” v.1.6 (June 24 <sup>th</sup> , 2016)	<p>Cocreation data room:</p> <p>Co-creation data room is a collaborative, real-time spreadsheet; It combines the ease of authoring and multi-user editing of spreadsheet enriched with export/import csv files. It allows citizens to collaborate on a spreadsheets over SPOD platform in real-time to create dataset with useful criteria for data validation, and to publish a dataset to share for further discussion in social collaboration tools in SPOD (e.g., in What’s New, Agora ).</p> <p>Some functionalities:</p> <ul style="list-style-type: none"> <li>• Create a New Sheet, Export CSV and Import CSV files, list of function (SUM, PRODUCT, AVG etc.);</li> <li>• Collaboration on the same dataset with support of validation criteria: users can edit the same document at the same time. Everybody's changes are instantly reflected on all screens. Finally, the dataset will be publish it in a new provider, named SPOD provider available for Controllet to share it in social tools (e.g., Agora, What’s new);</li> <li>• Each dataset can be authored providing also sources (in a shared text editor) and metadata</li> <li>• Each published dataset is accessible to build a datalet in the traditional ways (buttons in What’s new, Agora’s public rooms, etc.);</li> <li>• “View all dataset” is a new feature to show the datasets created and exported by groups in the cocreation data rooms and it allows to see the dataset, the notes and the link of the dataset and more..</li> </ul> <p>Controllet:</p> <ul style="list-style-type: none"> <li>• User Interface clean up (possible work on support of creation of charts);</li> <li>• Data quality of a dataset: in step 2) a further check on the data quality of an uploaded dataset highlights the error types in each field of the dataset and “?” white question mark appears over the fields of a dataset with the suggestions about field type;</li> <li>• Multiseries: In step 3, multi series charts are easier to create (this is a first step toward further improvements available in the next release).</li> </ul>
SPOD “Lesath” v.1.7 (July 27 <sup>th</sup> , 2016)	<p>Controllet: User Interface cleanup;</p> <p>Roles: now in SPOD it is possible to choose a role for a user. Currently, it is possible to choose the moderator role, that in Agora allows the creation of new rooms, only the moderators can create (configurable) a new room</p> <p>SPOD Welcome Page and “What’s new” page : a new user interface:</p> <ul style="list-style-type: none"> <li>• Log In: log into your account from the same page;</li> <li>• Sign up (Registration in SPOD);</li> <li>• Possible to have users enrolling by themselves checking only the email address (requires the configuration of a file);</li> </ul> <p>Open Wall replaced with new User Interface of the Welcome page:</p> <ul style="list-style-type: none"> <li>• “What’s new” page: custom and configurable widgets (Tweets, Agora Activities, SPOD Help RSS feeds, Static content, etc....);</li> </ul> <p>Improved stability and usability in Agora:</p> <ul style="list-style-type: none"> <li>• Improvement of the visualization of the Comments graph;</li> <li>• Improved the synchronization between the graph and the comments, when many comments are introduced.</li> </ul> <p>Cocreation data room introduces new features:</p> <p>Co-creation data room is a collaborative, real-time spreadsheet; It combines the ease of authoring and multi-user editing of spreadsheet enriched with export/import csv files. It allows citizens to collaborate on spreadsheets over</p>

	<p>SPOD platform in real-time to create dataset with useful criteria for data validation, and to publish a dataset to share for further discussion in social collaboration tools in SPOD (e.g. in What's New, Agora ).</p> <ul style="list-style-type: none"> <li>• Input from a Map from within the spreadsheet</li> <li>• Localization</li> <li>• Interactive map</li> <li>• Map locations</li> <li>• Notification in a discussion: in discussion room, a new message is notified with an icon with the counter of a number of received and not read messages.</li> </ul>
SPOD "Mira" v.1.8 (September 12 <sup>th</sup> , 2016)	<p>Registration policy: a new user can registers on SPOD platform only if administrator enables this feature.</p> <p>Agora: Add a new "send" button to post a message</p> <p>Cocreation data room: enhancement for export and import of a csv files (in particular, import of CSV file with delimiters as semicolons and tab)</p>
SPOD "Nash" v.1.9 (September 30 <sup>th</sup> , 2016)	<p>Migration of SPOD project (source code ) towards Git Hub platform, an open source software (<a href="https://github.com/routetopa/">https://github.com/routetopa/</a> )</p> <ul style="list-style-type: none"> <li>• Source code of the project</li> <li>• Initial analysis of the feedback previously sent.</li> </ul> <p>Cocreation data room: a new management of the services allows to re-connect each service when they are shut down.</p>
SPOD "Okul" v.1.10 (November 14 <sup>th</sup> , 2016)	<p>Cocreation: a new User Interface design with refactoring</p> <p>Privacy settings: it is now possible to configure the SPOD so that Agora rooms are visible (read-only) from an unauthenticated user. By default, the situation is the same as before, i.e., the rooms are visible only to authenticated users, but it is possible to change it in the admin plugin, per site.</p> <p>Upgrade Polymer Library in SPOD project</p> <p>A new plugin for Blogs is created. It allows to add post quick blog entries, with datalets embedded, and is public to all (even to unauthenticated users). Only a user with "blogger" role can create a blog. Authenticated users can vote (from 1 to 5) the blogs. By default the component is NOT activated and can be very quickly activated by the admin (or ask the UniSa team). For an example, see <a href="http://prato.routetopa.eu/blogs">http://prato.routetopa.eu/blogs</a>.</p>
SPOD "Polaris" v.1.11 (December 5 <sup>th</sup> , 2016)	<p>Integration of multilayer Ancitel map datalet and map creator in SPOD: Ancitel created a new map datalet based on openlayer that is capable to visualize KML, WMS and GeoJSON data on a multi layer map (such as OSM Standard, Cycle, Transport, Landscape and Humanitarian). Unisa team integrated the authoring system and the datalet inside SPOD.</p> <p>A new button has been introduced in My friends' activities/What's new that opens the map authoring system.</p> <p>Attach image/document: The discussion in the Cocreation data room allows file attachments to a post. All file types permitted for upload are available from admin panel.</p> <p>Cocreation data room with copy and paste:</p> <ul style="list-style-type: none"> <li>• User can move in the spreadsheet with arrows keys. The selected cell will be highlighted. User can edit the selected cell by pressing enter key and exit from edit feature with esc key;</li> <li>• By using shift and arrows keys at same time the user can select multiple cells. The selection will be highlighted;</li> <li>• By pressing ctrl + c the user can copy the selected portion of the sheet and, after selecting the destination cell by pressing ctrl + v, s/he can copy the content;</li> <li>• The user can copy from excel file cells into spreadsheet of data room cocreation;</li> <li>• Copy between different spreadsheets of the different cocreation data rooms.</li> </ul> <p>Caching server's settings improves performances</p> <p>Mail notification: each user profile can configure real-time an e-mail notification about site activity (e.g., Base, Comments, Messages, Newsfeed, Blogs)</p> <p>Decision Tree visualization in SPOD platform</p> <p>The "SilverDecisions" project is developed at Decision Support and Analysis Division, Warsaw School of Economics. SilverDecisions is a software for creating and analysing decision trees. SilverDecisions is available visiting its website: <a href="https://github.com/bkamins/SilverDecisions/wiki">https://github.com/bkamins/SilverDecisions/wiki</a>. SilverDecisions can be used in My Space of the SPOD platform, clicking on the new button .</p> <p>User email verification: When a new user registers, his/her e-mail address must be verified with an e-mail message</p> <ul style="list-style-type: none"> <li>• Re-send verification e-mail if user lose it</li> <li>• Allow authentication only for users with verified e-mail</li> </ul>
SPOD "Regulus" v.1.12	<p>Event: An event in SPOD can be planned only by the user with assigned "Events" user role. Any user can see upcoming and past events, invitations to the events.</p>

(December 22 <sup>nd</sup> , 2016)	<p>Other available features are:</p> <ul style="list-style-type: none"> <li>• A datalet (a new datalet or select a datalet from private space (i.e., My space page) can be embedded in the description of the event;</li> <li>• Add link to google maps at a particular location with a marker on a particular point (for example, location of the event);</li> </ul> <p>Activity notification for Cocreation room:</p> <ul style="list-style-type: none"> <li>• Only administrators receive e-mail notification when a new data room will be created</li> </ul> <p>User Mass Mailing: In admin area, an administrator can send emails to several users at once. Some features:</p> <ul style="list-style-type: none"> <li>• Email format - allows choosing "plain text" or "HTML" format. Note that the actions menu of Email body block changes depending on selected format;</li> <li>• User roles- allows to send email to users with certain User Role;</li> <li>• Subject - subject of the email;</li> <li>• Email body - body of the email.</li> </ul>
SPOD "Beta" v. 2.0 (January 31 <sup>st</sup> , 2017)	Fixed bugs, Updated Documentation (i.e., Quick Helper on line, User Manual, Administrator's Guide)

All the release notes (with more details) are available at <http://service.routetopa.eu:8000/d/481c8b5f9c/>.

## 2.1.2 FEEDBACK RECEIVED FROM PILOTS

Each Pilot has submitted feedback through template document that had the following structure (see table 1 below): the document contains a specific section for each issues (i.e., Bugs, Enhancement, New features, Others). To each issue is assigned a tag to rate both the *Severity* (i.e. something that ranges from cosmetic defects to blocking issues) and *Priority* (i.e., to indicate how important an issues for the current release, milestone and it used for planning) .

The available tags for *Severity* and *Priority* are High/Low/Normal to share the importance of a bug.

Bugs	Description	Severity (High/Low/Normal)	Priority (Normal/Minor/Wishlist, Important, Critical)
Enhancement			
New Features			
Others			
Questions	Description		

Table 2: Template document of feedback

The feedback, received during the second year of activities of the Pilots, are summarized as follows:

Pilot	Version	Date	Bugs			Enhancements			New features			Questions			Other		
			C	IP	O	C	IP	O	C	IP	O	C	IP	O	C	IP	O
Den Haag	1.3	11/3/2016	4			9		2		1	1			1			
	1.3	30/3/2016	3			2		1	1					1			
	1.6	07/07/2016	1			1		5									
	1.10	22/11/2016	2					2						7			3
	1.2	11/03/2016	4			9		2		1	1	1					
	1.6	07/07/2016	1			1		5									
	1.10	22/11/2016	2		1			2				7		4			
Issy	1.2	7/03/2016	6					1									
	1.6	7/7/2016	4			4		2							3		1
	1.7, 1.8	20/09/2016	6		1	1	3	8		1	3						
	1.9	04/10/2016	6			2											

	1.10	16/12/2016	1		2		1	2								1
	1.11	14/12/2016	2													
Prato	1.1	1/03/2016	4													
	1.3	06/04/2016	1			1								1		
	1.5	02/06/2016	7			2		1	1		3					
	1.6	30/05/2016						1	2		2					
	1.6	04/07/2016	8					2	2		1					
	1.7	17/09/2016	2			3								1		
	1.9	18/10/2016	1			5						1				
	1.9	20/10/2016				2		1	1			1			3	
	1.11	06/12/2016	1		2			4								1
	1.11	20/12/2016			7											
Dublin	1.4	29/04/2016				10		10				1				
	1.9	13/10/2016	2			1						2		1		2
Groningen	1.1	04/02/2016				2			5		1					
	1.1	18/02/2016				1			3							
	1.3	29/03/2016	2			5			2		1	2				
	1.4	April/May 2016	6			7	3	4			1				1	
	1.6	11/07/2016	7			4		3							1	
	1.8	26/09/2016	2			2		1								

Table 3: A summary of feedback received from Pilots activities during the second year

To each feedback, SPOD team replied with “Reactions to Feedback” that summarized the Bugs/Enhancement/New Features/Others/Questions and ticket/issues on GitHub (or internal Taiga before September 2016) and are available at <http://service.routetopa.eu:8000/d/481c8b5f9c/>.

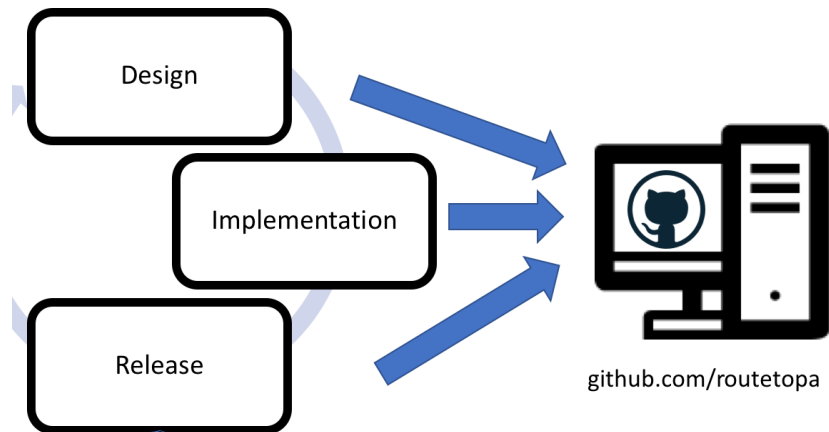
## 2.2 THE DEVELOPMENT PROCESS: DESIGN, IMPLEMENTATION AND RELEASE

The source code and development management of the SPOD project migrated from an internal system to GitHub [1] on September 2016<sup>1</sup> with the address <https://github.com/routetopa> for the entire project.

GitHub is the largest source code repository in the world, for hosting projects with a Git revision control system for the management of the source code. In addition, GitHub includes many features, like milestones, tracking and issue management; it also supports agile development models.

Here we provide more detail on the internal part of the cycle, zooming in the process shown in the figure below:

<sup>1</sup> Before September, the same methodology was employed by using the mechanisms, described later, for collecting feedback, raising issues (by using a different internal system, Taiga) and managing the versioning (GitLab, at [http://service.routetopa.eu:7480/users/sign\\_in](http://service.routetopa.eu:7480/users/sign_in)).



*Figure 2: The interactions with GitHub during the cycle*

GitHub is used in the development to:

- Track the issues raised by Pilots' feedback, requirements, internal testing, etc.
- Monitor development through the *Projects* in GitHub
- Manage all the source code, with versioning
- Managing the Releases of SPOD
- Manage the Technical Documentation

Now, we introduce briefly each topic, reminding that all of the following is absolutely open for access at <https://github.com/routetopa>.

#### 2.2.1 DESIGN: ISSUE TRACKING

As the testing is realized by the pilots, our decision was to let them provide the feedback by simply filling up templates feedback documents, which they are sent back to the design team, and then processed manually and inserted into the issue tracking that is offered by GitHub. In fact, GitHub allows to easily managing the issues to be dealt with during the development in an integrated issue tracker.

GitHub issue tracker (as shown in below *Figure 3*) shows three areas, i.e. default view, label view, milestone view.

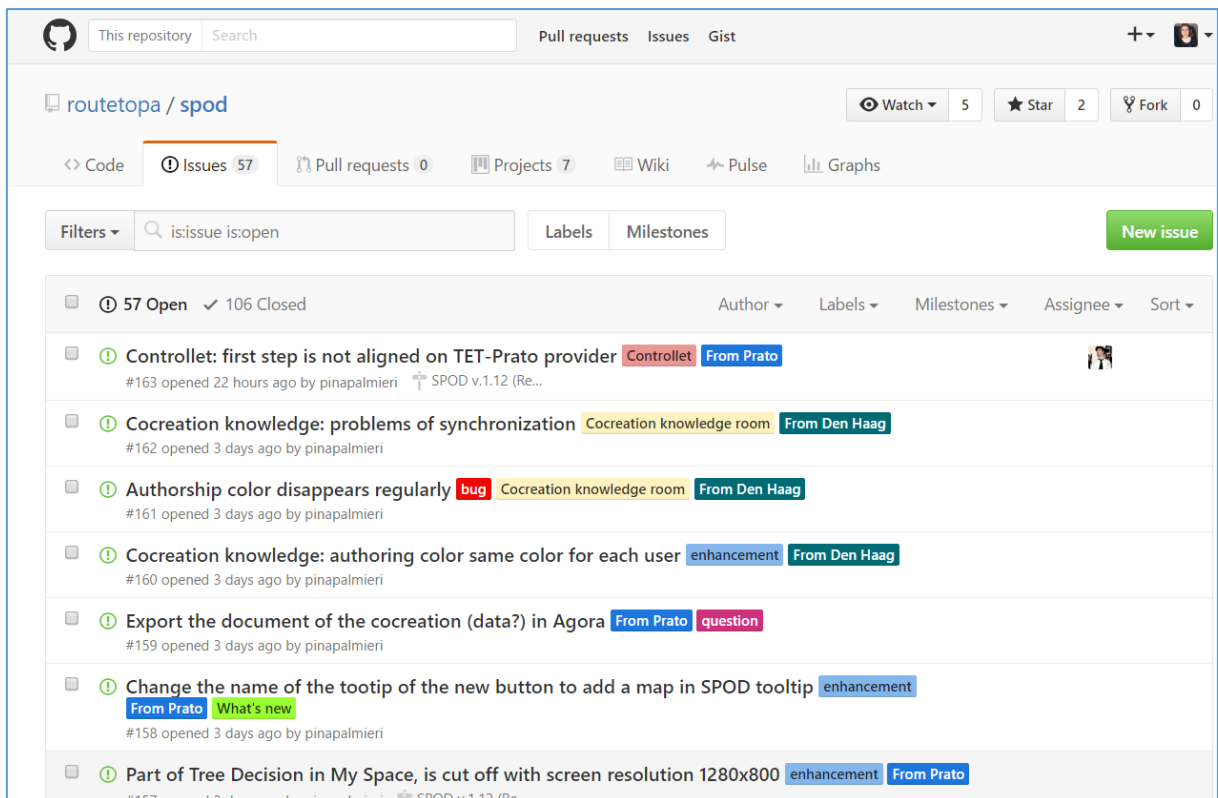


Figure 3: The issue tracker view in SPOD code GitHub

The management of each issue occurs in a different view to give a name and a description of the new issue, to assign it to the developers with small tags, named Labels, to plan in a milestone that is an iteration with issues linked to it (see the subsection on releases).

### 2.2.2 DESIGN/IMPLEMENTATION: MONITORING THE DEVELOPMENT

During the design and implementation, in the agile methodology, one has to select the issues to be currently included in the current cycle. This is done accordingly to the *Projects*<sup>2</sup> management of GitHub, that is a quite recent feature to help the team to organized into custom columns with tasks/issues into a board interface, where each issue, the task will be represented by a card. The team can organize these cards. The word *Kanban* is Japanese for “visual signal” or card; this methodology is a way to implement Agile and Lean management methods, introducing a visual nature to help the team to communicate more easily and to plan when the work must be done. The team can follow the flow of work in each project on GitHub.

For each project for SPOD project on GitHub, the flow of work (e.g. all tasks, issues) has been organized into a board interface, representing each issue with a card. Those cards can be rearranged into custom columns, so each developer can track how different project components are moving through a workflow.

All requirements received from each Pilot become new issues with a name, a description and assigned to a UNISA team developer, also assigned to milestone and labels. The labels are small tags, such as Pilots naming convention (e.g. From Prato etc.), and organized in different groups of labels like issue types (e.g., Agora etc.), issue priorities (e.g., High Priority etc.) and issue statuses (e.g., Open Closed etc...)

<sup>2</sup> GitHub launches built-in project management boards for its code repos (Sep 14, 2016)

<http://www.pcworld.com/article/3120302/github-launches-built-in-project-management-boards-for-its-code-repos.html>

Requirements (commonly defined as a user feedback ) are ordered and prioritized in a Backlogs, which should be released to pilots team (or be at least in a releasable state), every iteration (a.k.a. Iteration - a fixed duration, 1,2,3,4 or 5 weeks, for working on fixed requirements defined at the begging of the iteration).

In SPOD project, UNISA team, after collecting all requirements, stored in document of feedback, received from each Pilot, after creating all issues on GitHub, writes them into a BackLogs session of the respective Project, previously created. The coordinator of SPOD project initiates an estimation and prioritization session with all UNISA developers team. As a result of this session, the items in the Backlog have an initial estimation and a prioritization. The development process is iterative and all steps of each Project has been organized in the followings steps:

- BackLogs: it is a list of all requirement received from each Pilot, functions and feature of SPOD project. It is created and classified in the early phase.
- TODOs: the selected requirements on which work to release
- In progress: all issues to develop
- Testing: in each iteration, a testing phase occurs
- Deployed: when an iteration ends, the SPOD platform with new version is delivered to each Pilot

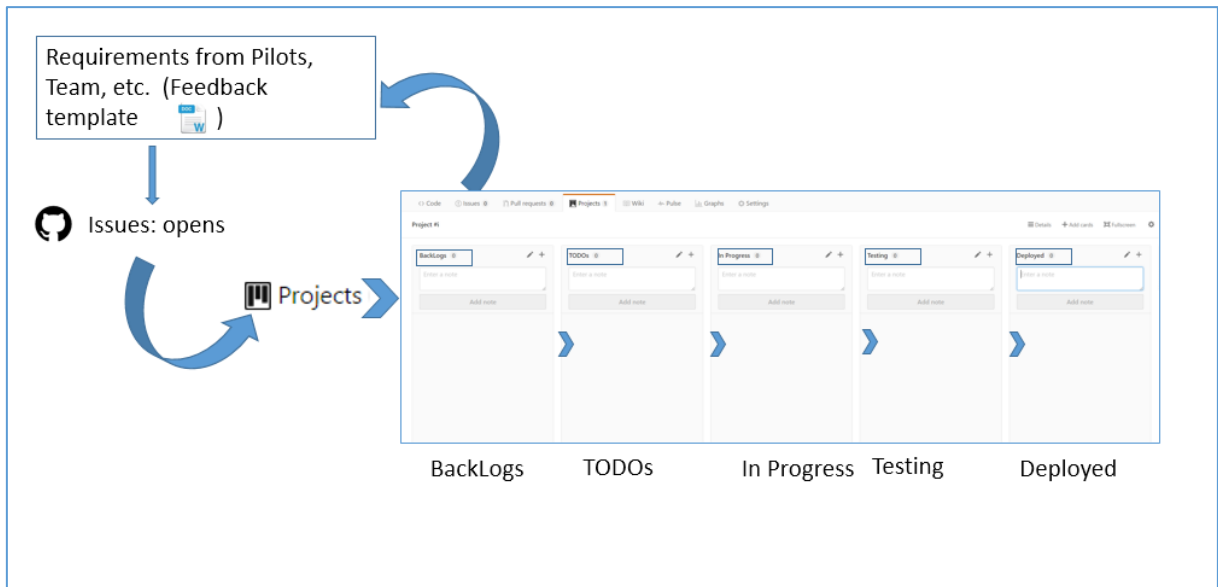


Figure 4: A SPOD development life cycle towards next release



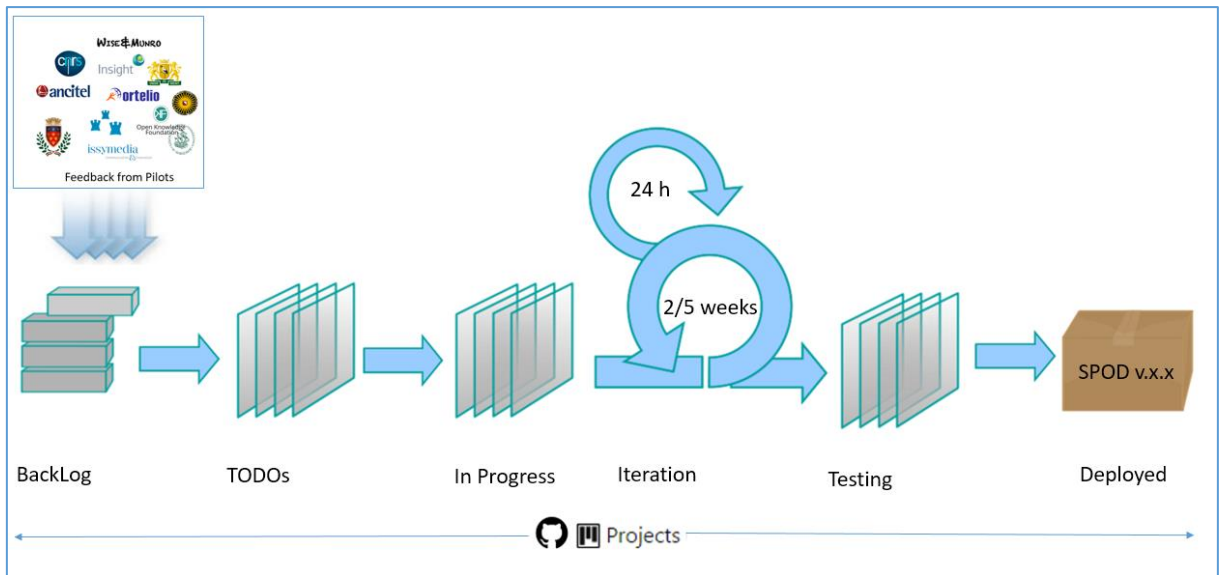


Figure 5: Tracking progress of each GitHub project for SPOD platform

## Tracking progress of a SPOD project

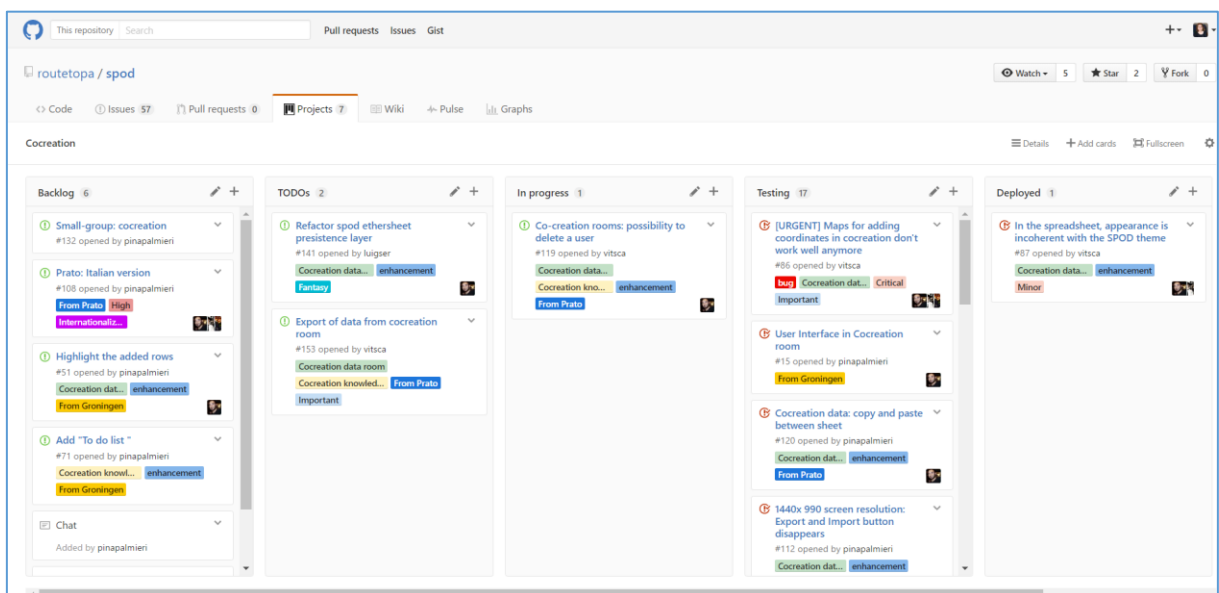


Figure 6: Kanban board of SPOD code source on GitHub, selecting a project, e.g. "Cocreation"

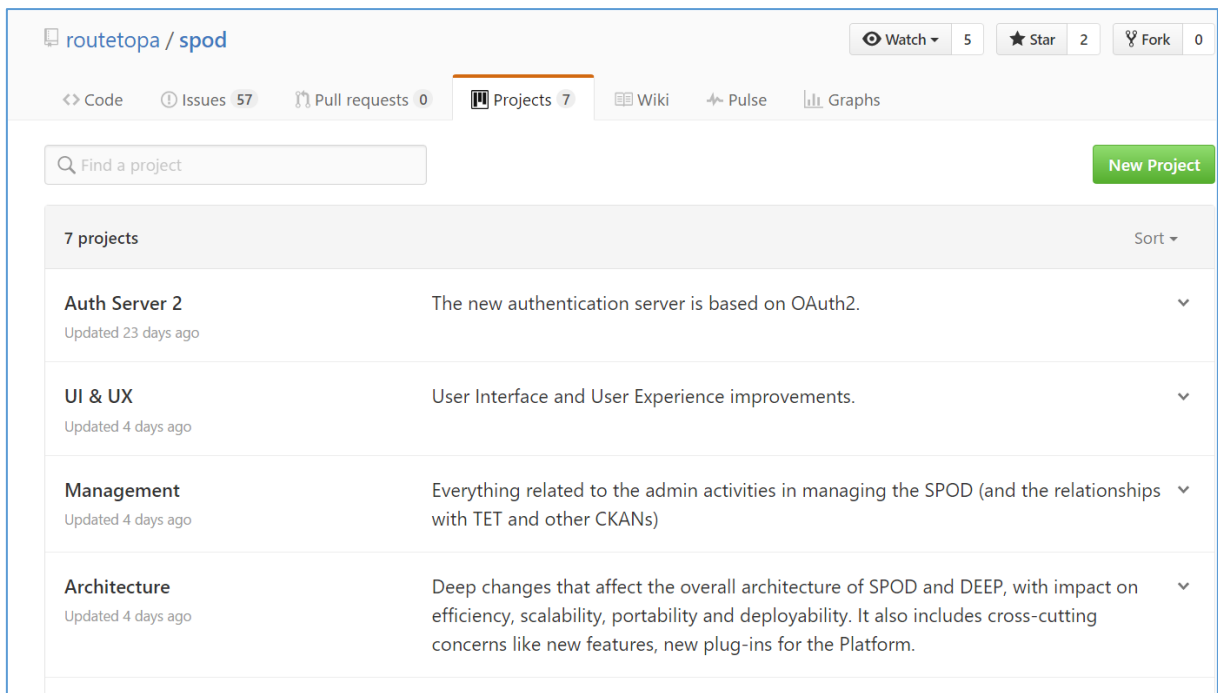


Figure 7: Some of the current projects for SPOD on GitHub

### 2.2.3 SOURCE CODE REPOSITORIES AND METRICS

The SPOD source code has been organized in the following repositories:

- Spod: Placeholder project with no source code, where all the issues, projects, milestones, releases, wiki, etc. are managed
- Deep-components: It is a repository of datalet and controllet
- Spod-plugin-cocreation: the "SPOD CoCreation" plugin allows user to collaborate in creating shared knowledge and datasets.
- Spod-theme-matter: The official theme for SPOD.
- Spod-plugin-tchat: The "SPOD Threaded chat" plugins enables nested comments.
- Spod-plugin-myspace: The "SPOD MySpace" plugin offers a space where users can save their datalets, links, images and notes.
- Spod-plugin-ode: The "SPOD Open Data Enabler" plugin enables the use of datalets and data providers in SPOD.
- Deep: The Datalet-Ecosystem Provider (DEEP) provides the list of available datalets, and a repository of their implementation.
- Spod-plugin-agera: The "SPOD Agora" plugin enables discussion within public rooms called Agora (an ancient greek word for "gathering place").
- Auth-server-2: It provides an authorization server via OAuth2 and identity server via OpenID Connect.
- Spod-core: A fork of Oxwall adapted for SPOD
- Spod-plugin-blog: Blogs plugin for SPOD. User blogs with archives, tags, comments and rates
- Auth-server: An OpenID Server in PHP
- Spod-plugin-privacy: It allows hiding certain pages from guest users.
- Wordpress-datalet: A Wordpress plugin that allows embedding a datalet into a blog post.
- Spod-plugin-agera-exporter: The "SPOD Agora Exporter" allows administrators to export comments and datalets from Agora rooms.

- Spod-plugin-widgets: The "SPOD Widgets" plugin allows administrators to create widgets in SPOD.
- Spod-plugin-notification-system: The "SPOD Notification System" is used by several SPOD plugins for real-time notifications and back-end features.
- Spod-plugin-openwall: The "SPOD Openwall" plugin serves as front page for SPOD.
- Spod-plugin-openid: The "SPOD OpenID" plugin enables users' authentication via an OpenID endpoint.
- Deep-client: It is Javascript library that allows the connection with DEEP and the inclusion of a datalet within an HTML page.
- Spod-plugin-api: The "SPOD API" plugin provides interoperability between SPOD and TET.

A specific section (i.e., the Appendix B) provides details to show the project statistic<sup>3</sup>, showing for each component of the project (e.g., SPOD plugins or DEEP plugins) the number of lines of code of each type (HTML, PHP, XML, JavaScript, etc.).

The tool available for calculating the software metrics for SPOD project is available at <https://plugins.jetbrains.com/idea/plugin/4509-statistic> allowing to see the metrics results for the entire project or to select plugins/modules of the project (e.g., Cocreation, Notification system ODE etc.)

#### 2.2.4 MANAGING THE RELEASES

On GitHub, each release is characterized by the source milestone, the issues closed as shown below:

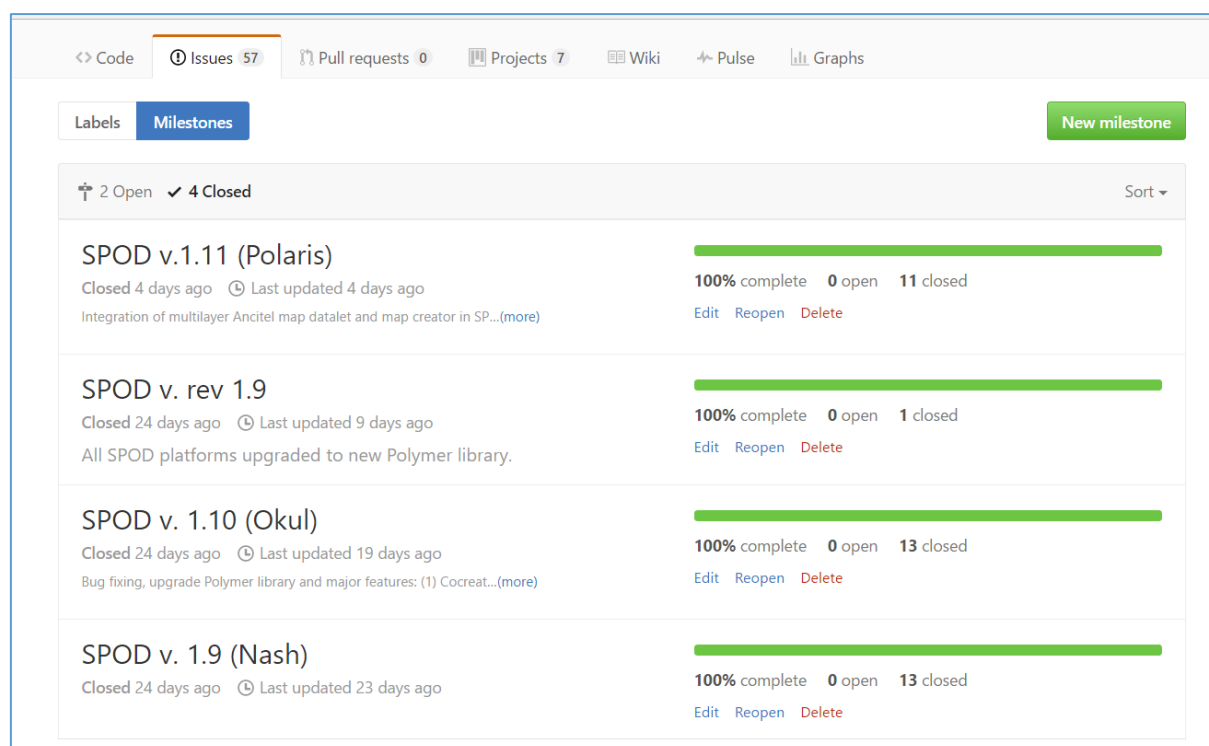


Figure 8: State of the Milestone: Open/Closed

All milestones listed from February 2016 until January 31th, 2017:

Release Name	SPOD platform version	Main release
<i>Furud</i>	1.1	February 05, 2016

<sup>3</sup> A "Statistic for IntelliJIDEA" plugin: <https://plugins.jetbrains.com/idea/plugin/4509-statistic>

<i>Giedi</i>	1.2	March 03, 2016
<i>Heze</i>	1.3	March 23, 2016
<i>Izar</i>	1.4	April 12, 2016
<i>Jabbah</i>	1.5	May 19, 2016
<i>Kuma</i>	1.6	June 24, 2016
<i>Lesath</i>	1.7	July 27, 2016
<i>Mira</i>	1.8	September 12, 2016
<i>Nash</i>	1.9	September 30, 2016
<i>Okul</i>	1.10	November 14 <sup>th</sup> , 2016
<i>Polaris</i>	1.11	December 5, 2016
<i>Regulus</i>	1.12	December 22, 2016
<i>Beta</i>	2.0 Beta	January 31, 2017

All the information about the Release Notes are available at <http://service.routetopa.eu:8000/d/481c8b5f9c/>.

#### 2.2.5 TECHNICAL DOCUMENTATION: THE WIKI PAGE

The technical documentation for SPOD, including detailed information on how to install it, is made available through the GitHub Wiki that is available at the following address: <https://github.com/routetopa/spod/wiki>:

## HOME

Pina Palmieri edited this page 4 days ago · 11 revisions

- SPOD
  - Getting Started
    - SPOD Installation
  - Architecture
    - DEEP
    - Oxwall
    - Plugins
      - Agora
      - Cocreation
      - Notification System
      - Openwall
      - My Space
      - Open ID
      - Agora Exporter
      - SPOD API
      - ODE
      - Widgets
      - TCHAT
      - Blogs
    - Theme
      - SPOD Theme
    - Components
      - Controllets
      - Datalets
  - Release Note
    - 1.9/Nash

► Pages 30

- SPOD
  - Getting Started
    - SPOD Installation
  - Architecture
    - DEEP
    - Oxwall
    - Plugins
      - Agora
      - Cocreation
      - Notification System
      - Openwall
      - My Space
      - Open ID
      - Agora Exporter
      - SPOD API
      - ODE
      - Widgets
      - TCHAT
      - Blogs
    - Theme
      - SPOD Theme
    - Components
      - Controllets
      - Datalets
  - Release Note
    - 1.9/Nash
    - 1.10/Okul
    - 1.11/Polaris

## 2.2.6 EXPLOITATION AND IMPLEMENTATION: CONTAINER TECHNOLOGY (ALL-IN-ONE MACHINE)

To make easy exploitation and dissemination, the introduction of container-based virtualization simplifies the deployment of the whole application suggesting a unique machine (Virtual Machine). This container-based virtualization technology as a lightweight solution allows to the developers to deploy and manage applications with many advantages:

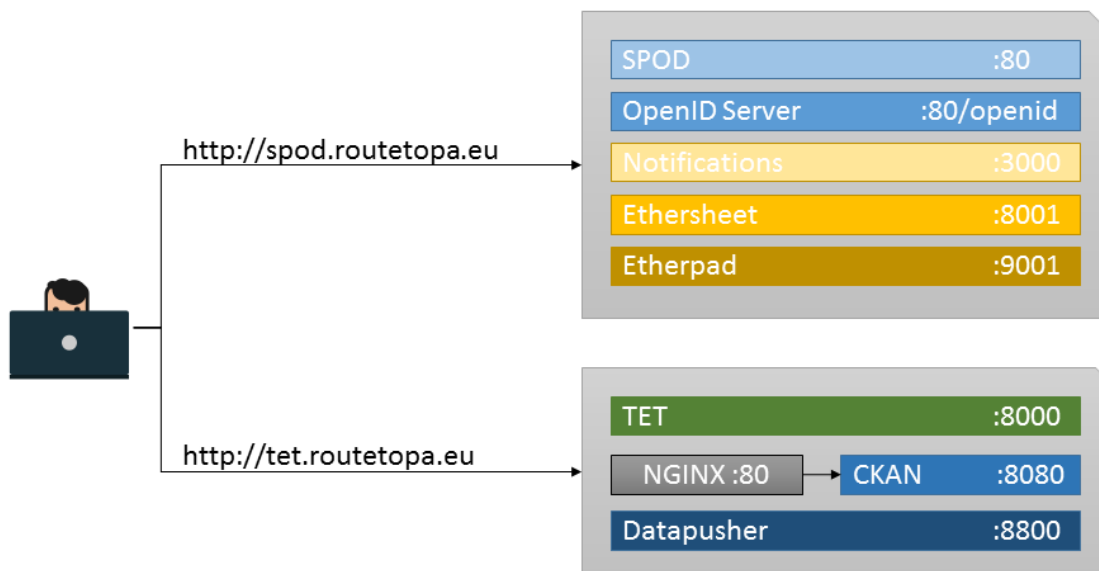
- Low resource consumption (usually half a day of work)
- Portability of container eliminates bugs happens in the running environment
- Lightweight: the developers can run dozens of containers at the same time

This container solution allows by packaging each application with its dependences, configurations files into coherent and self-contained units to help the deployment with few clicks to use in few minutes and it can be moved between different hosts providing a secure and consistent runtime environment between test and production versions of the application.

The advantage using container technology is to avoid to the developers and the administrators to face with challenging web of the compatibility restrictions addressed a specific environment; it provides a common interface for migrating application between environments.

The container-based virtualization for ROUTE-TO-PA project includes SPOD platform, TET-enabled CKAN and Wordpress.

The figure below depicts a “standard” SPOD & TET installation. The platform spans on two different hosts, which is made evident to the user by the fact that he will reach SPOD using a URL (i.e. <http://spod.routetopa.eu>) and TET using a different URL (i.e. <http://tet.routetopa.eu>).



The SPOD machine hosts the following software and services:

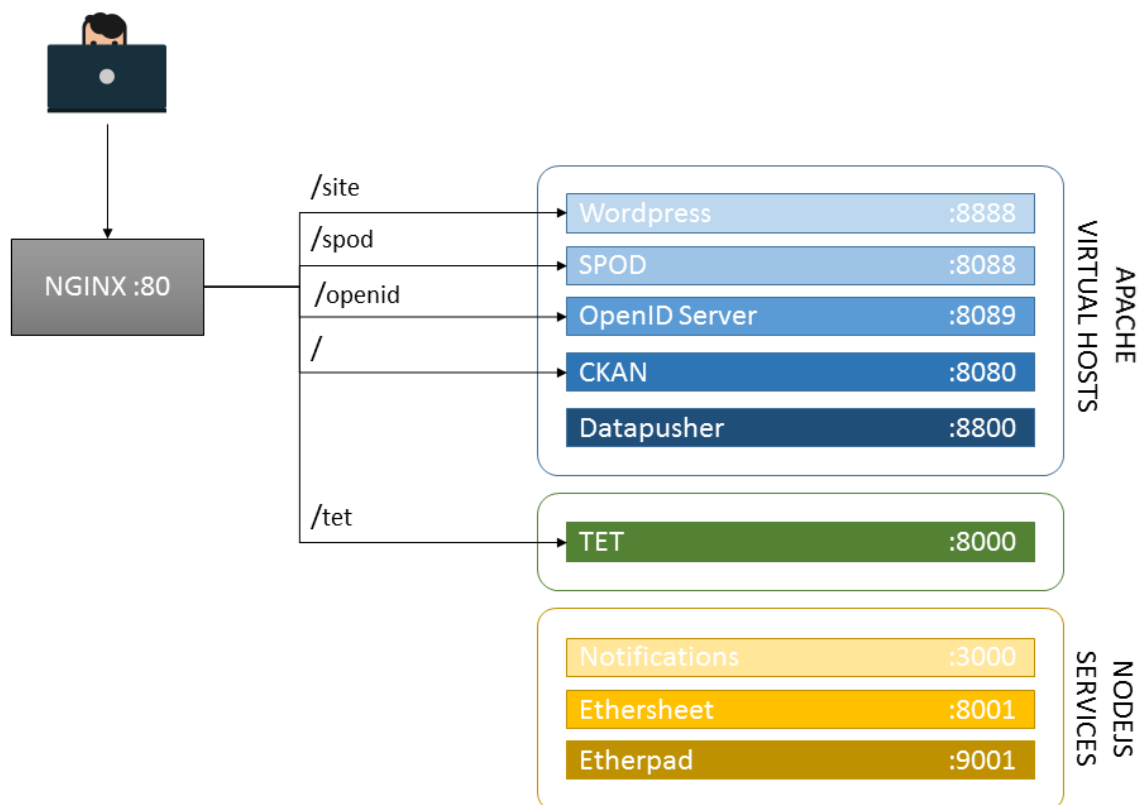
- SPOD on port 80 (the standard web port);
- The authentication server on port 80 under “openid” directory;
- Notifications service: enables real-time notification and updates in Agora, chat and co-creation;
- Ethersheet service: a component for data co-creation;

- Etherpad service: a component for knowledge co-creation.

The TET machine hosts the following software and services:

- TET on port 8000;
- CKAN on port 8080;
- Datapusher service: used by CKAN.
- NGINX on port 80: allows user to access CKAN (on TET) without having to remember and type the port number).

The next figure pictures the “All-in-one” appliance:



As one can see, there is a single URL that the user will type to reach the platforms. The NGINX server will intercept the request and forward it to the right service by analysing the URL typed by the user. Let's assume the All-in-one machine is deployed to an hypothetical <http://platform.routetopa.eu> URL.

NGINX recognized the following URL patterns:

- <http://platform.routetopa.eu/site>: forwards the request to Wordpress;
- <http://platform.routetopa.eu/spod>: forwards the request to SPOD;
- <http://platform.routetopa.eu/tet>: forwards the request to TET;
- All other requests are forwarded to CKAN.

CKAN can be configured so that requests made on “home” page of the platform are redirected to one of the other services. For instance, one could set-up CKAN so that when the user reaches <http://platform.routetopa.eu>, he is automatically redirected to the Wordpress site.

Note that the port number is never explicit: requests are all made on port 80 and that services not meant to be invoked manually by users are not accessible via the port 80. That is the case of Datapusher, Notifications,

Ethersheet and Etherpad. These services, as well the others behind NGINX, are accessible on their own listening post (8800 for Datapusher, 8000 for TET, and so on).

The All-in-one appliance also demonstrates that using NGINX as “reverse proxy” it is easy to deploy services on the machine. On the standard SPOD & TET installation, for instance, it is difficult to install SPOD and Wordpress on the same machine.

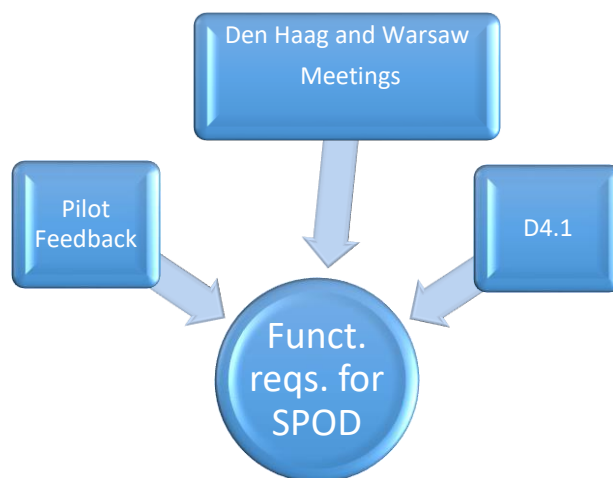
## 3 REQUIREMENTS

---

In this section, we describe **only** the additional requirements with respect to the section 3 of Deliverable D4.1.

### 3.1 FUNCTIONAL REQUIREMENTS

The realization of the Beta version of SPOD (WP4) was achieved considering as input the analysing requirements of the Deliverable D.4.1 "Alpha Version of SPOD", Pilot Feedback, Den Haag Meeting in February and Warsaw meetings in May 2016.



An overview of the functional requirements that were taken into account, grouped in categories, are described by the table below.

Requirement	Notes	DL 2.4	DL 3.1	DL 4.3	DoW	Pilot Feedback
Events						
Discuss about an off-line event	An event is added, with discussions and relevant datasets. Already planned for Beta	UC26: Add an Event	6.3 Activity System. SPOD offers links to off-line events, relevant to the community			
Collaborative space						

Small-group collaboration	Support to small, private (by invitation) group activities, including tools specifically designed for small groups. Already planned for Beta.		6.2. Organizational Level. SPOD, in the initial level (Collaborative Space) facilitates discussions around the meaning of data, and then, in the public room enables participation and collaboration		Small-group collaboration	
Blogs						
	Enable various groups of users to discuss remotely area of interest in different ways					Prato Pilot

More details are found in the specific sections where the functions are described. More info about the meetings (notes, presentations, etc.) can be found at <http://service.routetopa.eu:8000/d/481c8b5f9c/>.

## 3.2 NON-FUNCTIONAL REQUIREMENTS

### 3.2.1 PERFORMANCE

Cache management solution have been addressed to improves performance on SPOD platform

### 3.2.2 PRIVACY

Specific permission can be applied for settings to any page of the platform, so the access can be more restrictive at every level:

- *Public*: everyone view edit page
- *Protected*: every one can view pages, but only members can edit
- *Private*: only member can view or edit page

This kind of setting can be applied to all the pages in the SPOD and can be configured on site by the administrator. This is particularly useful as different policies are applied in different cases by the pilots, and SPOD can be accordingly configured.

## 4 DESIGN PRINCIPLES AND TECHNOLOGY

---

### 4.1 BASE SOCIAL PLATFORM: WHY OXWALL AND NOT ELGG

*This section is specifically meant to explain the choice of Oxwall with respect to Elgg as basis for our development of SPOD. Although in the DoW, the mentioned FOSS server for Social Network was Elgg, we decided to refine and review the choice, considering the time passed from the original design of the proposal to the effective implementation. Some presentation of arguments supporting our decision was present in the D4.1 in section 3.3.1. Here we refine and integrate the arguments.*

We have selected a set of essential feature that the social platform must have to fits all our needs.



The platform must be:

- FOSS (Free and Open Source Software)
- Licensed with a ROUTE-TO-PA compatible license.

The core code of the platform, must be easily readable and well commented, in order to decrease the amount of time necessary for custom modification. We looked to an active and popular project. Popular project has large community of developer that increase the quality of code, perform large tests, solve bugs and keep the project lined up with the last technology. The platform must have a plug-in architecture for rapid functionality extension, moreover the plug-in creation and deploy must be free and easy. Since SPOD will be publicly available in several municipalities, we need a platform that potentially can handle hundreds of simultaneous user on a common server. One of the main goal of the project is to be available for a large audience (of people and devices) we need a responsive UI capable to adapt to all kind of operative system, browsers, screen-resolution. From a technical point of view the platform must have a well-structured layer for data manipulation and persistency, an ORM (Object-relational mapping) capable of communication with the most common DBMS.

Finally, the platform technology stack must be easily deployable or must correspond to a common server configuration, in order to be easily and economically hosted on a server host.

At the beginning of ROUTE-TO-PA project, only two projects meet all the selected requirements: Elgg and Oxwall. Both Elgg and Oxwall are web based MVC project based on LAMP (or WAMP) stack. From a technological point of view the two projects are similar.

Here we will present the motivation that led us to choose Oxwall instead of Elgg.

All the motivation are technical and code related.

In particular comparing code structure you can see that Elgg has a mixed approach for its codebase Oxwall, instead, has a really strong structured approach. Elgg core is written with an Object Oriented paradigm, while plug-in developer can choose between an Object Oriented and Procedural approach. This particularity of Elgg delegate to the plug-in developer the selection of coding style, this could lead to low quality code or mixed approach in plug-in development and the entire platform could suffer in terms of maintenance.

Another important aspect in favour of Oxwall is the presence of a template engine. Smarty is the template engine of Oxwall that allows an easier software layer separation inside the MVC paradigm. Smarty is based on master/slave page mechanism. A plug-in view, can inherit the general layout from a master page and implement specific part of the view for plug-in related functionalities. This feature permits to create a homogeneous layout through the entire platform, and the reuse of master page code in all the plug-ins. Moreover the developer has a set of built-in tools provided by Smarty (like form elements, list, table, UI controller) that can be used to create the plug-in view. These tools are declared in PHP by the plug-in developer and injected in the Smarty template view that, automatically, create the HTML/CSS/Javascript code to handle and style the element. The developer can ignore the layout and behaviour of a component (like input form validation), because is handled by the template engine. If we want to change the entire layout of the platform we only have to reengineering the master page and, eventually, the built-in tools style and with no code change in the plug-ins the entire platform will change aspect. This enormously increase the personalization capabilities of the platform (since SPOD has multiple installation, with some differences) minimizing the developing effort. Moreover, Smarty increases the platform overall performances due to views cache mechanism. The first time that a SPOD user sees a page Smarty create a simple HTML/CSS/Javascript representation of that page and provide this “compiled” page for all the next page view (only the dynamic content is calculated on every page view). This provides an increase of platform performance since once the page is “compiled”, only the static page is provided by the server. With page caching, PHP isn’t involved in page creation, so the server load is very low and restricted only to static page dispatch. Web servers like Apache or Engix are optimized for static content dispatch so the cache mechanism of Smarty fully exploits the web server capacity, providing better user experience and large simultaneous use of SPOD.

In Elgg view layer, instead, is based on plain PHP/HTML/CSS, this approach delegates to the programmer all the complexity of software layer separation.

Poor software design choices in plug-in implementation can easily lead mixing up MVC layers and, consequently, lead to poor quality code in Elgg.

Oxwall, lastly, has a built-in theme for mobile devices, and every plugin can handle special views for mobile device. Elgg doesn't provide a mobile specific theme, and we should have write an ex-novo platform theme for mobile (this is an intensive time-consuming task).

The table below summarizes the main criteria to choose as social platform, Oxwall.

Criteria	Elgg	Oxwall	✓	✗
<i>Code structure</i>	Mixed approach ( ✗ ) <ul style="list-style-type: none"> <li>Core Source with Object Oriented paradigm</li> <li>plug-in implementation (Object Oriented or a Procedural approach)</li> </ul>	Structured approach ( ✓ ) <ul style="list-style-type: none"> <li>core and plugin implementation are Object oriented</li> </ul>	Code maintenance and quality	<ul style="list-style-type: none"> <li>Low quality Code</li> <li>Problem for maintenance</li> </ul>
<i>Template engine</i>	View layer is based on plain PHP/HTML/CSS ( ✗ )	Smarty as PHP template engine ( ✓ )	An easier software layer separation and reuse of viewing code	Complexity of software layers separation
<i>Mobile friendly</i>	Write ex-novo theme for mobile devices ( ✗ )	Available Built-in theme for mobile devices and View for mobile device ( ✓ )	Available the support of its in-house Mobile Browser Version, optimized for performance in major mobile browsers	To write an ex-novo theme for mobile devices

Table 4: Main criteria to choose Oxwall instead of Elggs

## 4.2 REUSE RATHER THAN REINVENT: THE COLLABORATION PLATFORMS ETHERDOCS AND ETHERSHEETS

### 4.2.1 OVERVIEW

According to user requirements, one of the most important issue about the Cocreation room is to guarantee a real-time mechanics to allow users to edit at the same time, some content in a text editor, in the knowledge room, and in the cells of spreadsheet editor, in the data room. These requirements mean that it is necessary to choose whether to use some existing solutions or to implement each of editors from scratch.

Considering that SPOD is itself a very complex software to be maintained and expanded, composed by a lot of components that interact with each other, and that to build up a new collaborative real-time editor will takes a great amount of time in design, implementation and tested, a reasonable way to face on this issue is to take an existing software, or third-party libraries.

Of course, each potential choice was made considering that, for both text and spreadsheet editor:

- We need to customize the user interface to fit the SPOD look and feel
- We need to manage the user profile to align the information about the editor user with SPOD ones

- We need to expand the standard features to fit them according to the future user feedbacks and new SPOD requirements
- We need to interact with editor via API to get information and contents about the documents that we will use in some components or sections of SPOD
- We need to choose a relative limited set of functions to provide an agile, well fitted with requirements user interface. Possibly we don't to provide a complex interface with great number of functions that could be unnecessary for the final users
- We would maintain, locally, control of the servers and the contents created by users
- We need an Open source solution
- Since the final administrators will be some people from Public Administration who may not have a great computer science skills, we need a backend easy to manage in terms of installation and maintenance
- We need a well-maintained software with a relatively big community

#### 4.2.2 POSSIBLE EXISTING SOLUTIONS

This section describes analysed collaborative tools, i.e., web-based applications that allow for the co-edit of document documents in real time and store them on centralized servers, spreadsheet and all features above described.

##### 4.2.2.1 GOOGLE DOCS

The first prototype of Cocreation room was made with Google Docs to provide a collaborative text and spreadsheet editors. Based on pilots feedback we decide to discard this solution because:

- We don't have any control on server side
- The interface provides a great set of functions but a lot of them are unnecessary for our users, resulting in too much complex GUI
- We cannot customize the look and feel
- The users content reside on Google servers
- We cannot expand any functions provided by the editors
- Any modifications, even in the APIs, require a modification of SPOD code
- We cannot manage user profiles. It requires that all SPOD users should have a Google account

##### 4.2.2.2 FIREPAD AS TEXT EDITOR

Firepad allows synchronous document (with CodeMirror) and code (via ACE) co-editing. Firepad requires no server-side code. You can easily embed it into any web app using the available JavaScript files, so you can extend your web application capabilities by adding the simple document and code editor.

Due to Firepad uses a cloud approach to managing the contents and since it's based on well know open source software Etherpad, we decide to discard this solution.

##### 4.2.2.3 ONLYOFFICE

OnlyOffice is a multifunctional online office suite that features text, spreadsheet, and presentation editors working within a browser. It includes features similar to MS Office desktop editors, but also lets you to co-edit and comment on documents in real time. You can install it standalone and integrate with any application through API or together with OnlyOffice's collaborative system that offers additional possibilities for document management. OnlyOffice seems to be the most complete solution but it has some disadvantages:

- It is really complex to install and manage
- It requires a dedicated machines because it's a very big project
- It requires a great amount of time to be extend with new features
- It is not simple to manage user profiles in order to align them with SPOD users
- Provide a lot of functions, similar to Microsoft Office and Google Docs, but a lot of them are unnecessary for our users.

#### 4.2.2.4 OWNCLOUD DOCUMENT

OwnCloud Documents is a part of file sync-and-share server ownCloud that lets up to five people collaborate on rich text documents in .odt, .doc, or .docx formats within the browser. You can share your document inside ownCloud or use the public link, setting up an expiration date and password, if needed.

Since it doesn't have support for a lot of amount of users and, furthermore, has the same problems of the previously mentioned software, we decide to discard it.

#### 4.2.2.5 ETHERCALC, ETHERGRID AND ZOHOSHEET

Each of this software provides a real-time collaborative spreadsheet based on cloud approach to managing the contents. For each it is not possible to control the server side and have the same problems discussed in the previously sections, so we decide to discard them.

### Overview

A summary below table lists features of each collaborative tool discussed in the previously sections, excluding the EtherCalc, Ethergrid and Zohosheet tools for the motivation explained in the previous section:

Features\SW	Etherpad	Google docs	Firepad	OnlyOffice	OwnCloud
No Registration required	✓	✗	✓	✓	✓
Easy installation and management	✓	✗	✗	✗	✗
Available features not useful	✓	✗	✗	✗	✗
Easy management user profiles	✓	✗	✗	✗	✗
Privacy (e.g. storing your data on a third parties servers)	✓	✗	✓	✓	✓
Self-hosted	✓	✗	✓	✓	✓

#### 4.2.3 SPOD COLLABORATION PLATFORM

For SPOD collaboration platform we decided to use a well known solution that fit with ours constrains: Etherpad as a collaborative text editor and Ethersheet for collaborative spreadsheet editor.

Etherpad is collaborative real time open source document editor maintained by the Etherpad Community.

It's written in JavaScript (99.9%), on both the server (nodejs) and client. It is designed to be easily embeddable and provide support to manage user profiles and groups even through HTTP API. It has a full-featured plugin framework that allows the administrator to select the features needed for to final users.

By default, your Etherpad is rather sparse and because Etherpad takes a lot of its inspiration from Wordpress plugins are really easy to install and update. Once you have Etherpad installed, you should visit the plugin page, placed in the administrator section, and take control. Finally, Etherpad comes with translations in most languages! Users are automatically delivered the correct language for their local settings.

EtherSheet is a collaborative, real-time open source spreadsheet. It allows people to collaborate on spreadsheets over the web in real time. It's written mainly in JavaScript on both the server (*nodejs*) and client side.

The two software are totally open source and they have a well-maintained Github repository and a good base code we can use as a starting point for face on our needs. In particular, they have:

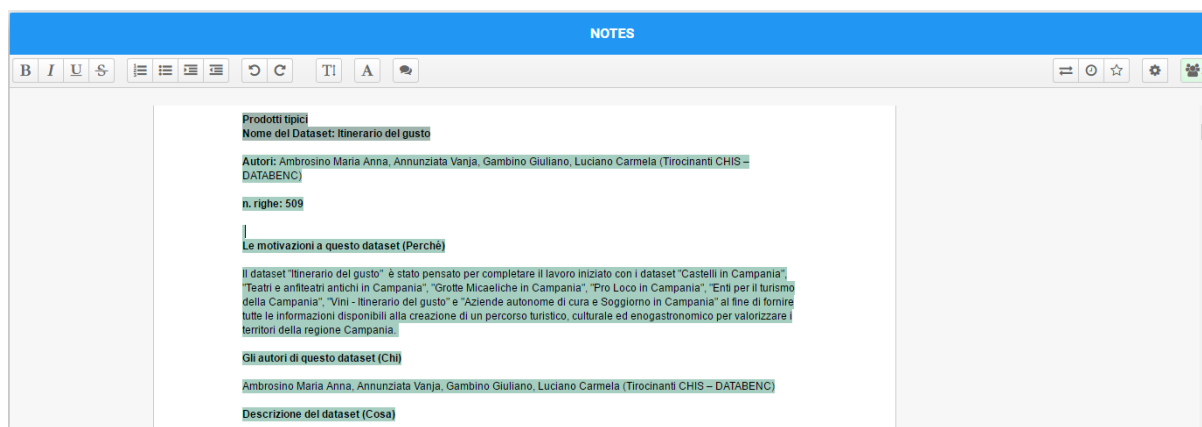
- a plugin base set of functionality we can use to provide new features in a structured way
- a way to align the editor users to SPOD ones
- a server side based on Nodejs technology we can modify in a simple way
- a small, but useful, set of features we can use as starting point
- a local repository to manage the users content
- a set of APIs to interact with contents from code
- customizable user interface with style sheet

In order to make this software able to integrated in SPOD according to the required features, a lot of modifications has been done in the core section and existing functionality and furthermore a new set a completely new tools has been provided, resulting in the creation of SPOD version of Ethersheet and Etherpad.

#### 4.2.3.1 CUSTOM MODIFICATION TO ETHERPAD

The Etherpad instance that runs on SPOD is pretty similar to the current stable version available on Github but it has some custom additional modifications:

- Custom installation is provided to make easily integration of core component, style and selected plugins with related support third-party library
- Etherpad as kernel service to improve stability, manage the fault, and allow to web administrator to start and stop it.
- Complete restyling of user interface to fix the SPOD look and feel
- Plugins selection and adaptation of third-party library in order to fit software compatibility issue with other server parts of SPOD.
- User profiles management to create one to one relation from SPOD user to Etherpad user



#### 4.2.3.2 ETHERSHEET

By default, Ethesheet provides a very simple set of functions on both server and client side so it has been necessary a lot of change to fit it with SPOD requirements. These changes involve the creation of new features do not implement in the default stable version of Ethersheet and the modifications to the core components of the software. In particular:

- Custom installation is provided to make easily integration of core component, style and the related support third-party library
- Ethersheet as kernel service to improve stability, manage the fault, and allow to web administrator to start and stop it.
- Complete restyling of user interface to fix the SPOD look and feel

DATASET									
	Vino	Denominazione	Colore	Tipologia	Produttore	Descrizione Organolettica	Vitigno	Gradazione Alcolica	Certifica
1	Vino	Denominazione	Colore	Tipologia	Produttore	Descrizione Organolettica	Vitigno	Gradazione Alcolica	Certifica
2					Torre Gala				
3	Trama	Beneventano	Bianco	Falanghina	Terre Stregate	Colore paglierino carico, profumo intenso con spiccate note floreali	100% Falanghina		I.G.T.
4	Genius Loci	Sannio	Bianco	Fiano	Terre Stregate	Colore paglierino carico, con spiccati sentori floreali, intensi e persistenti	100% Fiano		D.O.C.
5	Svelato	Falanghina del Sannio	Bianco	Falanghina	Terre Stregate	Colore paglierino con riflessi verdolini, ricco di sentori fragranti e floreali	100% Falanghina		D.O.C.
6	Svelato sur lies	Falanghina del Sannio	Bianco	Falanghina	Terre Stregate	Colore paglierino con riflessi verdolini, ricco di sentori fragranti e floreali	100% Falanghina		D.O.C.
7	Aurora	Sannio	Bianco	Greco	Terre Stregate	Colore giallo paglierino con intensi riflessi verdolini	100% Greco		D.O.C.
8	Fabula	Falanghina del Sannio	Bianco	Falanghina	Terre Stregate	Colore paglierino, con riflessi verdolini, ricco di sentori fragranti e floreali	100% Falanghina		D.O.C.
9	Idillio Cantine Iacobucci	Beneventano	Rosso	Aglanico	Terre Stregate	Colore rosso rubino, ricco di sentori fragranti e fruttati	100% Aglianico		I.G.P.
10	Manent	Beneventano	Rosso	Aglanico	Terre Stregate	Colore rosso rubino, ricco di sentori fragranti e fruttati	100% Aglianico		I.G.T.

Based on users feedback:

- Redefinition of the concurrency mechanism to allow the users to edit the spreadsheet at the same time: this is a modification to the core since the default version is able only to send the changes to the other users but it allows to edit only one cell at same time for all users, resulting in lost of focus on the cell every time one more user try to edit another one.
- Use the first row as header to maintain the header information during the vertical scrolling

DATASET									
	Vino	Denominazione	Colore	Tipologia	Produttore	Descrizione Organolettica	Vitigno	Gradazione	Certifica
30	Fuodo di Nevera	Castelli Presigri	Bianco	Falanghina	Terranera	Freschezza armonica.	Falanghina	13%	
31	Fiano di Avellino - Feudo di Nevera	Fiano di Avellino	Bianco	Fiano di Avellino	Cantine Terranera	Paglierino con riflessi verdognoli. Odore fruttato, delicato e caratteristico. Gusto vellutato, armonico e pieno.	100% Fiano	13%	D.O.P.

- Redefinition of the import / export to csv to fix bugs present in version default and fit to specific needs:
  - import / export from Microsoft Excel
  - manage big csv file
  - improve performance
  - manage csv with different separator
- Update of the communication layer with last version of third-party library, resulting in several modification in the *base code*
- The addition of a mechanism, based on a debouncing timer queue, to communicate with external components global changes to the spreadsheet: we use this to update the visualizations related to the dataset in the data cocreation room
- The addition of Copy and paste and arrow keys feature:

- The user can move in the sheet by using the arrow keys. The selected cell will be highlighted. By pressing the enter key he can edit the cell content and with the esc key can return to selection feature.
- The user can select multiple cells by using at same time the shift and the arrow keys. The selected portion of the sheet will be highlighted.
- After selection the portion of the sheet, the user can use the arrow keys to select the destination cell to paste (ctrl + v) the content previously copied(ctrl + c). If the content copied overflows the destination boundaries it will be truncated to fit the destination space.
- The user can copy/paste some content from Microsoft Excel or from another sheet in a different room.
- The user can use the arrow keys as well as the mouse to select the content to copy and paste.

DATASET									
	Vino	Denominazione	Colore	Tipologia	Produttore	Descrizione Organolettica	Vitigno	Gradazione	Certifica
11	Manent Rosé	Beneventano	Rosato	Aglianico	Terre Stregate	Colore rosa cerasuolo, con sentori fragranti e floreali	100% Aglianico	alcuna	I.G.T.
12	Idillio	Sannio	Rosso	Aglianico	Terre Stregate	Colore rosso rubino, ricco di sentori fragranti e fruttati	100% Aglianico		D.O.C.
13	Scrypta	Sannio Guardia Sanframondi	Rosso	Aglianico	Terre Stregate	Colore rosso rubino, ricco di profumi di mora e lamponi	100% Aglianico		D.O.C.
14	Arcano	Sannio Guardia Sanframondi	Rosso	Aglianico Riserva	Terre Stregate	Colore rosso rubino, con profumi intensi e persistenti di frutta matura con spiccate note balsamiche	100% Aglianico		D.O.C.
15	Malaca	Falanghina del Sannio	Bianco	Falanghina Passito	Terre Stregate	Di colore giallo oro con un importante residuo zuccherino mai stucchevole. I profumi ricordano la frutta essicata ed il confetto, oltre a sensazioni di frutta tropicale matura	100% Falanghina		I.G.T.
16	Spumante Falanghina	Falanghina del Sannio Spumante	Bianco	Brut	Terre Stregate	Colore trasparente con riflessi verdolini. Al palato è elegante, fresco, sapido con buona acidità. Il perlage è fine e persistente.	100% Falanghina		D.O.P.

- The addition of the contextual menu on each cell to add geographic information in SPOD such as a point related to a specific location or shapes to highlight some area on a map. The system is able to save and manage geographic information in GeoJson format.

DATASET							
Id	Supporti per l	Vincolato	Tipo di Vincolo	Coordinate	Affidabilità	DD	EE
524				40.668480545282364, 14.72260370850563	Si		
525							
526							
527							
528							
529							
530							
531							
532							
533							

Aggiungi coordinate geografiche

**Masseria la Cicerella**  
 Erchie  
 Apulia Italy

**Torre La Cerniola**  
 Maiori  
 Campania Italy

**Rue de la Chapelle**  
 57720 Erching  
 Lorraine France

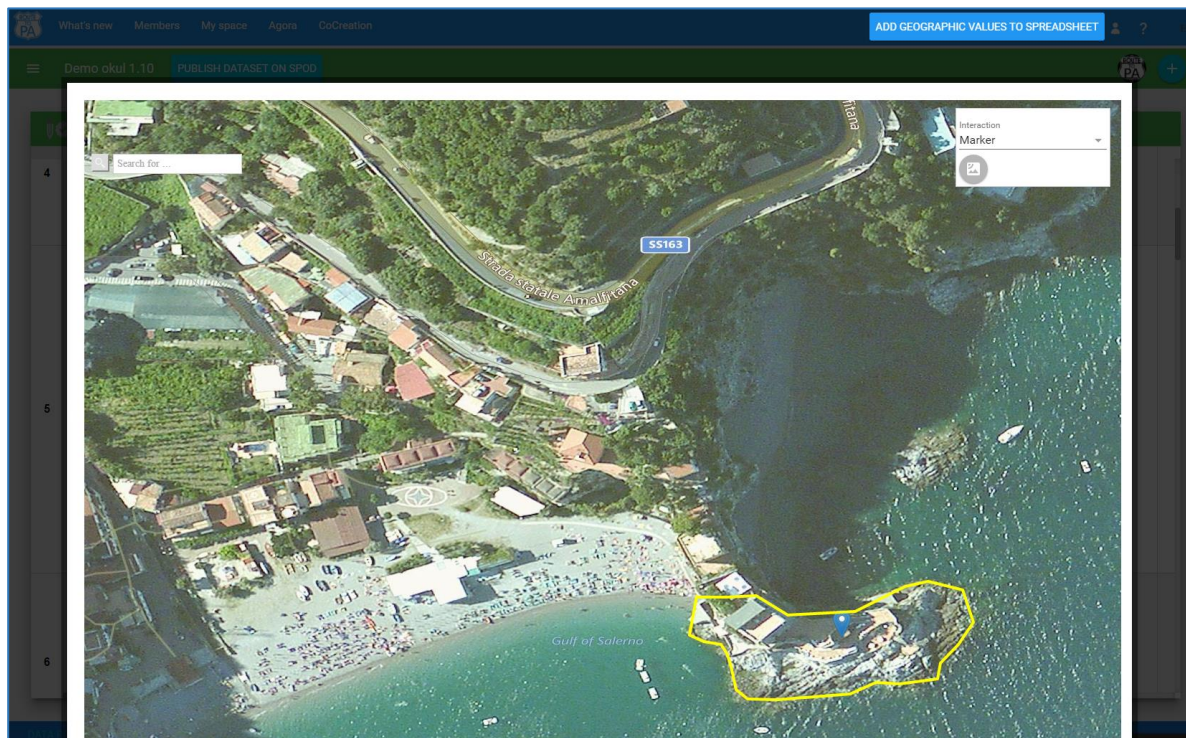
**Rue de la Montagne**  
 57720 Erching  
 Lorraine France

**Rue de la Grotte**  
 59169 Erchin  
 Nord-Pas-de-Calais France

Interaction

Marker





DATASET						
	Long	Lat	Categoria	Biblioteca	E	F
1	Long	Lat	Categoria	Biblioteca		
2	11.09269343...	43.88009381...	Biblioteca	Biblioteca dell'Archivio Fotografico Toscano (A.F.T.)		
3	11.10893419...	43.86103285...	Biblioteca	Biblioteca del Centro per l'Arte Contemporanea "L. Pecci" CID/Arti		
4	11.11821445...	43.87049253...	Biblioteca	Punto lettura Circostrizione Est		
5	11.08224767...	43.87869929...	Biblioteca	Biblioteca di Scienze Geoambientali e Planetarie (ex IGT)		
6				Istituto culturale e di		

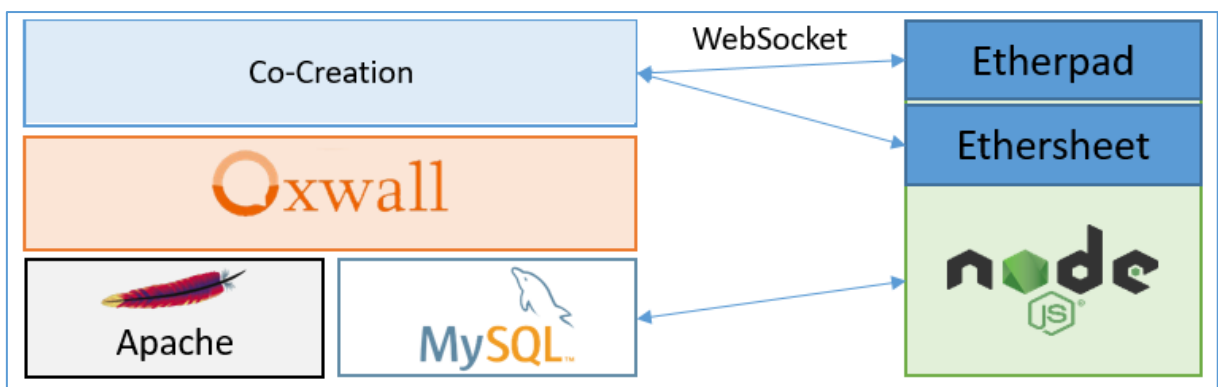


We proceeded to make a pull request to the Ethersheet community to create a SPOD version of Ethersheet as a spin-off product of ROUTE-TO-PA.

## 5 ARCHITECTURE DESIGN OF SPOD

### 5.1 THE ORGANIZATION OF COCREATION SERVERS

Due to the real-time nature of cocreation communication, we need specialized servers for Etherpad and Ethersheet. Both servers are written in Node.js and exploit the mechanism of web-socket for real-time client-server communication. Servers are installed as UNIX services and can be deployed on the same machine that hosts the web-server of SPOD since every server listens on a specific TCP port. Etherpad and Ethersheet servers' persistence layer is based on MySQL DBMS that store specific schema for Etherpad and Ethersheet. These schemas are different from SPOD database so can be deployed on dedicated MySQL server. In order to reduce the SPOD deploy difficulties we have tested and made available a mechanism that allows the cohabitation of all the databases in the same MySQL installation. Moreover, we have developed a series of shell scripts that allows a simple and complete installation of the servers from scratch with all the needed dependencies. The servers/service have an automatic respawn mechanism that monitors the server status and respawn it in the case of failure or machine reboot. This mechanism logs the error and respawn the server in order to have the maximum of uptime service.



## 5.2 NOTIFICATION COMPONENTS

Oxwall does not have a mechanism for real-time notification/message exchange in the discussion (agoras and cocreation rooms), so SPOD users don't know if someone added a comment to the discussions they are contributing. Only on page refresh, they can view the latest added comment/s. This limitation has been overcome by adding a real-time notification plugin. This plugin has two main parts, a JavaScript client browser-side and a Node.js server on server-side. Every time a user opens an agora room or a cocreation room automatically registers itself on the notification server for that room. When a user submits a message inside the room a notice (containing all the message details) is sent from JavaScript client to the Node.js server through a web-socket. The Node.js server forwards the message to all the users registered in that room, and the real-time notification plugin shows it to the user. This mechanism allows a real-time communication and a better experience for the user. Moreover, it reduces the server-side computation since the server forwards only a small message and not the full page for the single message. Notification server is installed as Unix service and can be deployed on the same machine that hosts the web-server of SPOD. The servers/service have an automatic respawn mechanism that monitors the server status and respawns it in the case of failure or machine reboot. We have modified the SPOD architecture presented in the D4.1 section 4.1 including the notification, Etherpad and Ethersheet servers as shown in Figure 9.

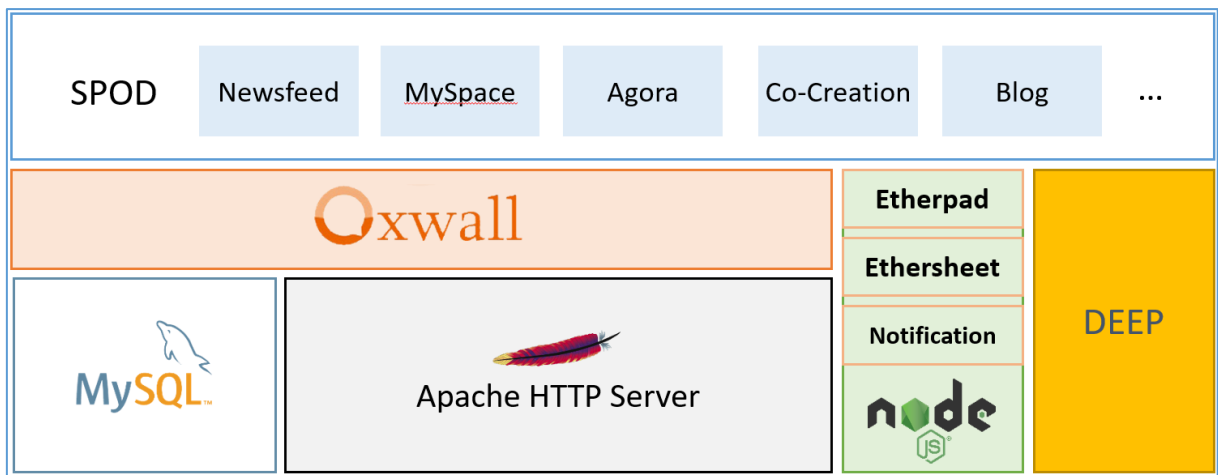


Figure 9: SPOD Architecture

## 6 SYSTEM DESIGN OF SPOD

In this section we include the novel features of the SPOD Beta, based on the feedback from users, the researchers meetings, the DoW and the functional and non-functional requirements specified in this document and in the Deliverable D4.1 SPOD Alpha.

### 6.1 COLLABORATIVE SPACE

During the second year of activity, the most important development of SPOD was the design and implementation of a collaborative space that allows citizens to collaborate closely to co-create together, based on Open Data. The requirements for this set of features was mainly developed at the Den Haag research meeting and Warsaw plenary meeting (minutes and presentations of the meeting are available at <http://service.routetopa.eu:8000/d/481c8b5f9c/>). The result was also published in a joint publication “Increasing Public Value through Co-Creation of Open Knowledge” at the International Conference on eDemocracy & eGovernment (ICEDEG 2017) (See Sec. 0).

In the design of SPOD platform, the aim is to study how to increase Public Value [2] for the citizen, available in the knowledge, through the collective participation, involving Public Administrations, stakeholders and citizens together. The goal is to link the Public Value to the Data Information-knowledge (DIK) hierarchy [3], which is drawn as a pyramid. A new process has been described to increase Public Value for citizen by proposing a variation of the hierarchy consisting of Data-Information-Knowledge.

The pyramid showed in Figure 10, shows the classic Data-Information-Knowledge Hierarchy, changed to host open data, open information and open knowledge. At the base level, there are the open data, processed data become information and subsequently knowledge. It essentially depicts processes to transform raw open data (i.e., tabular datasets) from the low level to the higher levels.

The DIK pyramid can serve also in the context of Public Value creation by the Public Administrations, with increasing Public Value as they go from bottom to top. In fact, public institutions and governments publish Open

Data [4] on their portals in favour of transparency. Open Data are raw data; they are often as simple as plain tables or scanned documents. In order to generate information, citizens could need data that are not actually available as they are not open or government did not release them. This opens the need for the citizens to produce, also collectively datasets.

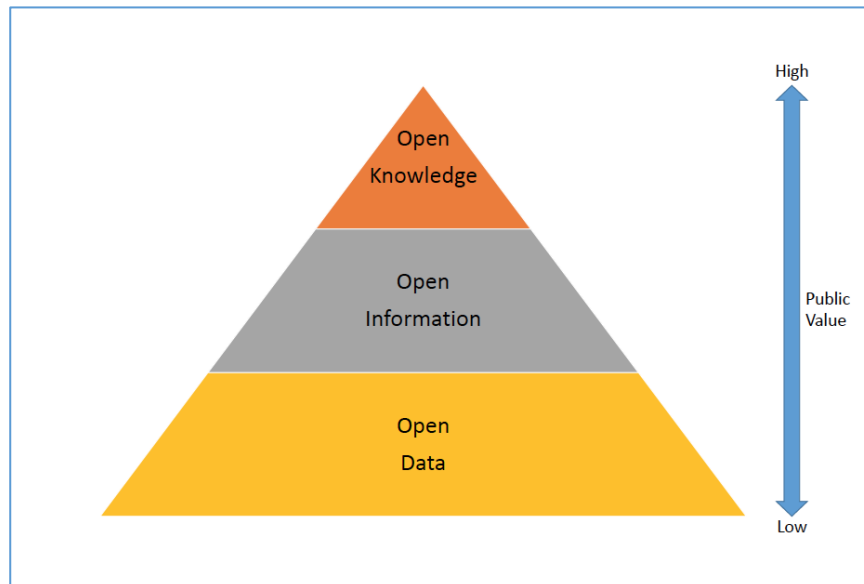


Figure 10: Pyramid highlights how not only data are open, but also generated information and gained knowledge.

### Process Overview

SPOD allows communication between citizens and their governments and all functionalities of the platform have been designed to enable discussion among citizens, foster interpretation, and transformation of open data; create data, information and knowledge. SPOD is a Social Platform for Open Data that supports the process of the information-processing pyramid.

Figure 11 depicts the main component of this process that are *Data Co-Creation Rooms* (described in the section), the *Agora* (this description can be found in Deliverable in DL 4.1<sup>4</sup>), the *Knowledge Co-Creation Rooms* (described in section 6.16.1.1), and the *Blog* (described in section 6.1.3).

Two parts have been highlighted in the process: the Social Platform for Open Data (SPOD), depicted within the dashed boundaries (see Figure 11), and what happens outside SPOD. The main SPOD feature is the Agora for the social discussions (shown in the middle of Figure 11), that enables discussions among small, medium and large groups of users that meet in rooms to publicly interpret data, create visualisations of Open Data, and collectively extract information. When needed, groups of citizens have the opportunity, based on their choice, to privately create a room and meet in small groups to co-create a dataset. At some point, they can decide to publish the co-created dataset on the platform, enlarging the audience and allowing other users to have a look at published datasets and exploit them.

Citizens can co-create information or knowledge using the Knowledge Co-Creation room. They can meet in small group to analyse, interpret, and transform data producing collaboratively a document. The knowledge document can be shared with other users within the SPOD platform to stimulate discussions or become part of a blog post outside the SPOD boundaries. Both Co-Created datasets and knowledge documents can be reused by citizens during the public discussions in the Agora rooms or be published outside the SPOD platform generating what we

<sup>4</sup> Deliverable: D4.1 Alpha version Of SPOD (January 31 2016)

call Open Data, when they are published on the government portals, or Open Knowledge when the knowledge documents become public, for instance in a blog post.

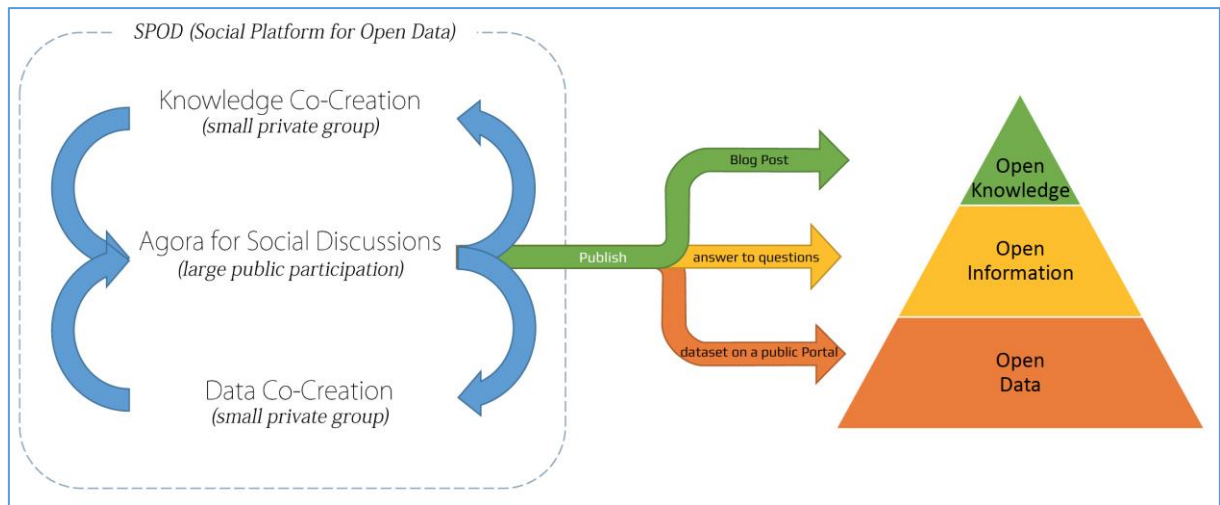


Figure 11: Co-curation activities: tools (i.e., Data Co-Creation, Agora and Knowledge Co-Creation) and process to increase value by interpreting, transforming, co-creating, and discussing data.

#### 6.1.1 KNOWLEDGE ROOM

##### Overview

The Knowledge Co-Creation room is the SPOD space where users can remotely meet and digest a particular topic together. This is the place where data and open datasets become piece of information or knowledge emerges. A SPOD user creates the room, providing the initial sparks to ignite the creation process. Invited users have a shared document where they can collaboratively write, contributing to the knowledge construction. The process is not straightforward, it requires creativity by interpreting, digesting, discerning, integrating, transforming and combining existing data.

The Knowledge Co-Creation has been designed on the principle to be more focused on the joint artefact, which is

not discussion, but some outcome such as a design, a plan, a solution, or some interpretation of a state of affairs, depending on the scenario. This implies that a tool should not simply record any contributions, but should afford some principles of co-creation such as:

- focus on ideas, content and data, in relation to the main issues: not all ideas contribute to the progress of the discussion;
- creative processes can be messy, so linearity is less productive: ideas do not come in a linear fashion, we should find a way to collate them at the sections where they belong, like in a written text rather than a list of contributions;
- reflection is important, and should be stressed by the interface;
- users can collectively add and modify each other's contributions, this is the essence of co-creation;
- less focus on ownership, but activity should be visible;
- the process ends when sufficient agreement on formulations of issues, ideas, and outcomes has been achieved. This requires explicit appreciation of those formulations.

These ideas were the basis for the design of the cocreation room. Shortly, it affords co-creating a text, based on created datalets and their interpretation. Multiple datalets can be created and discussed, and discussions can be

synthesised into a common text. At some point, when users are confident, they can decide to make the document available for all the SPOD users, enabling its reuse within other rooms.

As described, users can collaboratively Co-Create Knowledge documents. The content is initial private, but users can decide to share it in SPOD to be discussed by the community on SPOD. A further step is to open the Knowledge to a wider public, publishing it over the Internet. SPOD has the Blog, which can show the Knowledge as public to all Internet users as result of interactions within the platform, co-creations and discussions. It is crucial that during the activity, some of the knowledge of the participants becomes transparent; this is what we mean by sharing of knowledge. For the purposes of the current project, the technology should help to realise transparency that is more effective; there must be receptors capable of processing, digesting and using the information/open data.

### Knowledge room on SPOD

Knowledge room includes tools specifically designed for small groups such as:

- template *real-time collaborative documents* for each of the three phases *explore, ideas, outcome*
- a set of datalets, that can be annotated via post-it, to support the process of knowledge creation with visualization of open data
- a dataset library that collect a set of dataset useful for the process of knowledge creation
- a synchronized mechanism to notify to all room members when someone creates a new datalet, adds a new dataset, adds some post-it to a datalet and so on.

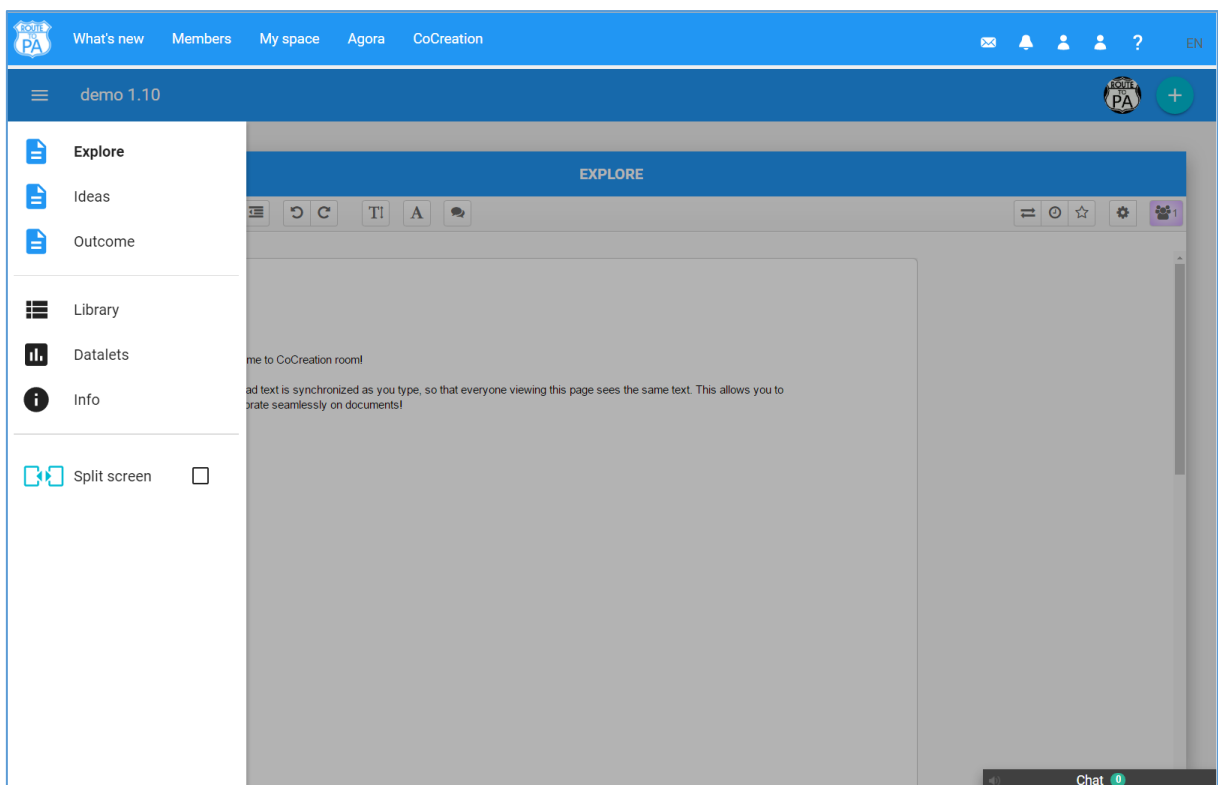


Figure 12: All features (show Explore, Ideas, Outcome, see Library, Add datalet, show Info, modify UI with Split screen) in a knowledge room



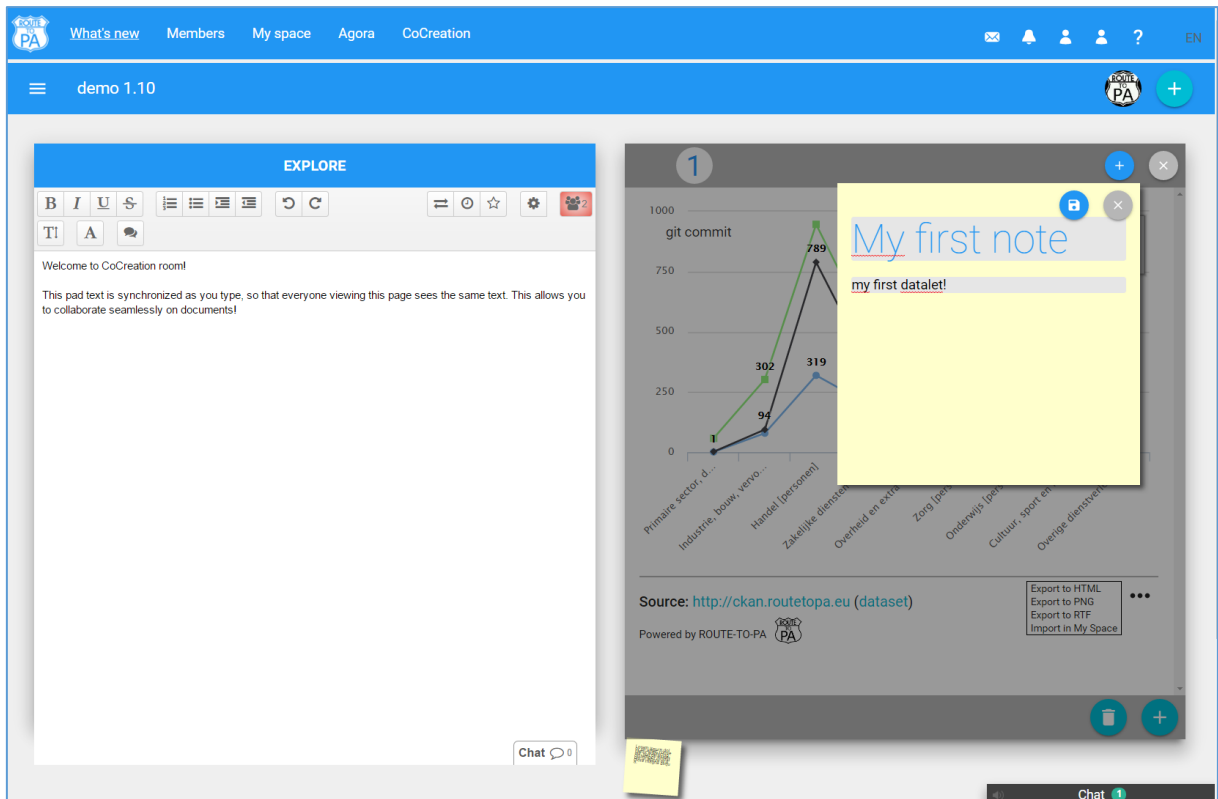


Figure 13: An example of the interface that show a document and annotated datalet with split screen feature.

For each document related to the different phases are available:

- Clear collaboration: it is possible to see changes being made by other users in real time and each participant and the related written text are identified by a color and name.
- Chat: a synchronous communication is available with other people who are using the document
- Recording Revisions: it is possible to save specific versions at any time. A time slider shows anyone the history of the shared document.
- Exporting: a document can be downloaded in plain text, HTML, Open Document, Microsoft Word or PDF format

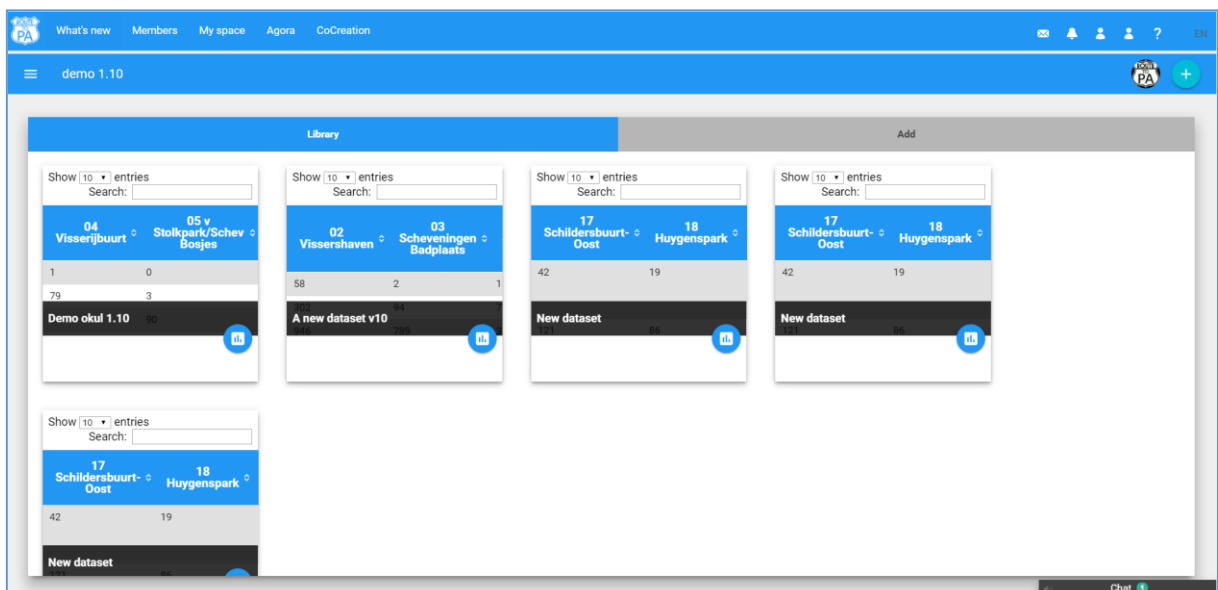


Figure 14: The Library section contains a set of datasets added from room members.

Each dataset added by room members will be placed in the controllet as *suggested dataset* and can be used to create a new visualization to support the knowledge creation process.

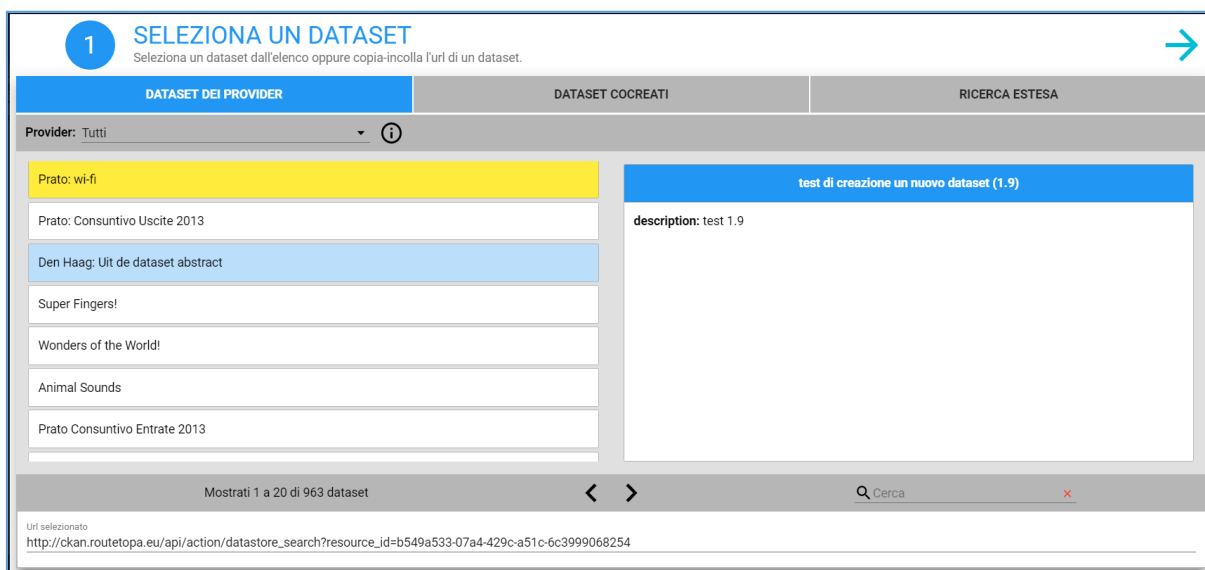


Figure 15: The suggested datasets will be highlighted in yellow

### 6.1.2 DATA COCREATION ROOM

#### Overview

The Data Co-Creation rooms are virtual places where small groups of participants meet together to collaboratively create new datasets. Citizens become active players, who do not merely access to open data but contribute to their creation. The Data Co-Creation is based on similar design principles and concepts already described in the previous sections. The room has a shared spreadsheet that allows users to progressively, collaboratively and interactively build a dataset from scratch or import an existing one to make improvements. A shared spreadsheet means that when a user is making a change to the dataset (e.g., column, cell), other users can see the change in real time. The coordination among users when needed is supported through two communication tools within the room: an instant chat and a threaded chat to allow the users to discuss. The room has a creator, who has the need to create a new dataset and initiate the room. For example, a data journalist working on her/his report initiates a room to collect the data in tabular form with her/his colleagues. Another example, a group of citizens could map all the historical places in the city.

The creation of a dataset includes some essential steps. The definition of the datalet structure in terms of how many columns should be included and how they should be named are preparatory choices before the collection and fill in of data. Of course, the structure of the table can evolve over the time, for example by adding other columns. A dataset within the data co-creation rooms is not public until a participant does not explicitly publishes it on the SPOD platform boundaries along with its metadata.

#### Data Cocreation on SPOD



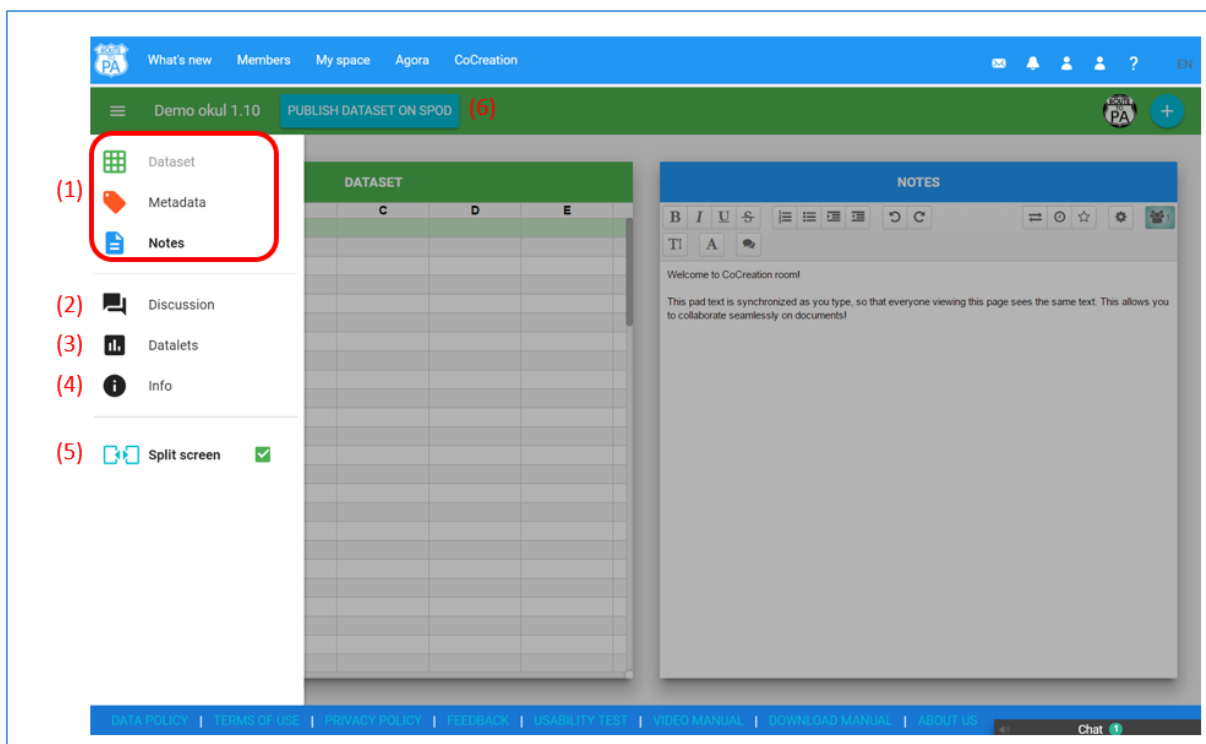


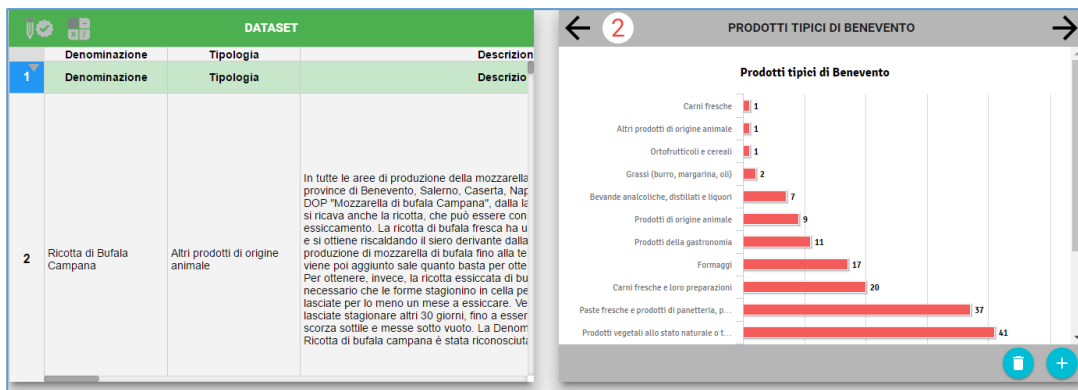
Figure 16: Create a dataset with data validation with notes and metadata for the dataset (1); social and collaborative discussion with attached file (e.g., doc, pdf and images document) (2); add a new datalet (3); show information about the room (4); split the user interface (5), finally publish the dataset to make it available to providers of the Controllet, named SPOD provider (6).

The Data Cocreation room provides these main functionalities:

- A real-time collaborative spreadsheet to create a dataset

DATASET										
	Vino	Denominazione	Colore	Tipologia	Produttore	Descrizione Organoleptica	Vitigno	Gradazione	Certificazione	Indirizzo
6	Sedici	Roccamon...	Rosso	Novello	Agricola San Teodoro	rubino, con sapore sapido fine ed equilibrato a lunga persistenza, dal profumo fruttato intenso, con sentori di ciliegia, amarena e frutti di bosco.	75% Aglianico; 15% Cabernet Sauvignon; 10% Piediroso	13%	I.G.T.	Via Provincie Loc. Martorello
7	Campapietro	Galluccio	Rosso	Galluccio	Agricola San Teodoro	Dal colore rosso rubino intenso con riflessi granato, è un vino ricco di sentori speziati di pepe nero, cuoio e tabacco, ben infusi con gli aromi del legno.	100% Aglianico	14%	D.O.C.	Via Provincie Loc. Martorello
8	Rose'	Roccamon...	Rosato		Agricola San Teodoro	Colore rosa tenue, dal sentore di rose con intenso profumo fruttato, sapore morbido, delicato e vellutato.	70% Cabernet Sauvignon; 20% Aglianico; 10% Piediroso	13%	I.G.T.	Via Provincie Loc. Martorello
9	Cannizzaro	Roccamon...	Bianco	Fiano	Agricola San Teodoro	Vino dal colore giallo paglierino con sapore sapido fine ed equilibrato a lunga persistenza, dal profumo fruttato intenso, con sentori di ananas, banana e mela acerba	100% Fiano	13%	I.G.P.	Via Provincie Loc. Martorello

- Real-time visualizations: users can create datalets using the dataset. Every time someone makes a change to the values in the spreadsheet, the visualizations will change.



- Notes attached to the dataset: users can use a collaborative editor to add additional information to the dataset such as the sources, considerations and so on.
- Metadata section: in order to make a truly open data dataset, it should be accompanied by a set of standard information that makes it linkable, searchable and useful for professional usage.

The screenshot shows a 'METADATA' form with the following sections:

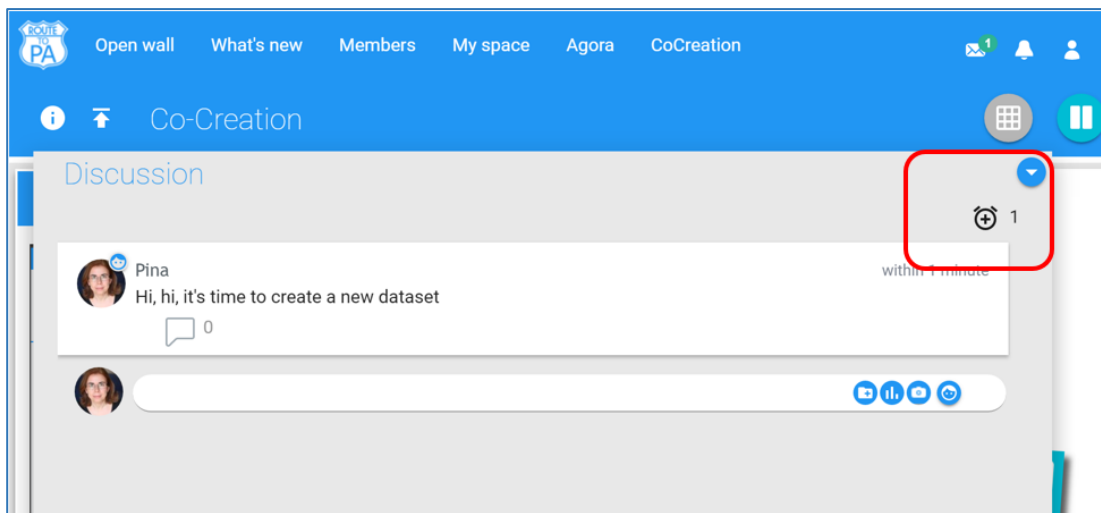
- Title**: Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery. Example: 'Itinerario del Gusto - Prodotti tipici della regione Campania'.
- Description**: Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest. Example: 'Elenco dei prodotti tipici della regione Campania, suddivisi per tipologie e certificazioni'.
- Tags**: Tags (or keywords) help users discover your dataset; please include terms that would be used by technical and non-technical users. Example: 'prodotti tipici, regione, Campania, prodotti tradizionali, prodotti agroalimentari'.

- Discussion: Through a threaded chat, users can discuss issue related to the dataset. It's possible to attach files (the permitted formats can be defined by an administrator from admin panel) and visualizations to a post.

The screenshot shows a 'DISCUSSION' forum with the following messages:

- Message 1**: <http://www.aziendebiocampania.it/...vincia=&pagina=6> (0 replies)
- Message 2**: Vittorio Scarano: Sousate ragazzi... innanzitutto complimenti.. per una demo ho inserito una datalet.. Pero' da questa datalet, ad esempio, si vedono alcune piccole imprecisioni, come ad esempio, il campo "Formagg... See more" (2 replies)
- Message 3**: Maria Anna Ambrosino: Grazie Prof. Non abbiamo ancora effettuato i controlli, e quindi uniformato tutte le diciture. Ora però provvedo a sistemare almeno questo particolare che è emerso => (0 replies)
- Message 4**: Vanja Annunziata: Grazie :D E' ancora un work in progress, man mano che emergono gli errori li stiamo correggendo :) (0 replies)

- A synchronized notification mechanism enables members of the room to be updated, in real-time, on the development of discussion.



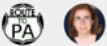














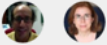




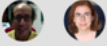




- Publish the dataset: when published, each dataset is accessible to build a datalet in the traditional ways (buttons in What's new, Agora's public rooms, etc.) inside SPOD. It is also available outside the SPOD through a URL provided by the system. Every time the dataset is published, a new version of it will be created, to avoid losing the previous, and potentially important, versions.

SELECTED DATA					
Nome Torre	Descrizione	Città	Provincia	Inizio Costruzione	Fine Costruzione
Crestarella (o Torre del Chiatamone) (Vietri sul Mare - Sa)	La prima torre che si incontra sulla strada costiera c...	Vietri sul mare	Salerno	1564	
Vito Bianchi (Vietri sul Mare - Sa)	Vito Bianchi è molto più grande perchè torre di difesa...	Vietri sul mare	Salerno	1569	
Seconda torre (Vietri sul mare - Sa)	Tale torre è adibita a torre di avvistamento e di dime...	Vietri sul mare	Salerno	1569	
Torre Cetara (Cetara - Sa)	Sul lato orientale della marina di Cetara sorge la torr...	Cetara	Salerno	1576	
Torre di Erchie (Erchie - Sa)	Anche il sistema difensivo di Erchie era costituito da...	Maiori	Salerno	1532	
Torre Capotumolo (Erchie - Sa)	Edificata intorno al 1570 oggi quasi totalmente distr...	Maiori	Salerno	1570	
Torre Lama del Cane (Erchie-Maiori (Sa) )	La torre Lama del Cane costruita intorno al 1532-33...	Maiori	Salerno	1532	1533
Torre Badia (Erchie - Maiori (Sa))	La torre Badia, anch'essa a due piani, costruita int...	Maiori	Salerno	1570	
Torre Cesare (Maiori - Sa)	La torre Cesare a pianta quadrata a tre piani, cost...	Maiori	Salerno	1532	1533
Torre Normanna (Maiori - Sa)	La torre Normanna costruita tra il 1534-90 a due pia...	Maiori	Salerno	1534	1590
Torre Mezzacapo (Maiori - Sa)	Torre Mezzacapo atata costruita nel 1584 ad un s...	Maiori	Salerno	1584	
Torre Paradiso (Minori - Sa)	Torre Paradiso restano attualmente solo i muri peri...	Minori	Salerno	1564	1599
Seconda torre (Minori - Sa)		Minori	Salerno	1565	
Terza torre (Maiori - Sa)		Minori	Salerno	1584	
Torre dello Scarpariello (Ravello - Sa)	La torre dello Scarpariello denominata anche Ficarol...	Ravello	Salerno	1533	
Torre d'Atrani (Atrani - Sa)	La torre d'Atrani a pianta circolare costruita tra il 154...	Atrani	1544	1549	
Torre dello Ziro (Amalfi - Sa)	La torre dello Ziro, sulla collina che domina Amalfi, f...	Amalfi	1278		

Showing 1 to 22 of 22 rows

PUBLISH DATASET ON SPOD

- View all dataset in Cocreation main page this feature allow users to look at the datasets created in the cocreation data rooms in SPOD. Each datasets, and its sub-versions, can be shown and downloaded and it is possible to get the relative URL to use it outside SPOD.

DATASET						<input checked="" type="checkbox"/> Last version
VER	USER	NAME	DATA	DESCRIPTION	TOOLS	
1			26/07/2016		   	
1		Wifi Prato	25/07/2016	Wifi Prato	   	
1			24/06/2016		   	
2		Humidity of 2016	24/06/2016	The first 5 months of 2016 are covered, ...	   	
2		Average temp	24/06/2016	The average temperature of the first 5 m...	   	
Showing 1 to 6 of 6 rows						<input type="text" value="Search"/>

- A synchronized notification mechanism enables users to know what is happening on room, such when someone adds a datalet, writes a comment, updates metadata and so on.

### 6.1.3 BLOGS

A new plugin for Blogs has been introduced. It allows adding post quick blog entries, with datalets embedded, and is public to all (even to unauthenticated users). Only a user with “blogger” role can create a blog. Authenticated users can vote (from 1 to 5) the blogs. By default the component is not activated and can be very quickly activated by the administrator (or ask the UNISA team). For an example, on the Prato platform the Blogs has been activated (see <http://prato.routetopa.eu/blogs>).

A user can pull all the recent posts from blogs, can read a set of latest post on Blogs, can filter the content by post blogs and display the posts from the set of tags.

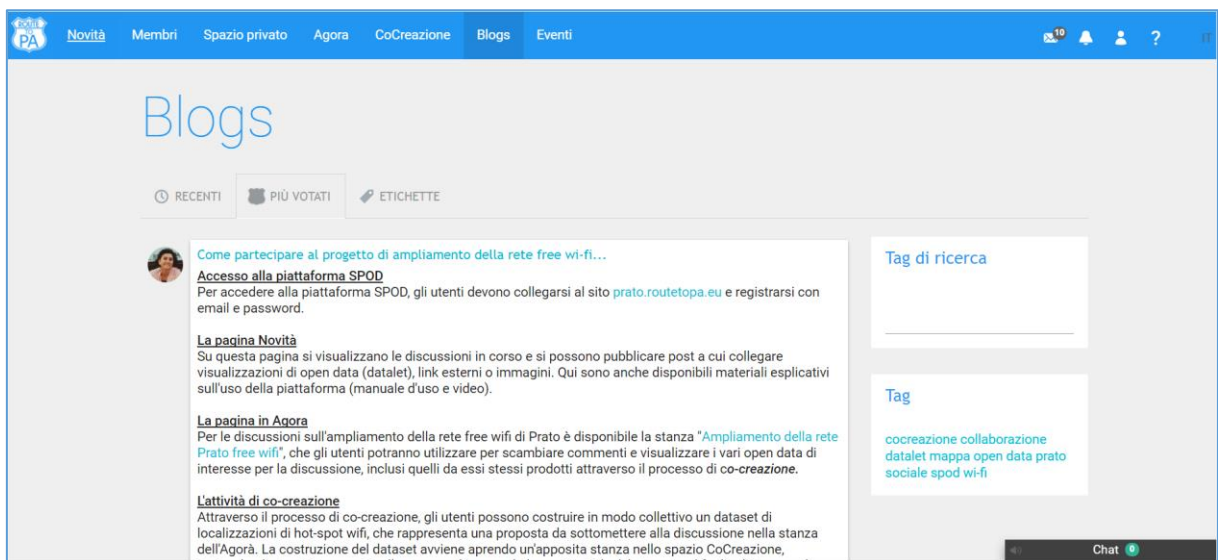


Figure 17: Blogs with Latest, Top rated, Browse tabs on the Prato platform (Italian platform)

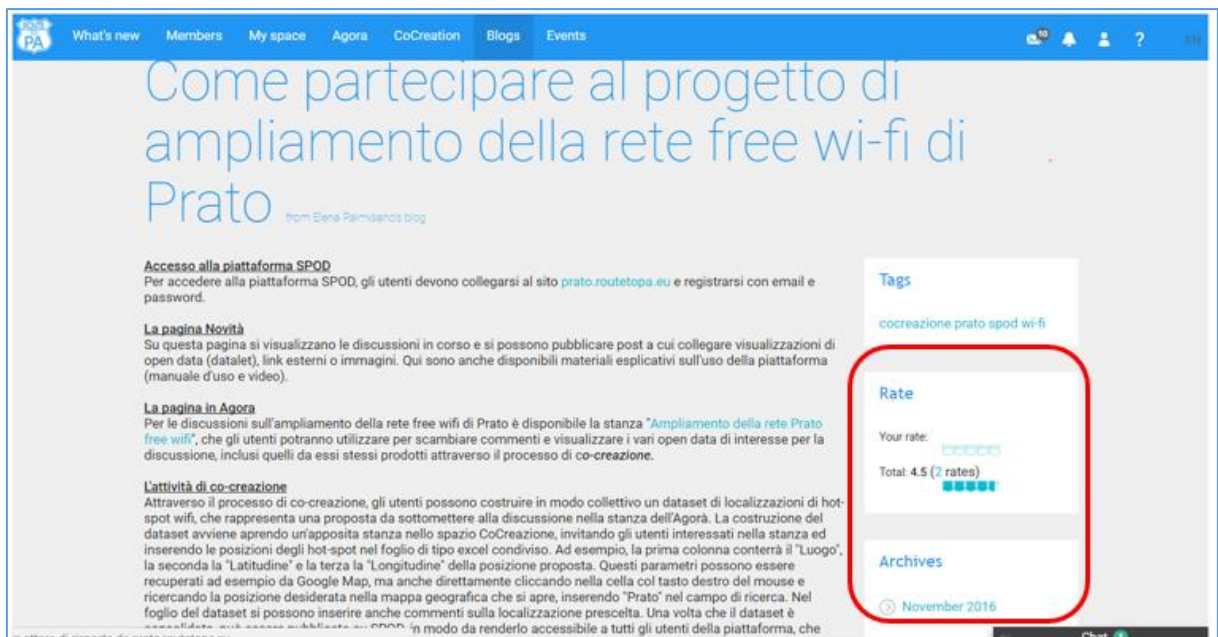


Figure 18: Blogs of Prato (Italian platform) allows to vote (from 1 to 5) the post on the right side

The SPOD's datalets can be statically embedded in a post of Blogs through a copy and paste of its source code. Every datalet can be exported in various formats (i.e., PNG, RTF, and HTML) to be inserted and reused in other places, for example, a post on Blogs (see manual for more details at <http://service.routetopa.eu:8000/d/481c8b5f9c/>).

## 6.2 ADDING MAPS FROM GEO-SERVERS

SPOD users can use this feature to create a cartographic project composed of multiple layers of public datasets that contain geographic territorial data, tied together by some common theme.

Users can correlate and visualize GIS information from different sources, from different typologies (points, lines, polygons, etc.), using navigation tools (pan, zoom) and query (information on features).

For example, it is possible to visualize municipal data relative to scholastic buildings and government data concerning the presence of asbestos in the territory, in order to encourage a public debate on the necessity of removing asbestos.

Many comments were received by the Pilots (see table below).

Pilot	Requirement
Groningen <sup>5</sup>	"It would be interesting to use this tool in relation to financial data and more specifically in which local neighborhoods the government spends money. It would be great if this tool could put that on a map".
Groningen <sup>6</sup>	A more easy function to make maps would be great. Now you need longitudinal data, but a function where you can make a map based on the datalist would be practical. Another student adds that making use of <b>kadaster</b> information and the possibility to color certain areas would enhance the mapmaking function.
Prato meeting <sup>7</sup>	For map datalet possibility to add a type category to differentiate different markers types inside a single dataset

<sup>5</sup> Open Government Action plan meeting on February 4, 2016

<sup>6</sup> Results Population Decline Challenge – User experiences (April, May 2016)

<sup>7</sup> Remote Meeting Prato, 30 June 2016

DenHaag <sup>8</sup>	Another thing that came up in the meeting yesterday: will there be more options for maps in SPOD? Now, we can create a map with points in SPOD. But we have a lot of data that needs a more GIS-like approach to a map. So a map, in which per neighbourhood a lot of data is shown (how many people unemployed, what kind of companies in the area, how many job openings etc. ). Our data guy made a map and shared the link in the first pilot, and SPOD recognized it as a datalet (you can see this in pilot 1), but it was not made in the wizard.
Issy <sup>9</sup>	<b>Decisions:</b> "On the forum will be organized a discussion about new kind of graph visualization for maps and geolocalized data."

Table 5: Pilots feedback on map

### 6.2.1 ARCHITECTURE

This feature was developed following the guidelines of Polymer (an open source library written in JavaScript from Google that aims at simplifying the creation and use of web-components).

The developed 'Maplet', uses a web component OpenLayers-map, created ad hoc, that allows access to the features furnished in the JavaScript library "OpenLayers 3" through the use of simple html elements.

The formats supported at the moment are GeoJSON, KML, WMS.

The maplet component is initialized through a JSON object, present in the attribute 'params', that contains, among others, information relative to the layers to display on a map.

During the phase of initialization, the maplet generates dynamically the necessary HTML elements.

The OpenLayers maplet component should be viewed as a basic implementation of some but not all features of the GIS library OpenLayers 3.

### 6.2.2 TECHNOLOGY

The libraries used to develop the web component OpenLayers-map are:

- [OpenLayers 3](#): GIS base library ( v3.17.1 )
- [ol3-layerswitcher](#) : utility for displaying the levels loaded on a map

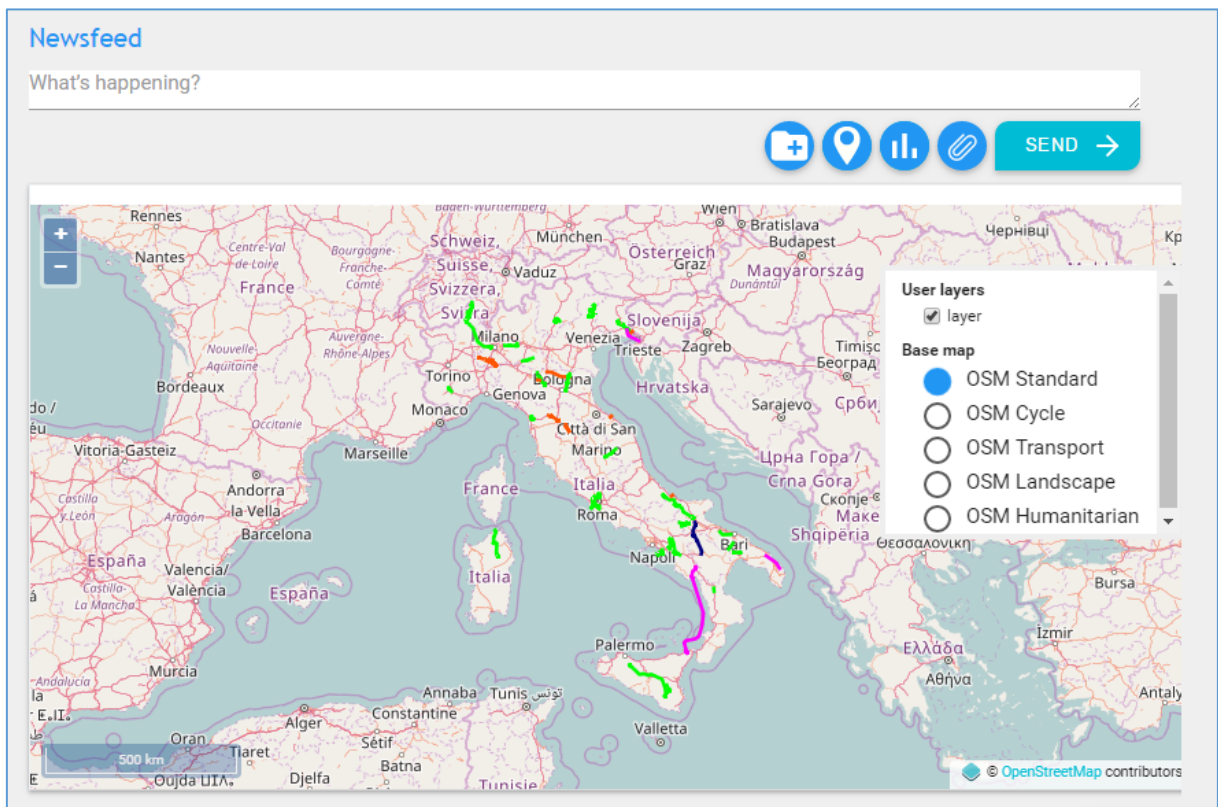
OpenLayers was developed to make it easy to put a dynamic map in any web page. It can display map tiles, vector data and markers loaded from any source. OpenLayers was developed to further the use of geographic information of all kinds. It is a free Open Source library written in JavaScript and released under the 2-clause BSD License (also known as the FreeBSD).

### 6.2.3 MAPS INTEGRATION IN SPOD

We have added a new button in "What's new" page to integrate maps inside the SPOD platform and a wizard help to create a user-generated street maps adding all kinds of useful details. In the next releases, once the tool will be consolidated, it will be accessible also in Agora and My Space.

<sup>8</sup> Meeting in May 19 2016

<sup>9</sup> Minutes Remote Meeting in April, 20<sup>th</sup>, 2016



### 6.3 EVENTS

The introduction of the events in SPOD was solicited by the use cases highlighted in the DL 2.4 and DL 3.1 deliverables<sup>10</sup>: the first deliverable underlines the needs of stakeholders took such as

- Report on local events & initiative;
- Events in neighbourhood (e.g., road works, environmental projects, information on grants);
- Events happening in the area – entertainment, sporting, charity;
- Information about free events and community venues/resources;
- Calendar of events to get people informed involved.

Instead, the second deliverable (i.e. DL 3.1) highlights

- SPOD could facilitate a link with off-line events regarding the object by e.g. the posting events or announce City meetings, NGO meetings, conferences etc. that might be relevant to the community.

Event management identifies planning the logistic and coordinating some aspects before launching the event identifying the target audience.


The management of roles in SPOD will allow creating a new an “Events” user role, and this event manager will allow him/her to execute actions specified in permissions of this role, such as for example create a new event. The event plugin shows the typical characteristics of an event (i.e. to set event time and duration, to add an image to the event, to set event Location (i.e. to add the geographic location of the event). In addition, another feature is to embed a new datalet or import a datalet, previously created, from private space (see Figure 19):

<sup>10</sup> “D2.4 Requirement Specification and Use Case Models for TET and SPOD Subsystems” (submitted in October 2015) and “D3.1: Societal Activity model of Open Data use” (submitted in January 2016)





My space   Agora   CoCreation   Blogs   **Events** 16


Location **UNISA Laboratory**

Image   Nessun file selezionato

Who can view ☐ Anybody ☒ By invitation only

Who can invite ☒ Participants ☐ Event creator only

**Add datalet**  

**EDIT** 

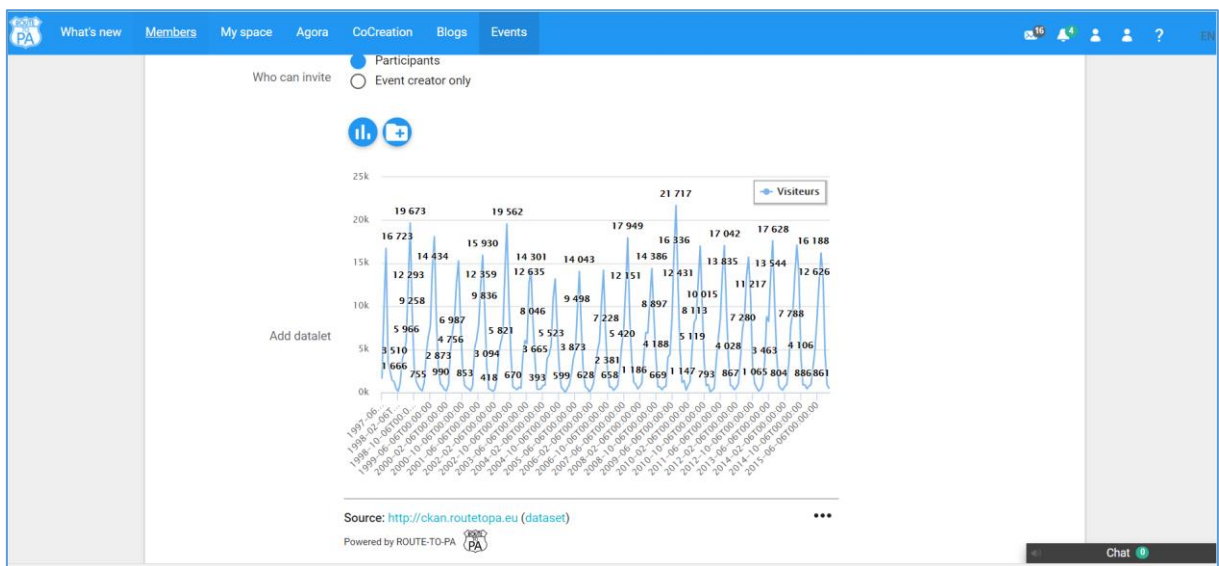


Figure 19: A new datalet embedded in the event

In the “What’s new” page, all upcoming events will be highlighted among the available widgets on the left side of the page (see Figure 20)



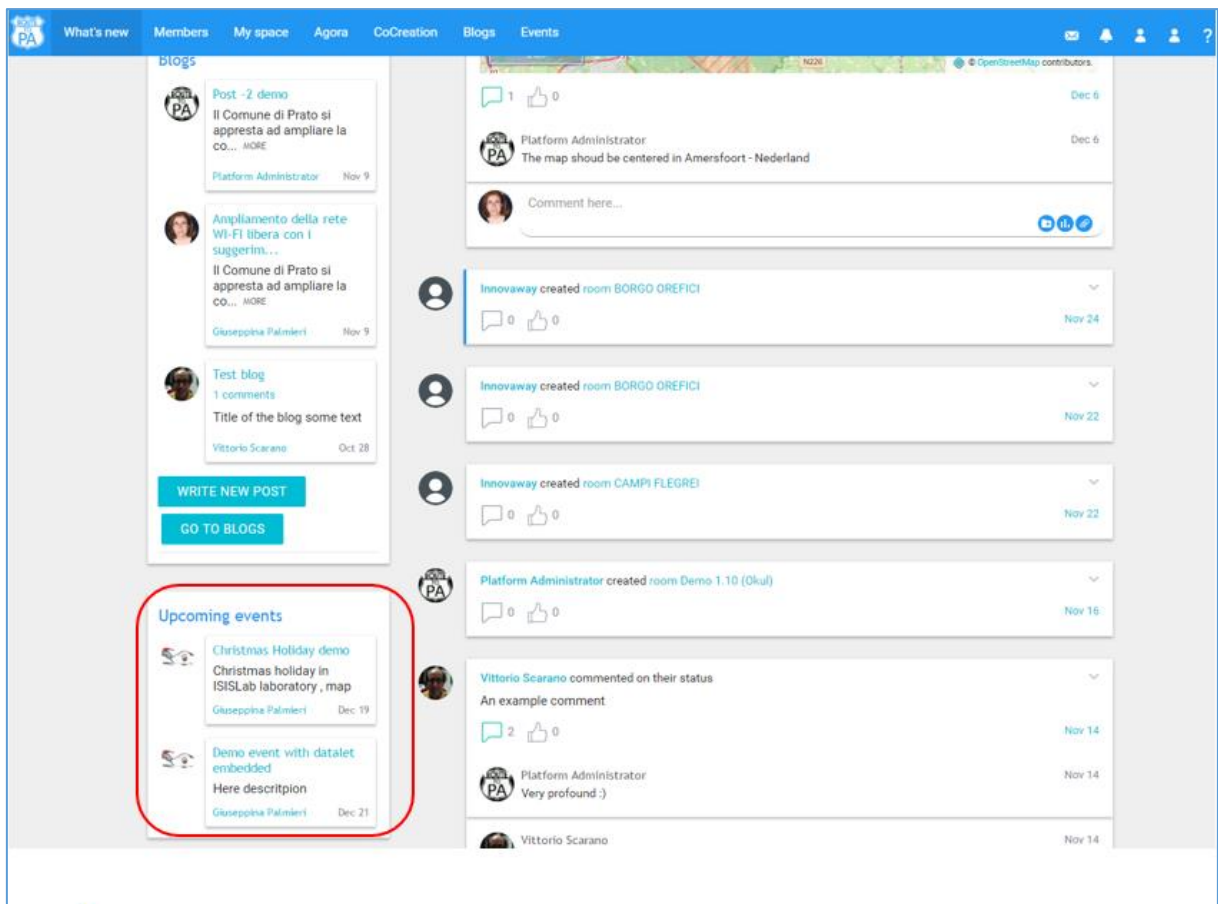


Figure 20: Upcoming events on the left page of the "What's new" page

More details are available on User Guide manual at <http://service.routetopa.eu:8000/d/481c8b5f9c/>.

## 6.4 OTHER FEATURES

### 6.4.1 ATTACH A DOCUMENT/IMAGE

In the SPOD platform, it is possible to post to attach a file in the collaborative discussion chat of the Cocreation data room.

The administrator panel allows to set parameters such as file types permitted for upload; it supports document files such as pdf, txt, ppt, xls, docx file, image files (e.g. jpeg, png, gif, bmp); it also allows to set the maximum upload file size.

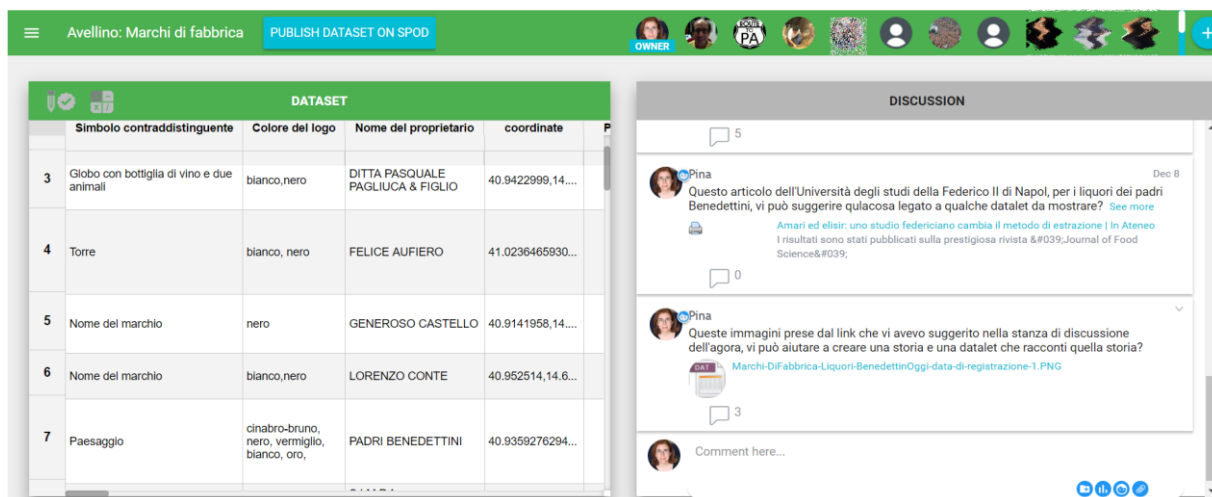


Figure 21: Cocreation data room: attached an image and pdf documents

**File attachments**

Maximum upload file size

250 MB (Server limit 250 Mb)

Allowed extensions

txt

doc

docx

sql

csv

xls

File types permitted for upload. Enter one type per line, format: xxx

SAVE

Figure 22: Administrator panel sets parameters to Maximum upload file size and allowed extensions

#### 6.4.2 CHANGING THE WELCOME PAGE (OPEN WALL)

The Open Wall provides the welcome information (such as contact email asking for information, web addresses of the project to watch activities on Facebook and Tweet) also to unauthenticated users, with embedded video to give a look at all features of the SPOD platform; furthermore, manual and videos are available to all unauthenticated users. An object of the blogging is to update others on activities of the SPOD platform.

Blogs has been introduced as a new tool for social activity to improve the quality of social interactivity: blogs posts may contain photo, multimedia content, but also interactive visualizations of chart, previously created in SPOD platform can be embed in a post. Blogs provide hypertext links to public room of the Agora to watch the started discussion.

It is available in the Prato platform at the following address: <http://prato.routetopa.eu/blogs>.

The tool “Events” shows a calendar of all upcoming events.

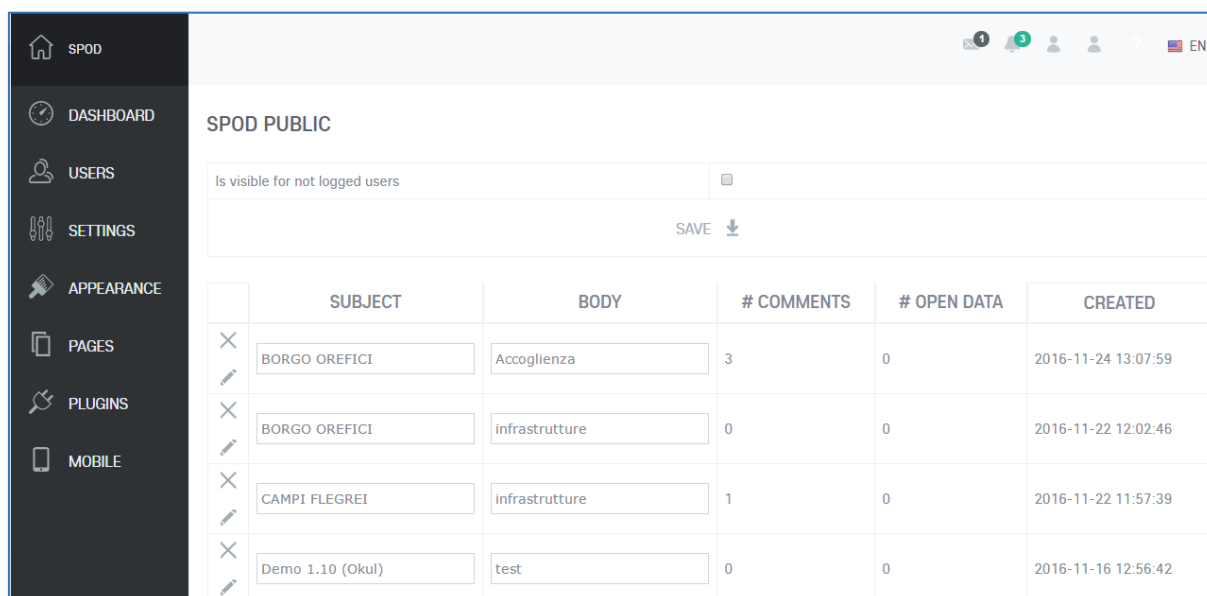
Changes in the Welcome Page have been personalized for each pilot, according to following feedback:









- Groningen feedback received in April 2016
- Dublin feedback received April 29 2016
- Issy feedback received in July 07 2016
- Groningen, Prato feedback are been collected from discussions that have taken place on the intranet's discussion forum <sup>11</sup> in May 2016

<sup>11</sup>Discussion in the forum of the Intranet addressed to an issue such “Usability #2: changing the landing page (Openwall) and What's new” discussion (<http://intranet.routetopa.eu/portal/intranet/forum>)

### 6.4.3 DELETE A PUBLIC ROOM

SPOD Platform Administrator can, through an administration page protected by log-in, delete any Agora Room. The administration panel show all the Agora Room in the SPOD installation with Subject, Body, number of comments, number of datalet and creation date. The Platform Administrator can modify the subject and the body of the room or delete it. A confirm is required to avoid unwanted deletion. Once a room is deleted, is impossible to restore its content.



	SUBJECT	BODY	# COMMENTS	# OPEN DATA	CREATED
 	BORGIO OREFICI	Accoglienza	3	0	2016-11-24 13:07:59
 	BORGIO OREFICI	infrastrutture	0	0	2016-11-22 12:02:46
 	CAMPI FLEGREI	infrastrutture	1	0	2016-11-22 11:57:39
 	Demo 1.10 (Okul)	test	0	0	2016-11-16 12:56:42

### 6.4.4 RECOMMENDED DATASETS

To stimulate discussion and simplify the use of Open Data all the users authorized to create Agora room can associate a list of the related dataset to the room itself. This dataset/s will be highlighted in the controllet. The room creator can modify (add, modify or delete) any item of the list in every moment. In the figures below the process of creation and exploitation of suggested datasets is shown.

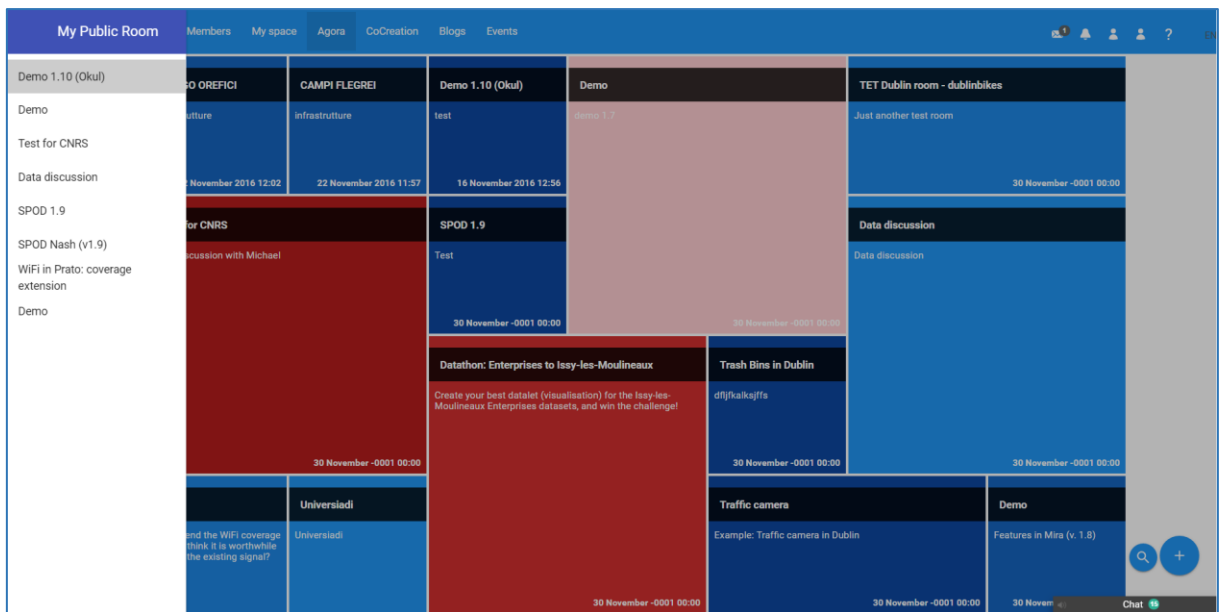


Figure 23 List of the user's Agora room

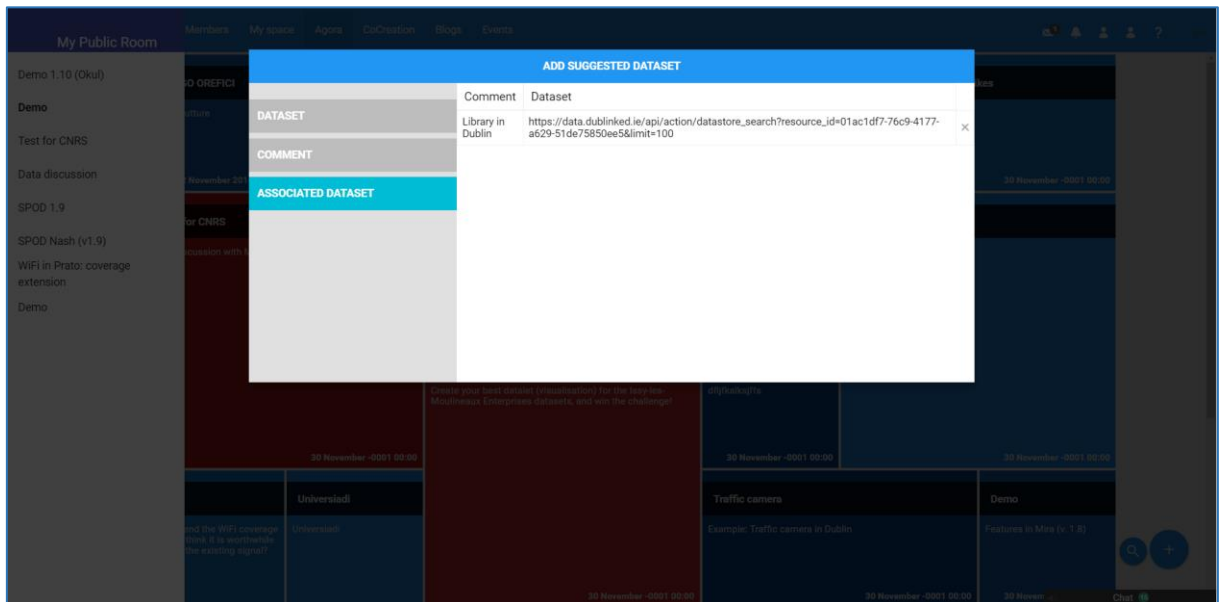


Figure 24: Agora room associated dataset list

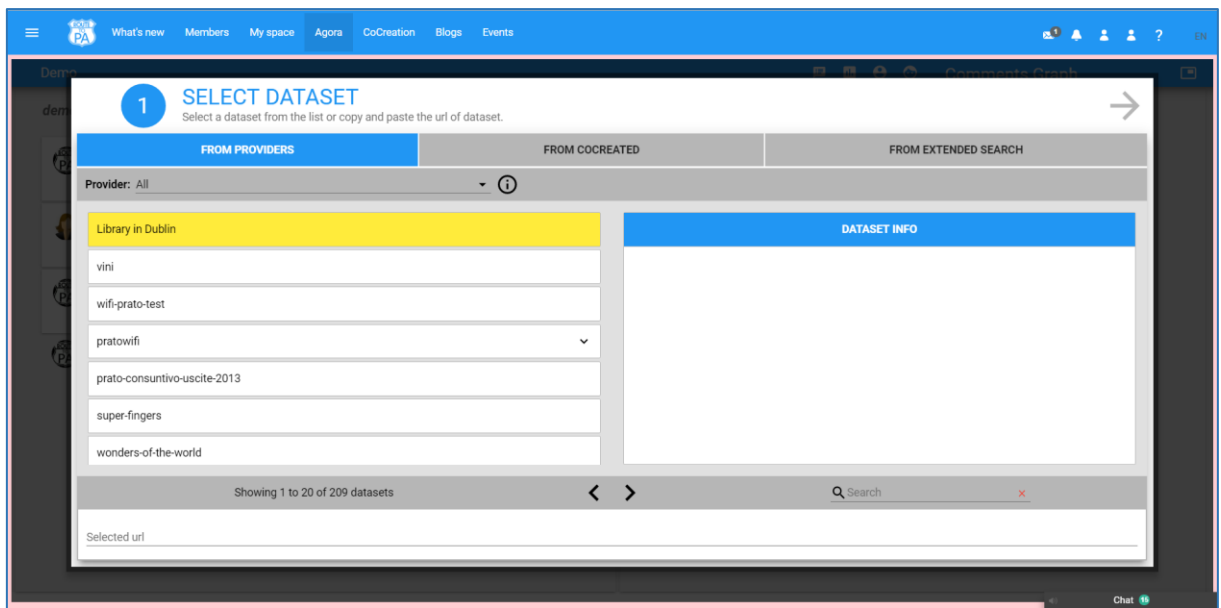
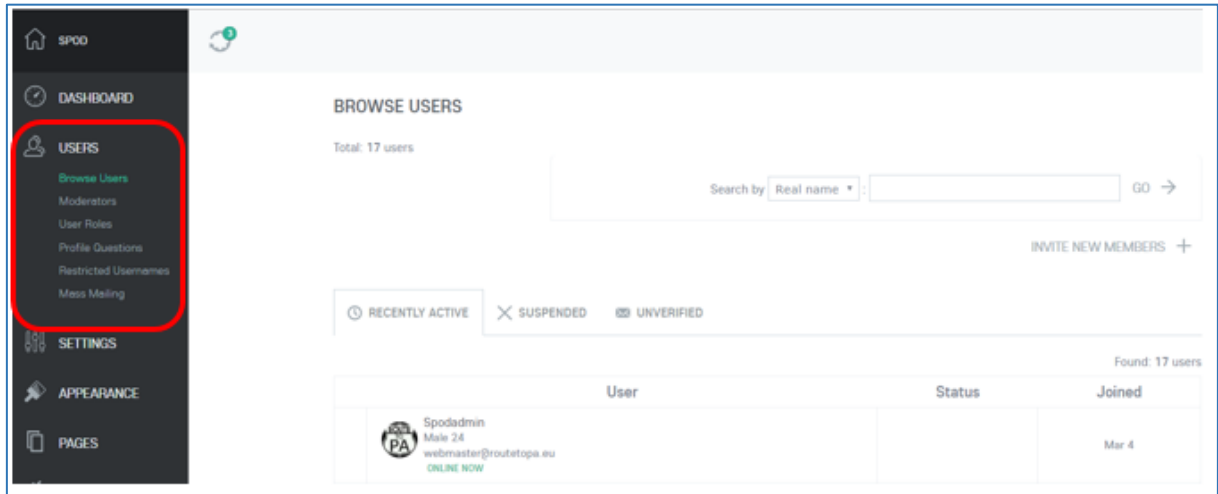


Figure 25: Suggested dataset highlighted in controllet

#### 6.4.5 ROLES

Oxwall natively implements roles, so the platform administrator can create custom role inside the platform. Each plug-in can reserve special action for a specific role. In particular, we have implemented particular action for Blog, Event and Agora plug-in. The platform administrator can enable a specific role to create an event, a blog post or an Agora room. The procedure is explained in detail in the “Administrator Guide” section 3.5 “How to manage your user”.



#### 6.4.6 CUSTOMIZING THE "WHAT'S NEW" PAGE WITH WIDGETS

The widgets are components that can be added to the widgets panel (e.g., *What's new* page) and only users as administrators can to customize specific pages of the SPOD platform.

In SPOD platform, the “What’s new” page has been enriched with custom and configurable widgets to provide extra information about what’s happening in SPOD platform, some examples:

In addition to traditional widgets (e.g., text, RSS feed, slideshow etc...), the following new widgets have been added to improve the communication between the citizens and the PA:

- “Agora activities” to notify the last created discussion room
- A list of specific online videos tutorials to help users to know the SPOD platform
- A suite of widgets integrating site with Twitter, such as Twitter of municipality.

We have enriched “What’s new page” according to following feedback from Pilots:

- Dublin Feedback received in April 2016
- Issy Feedback in July 2016
- Groningen, Prato Feedback discussed on the forum of Intranet112 in May 2016

<sup>12</sup> Discussion in the forum of the Intranet addressed to an issue such “Usability #2: changing the landing page (Openwall) and What's new” discussion (<http://intranet.routetopa.eu/portal/intranet/forum>)

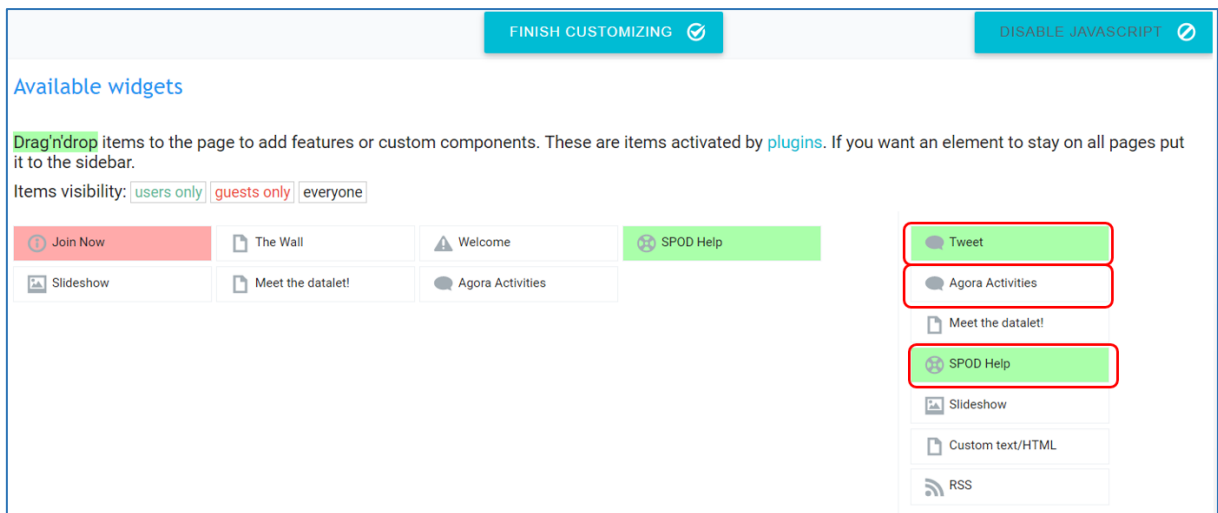


Figure 26: Some new available widgets

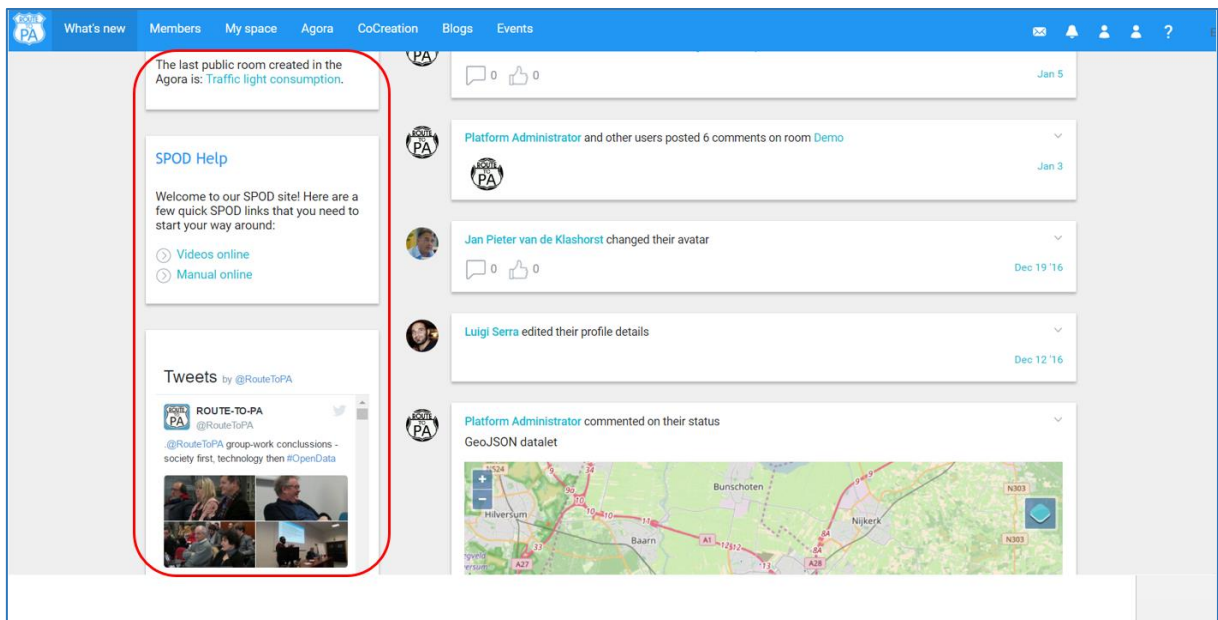


Figure 27: On the left side, a suite of new widgets

#### 6.4.7 DECISION TREES

SilverDecisions is a software for creating and analysing decision trees<sup>13</sup>. It is available visiting website: <https://github.com/bkamins/SilverDecisions/wiki> and it has been integrated in the *My Space* page of the SPOD platform.

The new feature has been implemented in collaboration with SGH Warsaw School of Economics team and has been introduced from SPOD platform v. 1.11 onwards.

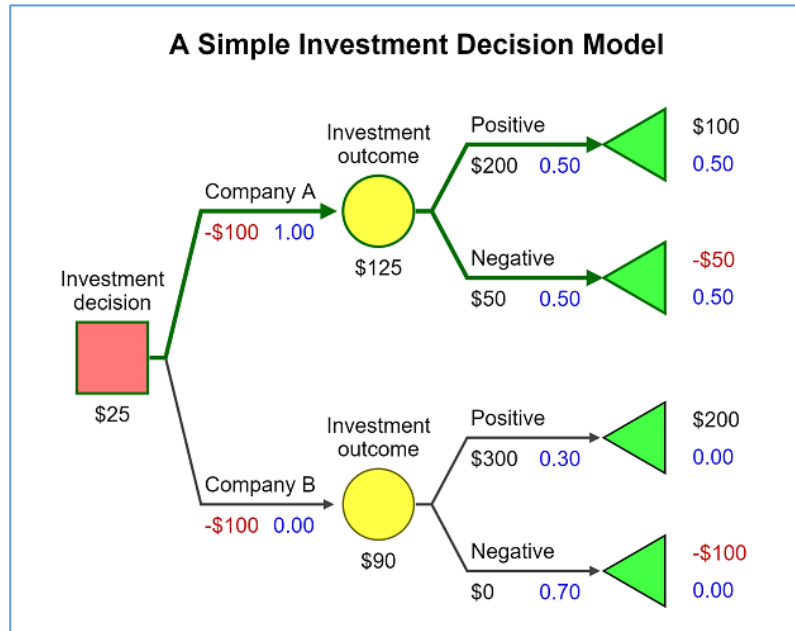
##### 6.4.7.1 WHAT IS A DECISION TREE?

Sequentiality and uncertainty are inherent in practical decision making. The former means that decision makers have to consider multi-staged strategies, encompassing several actions following one another, rather than just a single action. The latter means that decision makers' payoffs depend not only on actions but also on exogenous events (states of the world), which may often be perceived as random. The actions and reactions are usually

<sup>13</sup> [https://en.wikipedia.org/wiki/Decision\\_tree](https://en.wikipedia.org/wiki/Decision_tree)

intertwined, further complicating the picture. Decision trees are used as a model that helps to discover, understand and communicate the structure of such decision problems.

A simple decision tree created is presented below.



#### 6.4.7.2 DECISION TREE MODEL

A decision tree model describes and visualizes sequential decision problems under uncertainty in a tree-like diagram. This means that decision trees may be useful in such problems:

- the decision maker takes several actions following one another,
- the states of the world may differ based on the decisions that have already been made,
- some decisions may result in more accurate probability estimates of those states.

The tree-like diagram presents possible decisions to be made, independent events that may happen, and the outcomes associated with combinations of those decisions and events. There are two parameters that must be determined: the probabilities of events and the values. The former represents the probability that the specified state of the world happens. As the possible states of the world within one reaction are in fact competing events, the sum of their probabilities must be equal to 1. Then, the values stand for payoffs as the consequence of a decision or a state of the world. It might be either profit or loss. The decision tree model includes one more concept: expected value (or expected utility). It is calculated as the probability-weighted average of the values for competing decision-and-event sets. The expected value indicates how much one may earn or lose by making optimal decisions (this means such decisions that maximize gains and minimize losses). Finally, the outcome associated with decisions and events represents the total consequence of a set of decisions-and-events in the whole decision process. It might be interpreted as a decision maker's payoff - the result of both his decisions and the independent events that have occurred. Decision trees, with their easy to interpret structure, are excellent tools for decision analysis problems. They enable investigation of the possible decision outcomes and they help to choose between certain courses of action. A primary goal of the decision tree model is to determine the best possible decision, which represents the greatest payoff or the smallest loss.

#### 7.5.8.3 DECISION TREE STRUCTURE



A decision tree is constructed using a directed graph from left to right, with a set of nodes that split into three disjoint sets:

- decision nodes - typically represented as squares,
- chance nodes - represented as circles,
- terminal nodes - represented as triangles.

The leftmost node is called root node and it is the first decision node (first red square from the left - see A simple investment decision model above). In decision nodes it is the decision maker who makes the choice, i.e. in selecting exactly one of the branches emanating from this decision node. Those branches represent the set of available decision alternatives (actions). In a chance node (yellow circles - a sample tree above) each of the edges stemming from it - a reaction - is selected randomly with a given probability of event. Terminal nodes (blue triangles on a sample tree above) represent the outcome of a sequence of actions/reactions from the root node to that particular terminal node. The terminal node is the endpoint: no decisions can be made and no events may occur afterwards.

In the SilverDecisions application, the probabilities of events and the values associated with those events or decisions are defined in edges. The expected values calculated for every set of decisions-and-events are displayed in each decision/chance node, while the terminal nodes show the outcomes and the probabilities that one ends up in a specified terminal node.

Note that each edge is combined with two nodes: the one on the left, from which the edge comes out is called the parent node and the second one, on the right, is called the child node. Subtree is another term associated with decision trees - it represents that part of the tree which starts in any child node and each of them together with any descendants form a subtree. For instance, the subtree starting in the root node is the entire tree.

#### 6.4.7.3 DECISION TREE INTEGRATION IN SPOD

Decision tree is developed as stand-alone web tool. We have studied a mechanism to integrate Decision Tree inside SPOD's My Space.

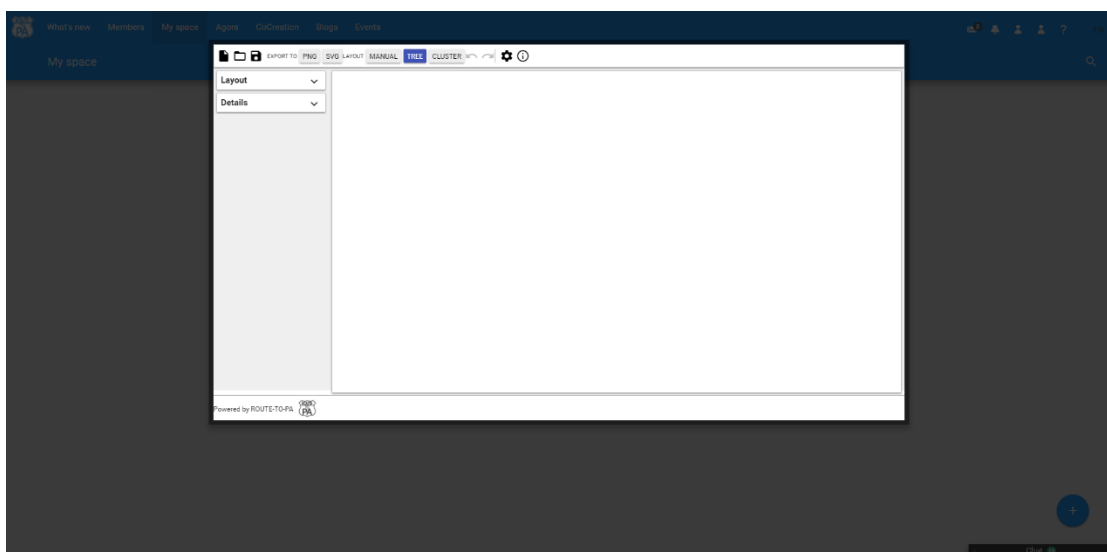


Figure 28 Create a Decision Tree into My Space

SPOD users can create a Decision Tree in their My Spaces and use the Decision Tree datalet inside What's New, Agora, My Friend Activity and Cocreation. When a Decision Tree is saved, a JSON representation of the tree is created containing all the details and layout of the tree. For the persistence of a tree inside SPOD, we save in the SPOD database it's JSON representation: when the decision tree datalet is requested we include the JSON as datalet parameter, and the tree is recreated dynamically.

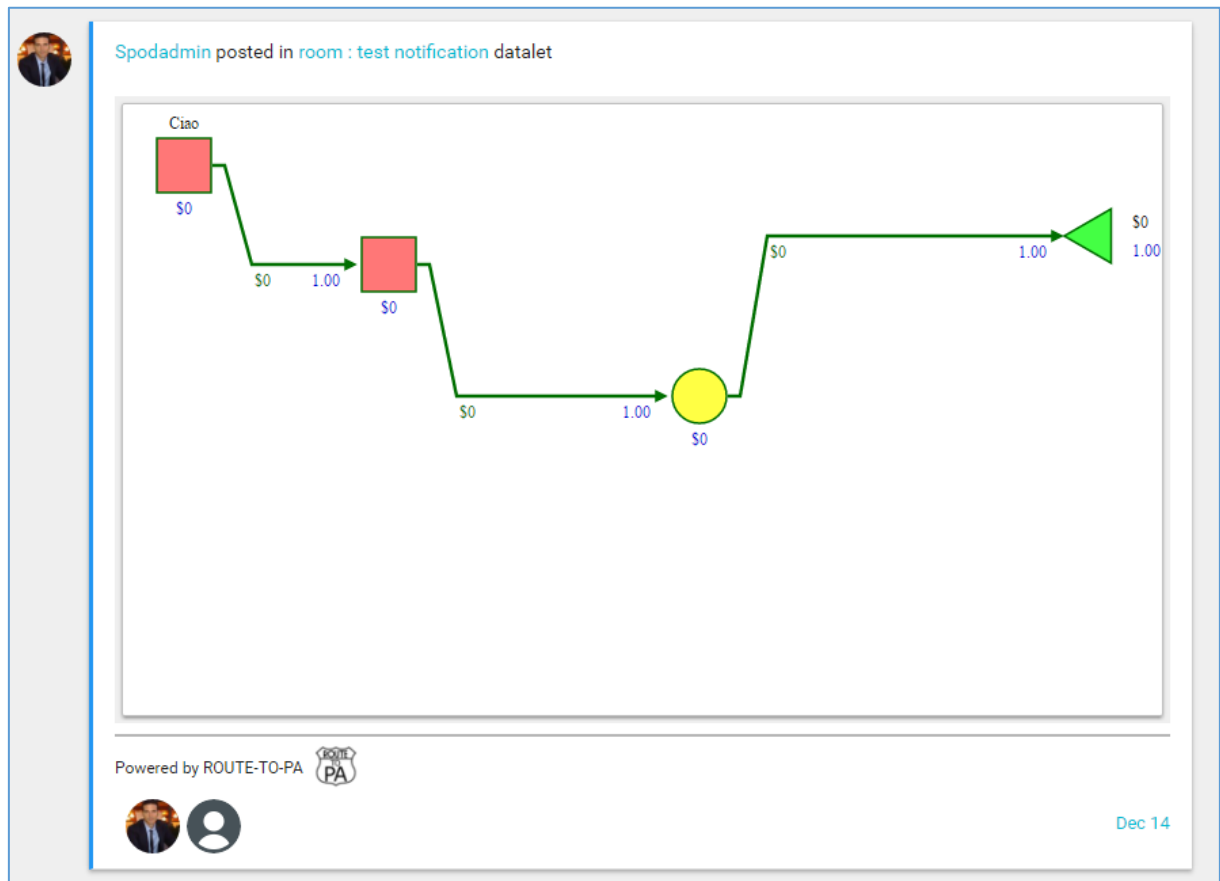


Figure 29: Decision Tree in Newsfeed

#### 6.4.8 INTERACTIONS WITH TET

##### 6.4.8.1 ROOMS USING A DATASET

Thanks to the DataFlow from SPOD to TET, when a user is browsing through a dataset details in TET 2.0, he is also presented with a list of Agora Rooms from SPOD where that dataset is used in form of a datalet. The picture below shows that the dataset “Punti Wi-Fi ad accesso libero Del Comune di Prato e della Provincia di Prato” is being used for discussion in a room called “Ampliamento della rete Prato free Wi-Fi”.



#### 6.4.8.2 ROOM SUMMARY

For each room listed by the DataFlow “Rooms using a dataset”, SPOD also exports another DataFlow called “Room summary”. In the same example as above, TET can also display a preview of the room:



In the picture above, SPOD exports a list of comments, authors and avatars from the Agora Room. The user can reach the discussion on SPOD just by clicking on the Room title (“Ampliamento della rete Prato free Wi-Fi”) or by clicking on the link “Join the discussion”.

#### 6.4.8.3 IMPORT DATALET

Users have the ability to create Pivot tables on datasets in TET. The generated Pivot Table can be seen as a derived dataset, and can be exported into user’s “My Space” area of SPOD. The example below shows a Pivot Table created in TET (notice the “Export to SPOD” button) and the resulting datalet in the user’s “My Space” in SPOD.

Organizations / Dublinked / Dublin City Council Prompt ... / DDC\_PromptPayment2013QTR2.csv

## DDC\_PromptPayment2013QTR2.csv

URL: <https://data.dublinked.ie/dataset/d13cc82e-273c-408e-bb2c-9f24861b5d44/resource/8e761a2a-f16f-4c9e-8d4e-2ebc42f56b4c/download/ddcpromptpayment2013qtr2.csv>

From the dataset abstract

Local Government Financial Reporting Prompt Payments schedule for Dublin City Council. Lists total quarterly invoices paid. This information refers to purchases in excess of...

Source: Dublin City Council Prompt Payments

Data Explorer Pivot Table

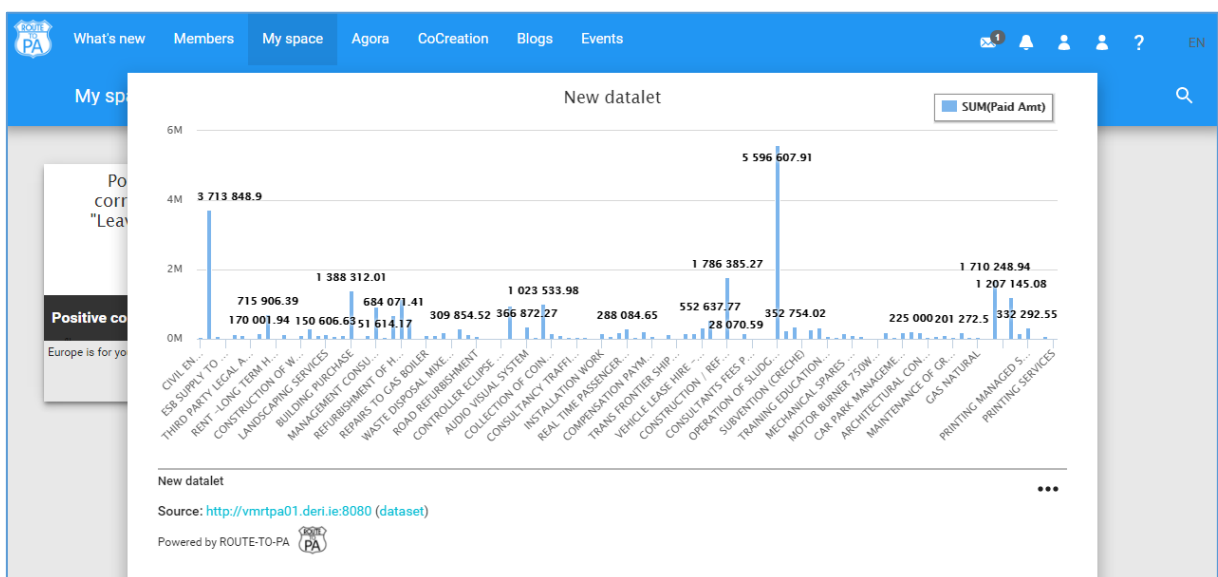
Export to SPOD

Table Vendor Name \_id Paid Amt

Sum Paid Amt

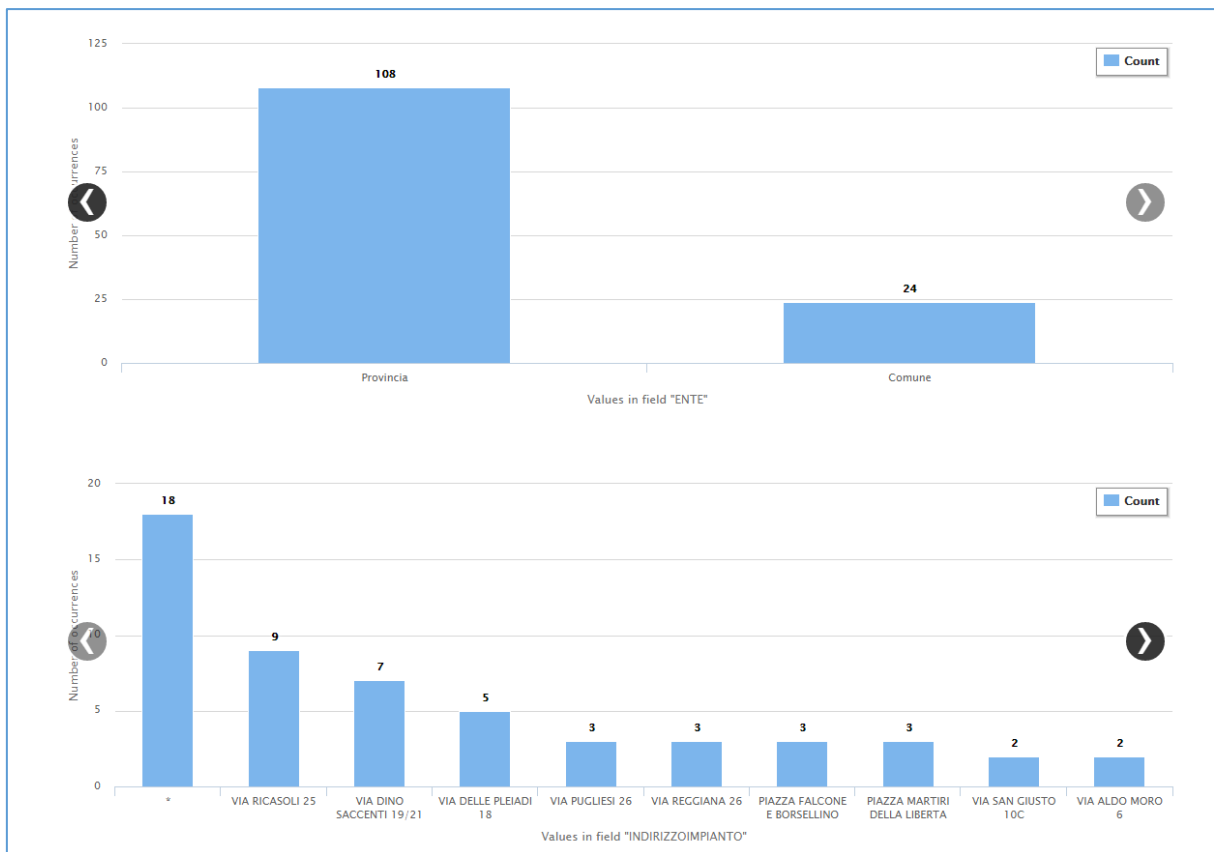
Item Description	Totals
ARCHITECTURAL PROFESSIONAL SERVICES	112,017.65
ATTIC INSULATION WORKS	96,047.35
AUDIT FEES	21,403.40
BUILDING PURCHASE	1,388,312.01
CATERING	50,779.44
CCTV CAMERA	34,620.20

Embed



#### 6.4.8.4 SPOD DATALETS IN TET

TET is now able to generate visualisation based on SPOD and DEEP datalets. The picture below shows two datasets automatically generated by TET on the same dataset. This feature gives a more seamless experience to the user.



#### 6.4.9 JOINT AUTHENTICATION

The OpenID Authentication Server was improved to accommodate requests from partners and the pilots.

- New user interface. The user interface for users as well as Platform Administrators is now coherent with the rest of the platforms and is designed in Material Design.
- Password reset. The login page has a link called "I forgot my password". Users may use this link to reset their password. This feature is a very common pattern on the Internet: a user can insert his e-mail address and ask a password reset. He will receive an e-mail with reset instructions (basically, he will need to open the provided link and choose a new password).

Sign in

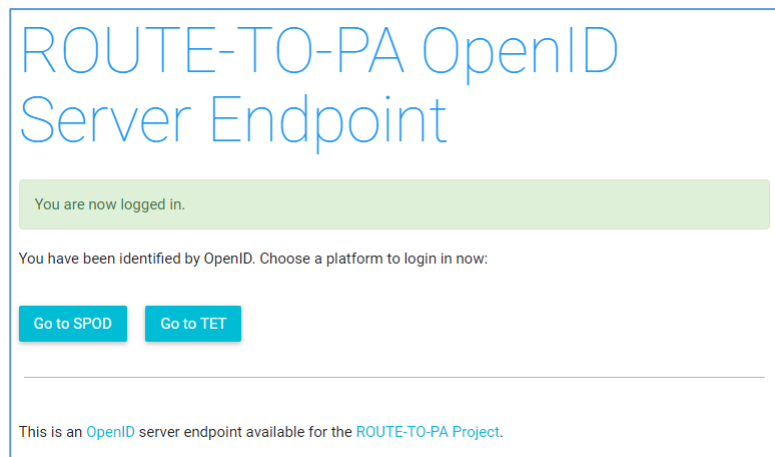
Please log in using your email and password.

E-mail:

Password:

[SIGN IN](#) [CANCEL](#) [I forgot my password](#)

- New welcome page. In the previous version, once the user was logged-in, he was presented with an empty page, not knowing what to do. Now he is automatically directed to the platform that requested the authentication (SPOD or TET). If the user accessed the Authentication directly, hence not from SPOD or TET, he is presented with links to the authorized platforms.



ROUTE-TO-PA OpenID  
Server Endpoint

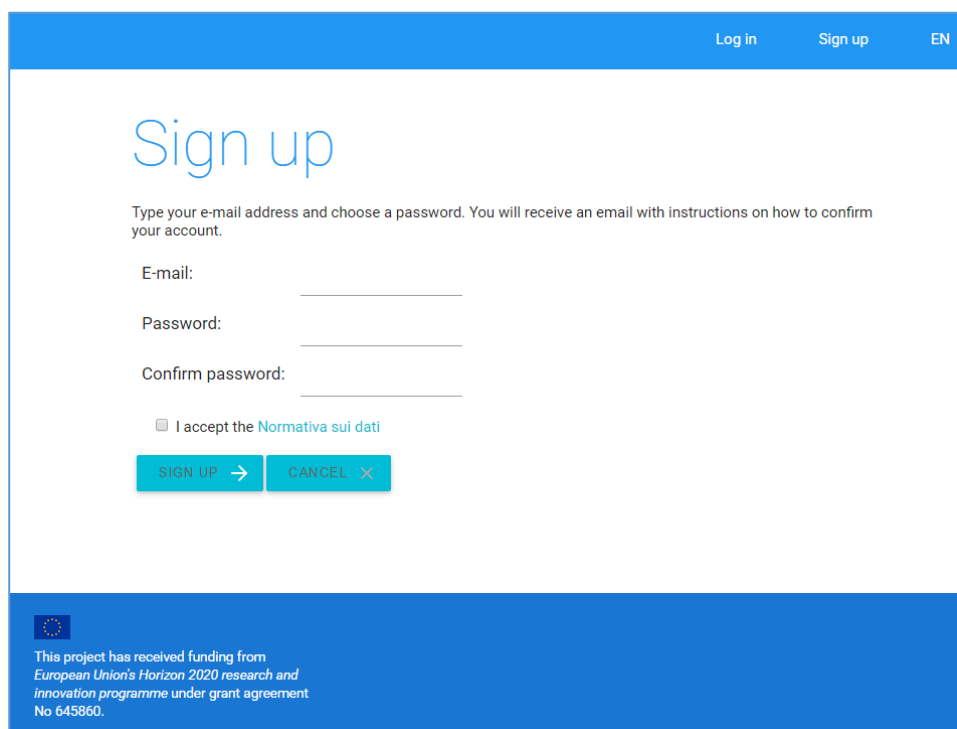
You are now logged in.

You have been identified by OpenID. Choose a platform to login in now:

Go to SPOD   Go to TET

This is an [OpenID](#) server endpoint available for the [ROUTE-TO-PA Project](#).

- User registration. Platform Administrators may choose between two form of user registration:
  - *Closed registration*: users can only be added by Platform Administrators.
  - *Open registration*: users can register on their own just using a valid e-mail address and picking a password. The user's address will be validated using a confirmation link sent by e-mail.



Log in   Sign up   EN

## Sign up

Type your e-mail address and choose a password. You will receive an email with instructions on how to confirm your account.


E-mail:

Password:

Confirm password:

☐ I accept the [Normativa sui dati](#)

SIGN UP →   CANCEL ✕

 This project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement No 645860.

#### 6.4.10 EXTENDED SEARCH

SPOD introduces an advanced search using UltraClarity as an additional utility to discover data sources. Feedback has been received from Prato Pilot. It has been implemented in collaboration with Ortelio team and has been introduced in SPOD platform v.1.5.

#### 6.4.10.1 OVERVIEW

UltraClarity is a search engine indexing the full text of datasets stored in CKAN repositories. UltraClarity improves greatly upon the standard CKAN search system, offering full-text indexing of the complete datasets in addition to metadata.

According to its website, CKAN offers ‘Google-style’ keyword search on dataset metadata fields. More specifically, it features: “Search on all dataset attributes – users can search on all dataset metadata, everything from title to tags to publisher name”<sup>14</sup>. This approach has a major shortcoming: there is no way to search the full text of the actual dataset files. CKAN supports some plain text and spreadsheet file indexing (CSV, txt) using the Data Store and Data API<sup>15</sup> but beyond that, you cannot perform search on the full text of documents such as Office files, PDF documents, and other binary files which include text.

UltraClarity introduces a pragmatic solution which indexes the full text of datasets stored in CKAN repositories and enables users to discover information from open data. UltraClarity key technology components are:

- The back-end system which retrieves, analyses and indexes all CKAN data.
- The front-end system which provides a search engine user interface and a web API which can be used by third parties such as SPOD to integrate search in their systems.

UltraClarity is using the following technologies:

- Linux and the btrfs filesystem to store all data and run applications.
- Scala programming language<sup>16</sup> to implement a system which can perform fast bulk data processing in parallel.
- Apache Tika toolkit<sup>17</sup> to detect and extract metadata and text from over a thousand different file types (such as PPT, XLS, and PDF). All of these file types can be parsed through a single interface. Apache Tika is used to extract text data from any kind of binary files hosted in CKAN repositories.
- Tesseract OCR engine<sup>18</sup> to extract text from images. Tesseract can recognise over 100 languages out of the box.
- MariaDB and the TokuDB database engine to store and organise information.
- Sphinx Search<sup>19</sup> open source search server to index and query all information.
- PHP/HHVM to implement a web application to enable search and a web API.

The UltraClarity operation can be summarised as follows:

- UltraClarity is configured to index a set of CKAN target repositories. The following steps are performed periodically in an automated way.
- UltraClarity uses the CKAN API to get a list of their datasets, groups or other CKAN objects. Example:
  - [http://demo.ckan.org/api/3/action/package\\_list](http://demo.ckan.org/api/3/action/package_list)
  - [http://demo.ckan.org/api/3/action/group\\_list](http://demo.ckan.org/api/3/action/group_list)
  - [http://demo.ckan.org/api/3/action/tag\\_list](http://demo.ckan.org/api/3/action/tag_list)

---

<sup>14</sup> CKAN FeatureTour, <http://ckan.org/features/>, accessed on 2017-01-15.

<sup>15</sup> CKAN Data Store and Data API, <http://ckan.org/2012/03/27/ckan-datastore-and-data-api/>, accessed on 2017-01-15

<sup>16</sup> Scala, <https://www.scala-lang.org/>, accessed on 2017-01-15

<sup>17</sup> Apache Tika Toolkit, <https://tika.apache.org/>, accessed on 2017-01-15

<sup>18</sup> Tesseract OCR, <https://github.com/tesseract-ocr/tesseract>, accessed on 2017-01-15

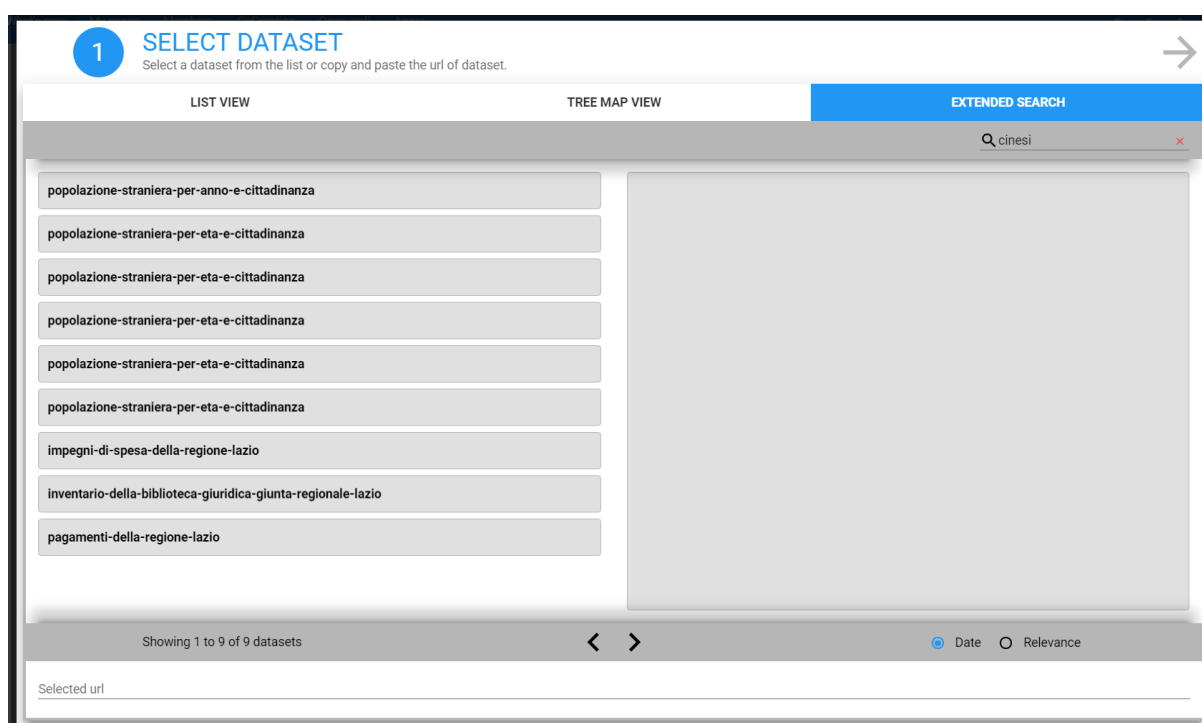
<sup>19</sup> Sphinx Search, <http://sphinxsearch.com/>, accessed on 2017-01-15



- UltraClarity uses the CKAN API to get complete metadata and binary files location for each dataset, resource or other object. Example:
  - [http://demo.ckan.org/api/3/action/package\\_show?id=adur\\_district\\_spending](http://demo.ckan.org/api/3/action/package_show?id=adur_district_spending)
  - [http://demo.ckan.org/api/3/action/tag\\_show?id=gold](http://demo.ckan.org/api/3/action/tag_show?id=gold)
  - [http://demo.ckan.org/api/3/action/group\\_show?id=data-explorer](http://demo.ckan.org/api/3/action/group_show?id=data-explorer)
- UltraClarity retrieves, analyses and indexes the metadata and binary files. The database and full text search index are updated.
- The web interface and API query the updated information to all users.

#### 6.4.10.2 SPOD INTEGRATION

SPOD is using UltraClarity as an additional utility to discover data sources. When a user tries to select a dataset in SPOD, he/she may use the “Extended Search” option which queries the full text of all datasets and returns rich and meaningful results that the simple search.



In the background, for each ROUTE-TO-PA case study, there is a matching UltraClarity instance indexing the relevant CKAN repositories. When a user types a query in the “Extended Search” text box, SPOD is contacting the UltraClarity API and performs a search in its full text index. The results are used by SPOD to present a list of datasets to the user.

#### 6.4.11 MANAGER VIEW

In this section, we will present some features for platform administrator, available from backend section of SPOD, as an integration of the previously described functionality. More details are available on the Administrator Guide at <http://service.routetopa.eu:8000/d/481c8b5f9c/>.

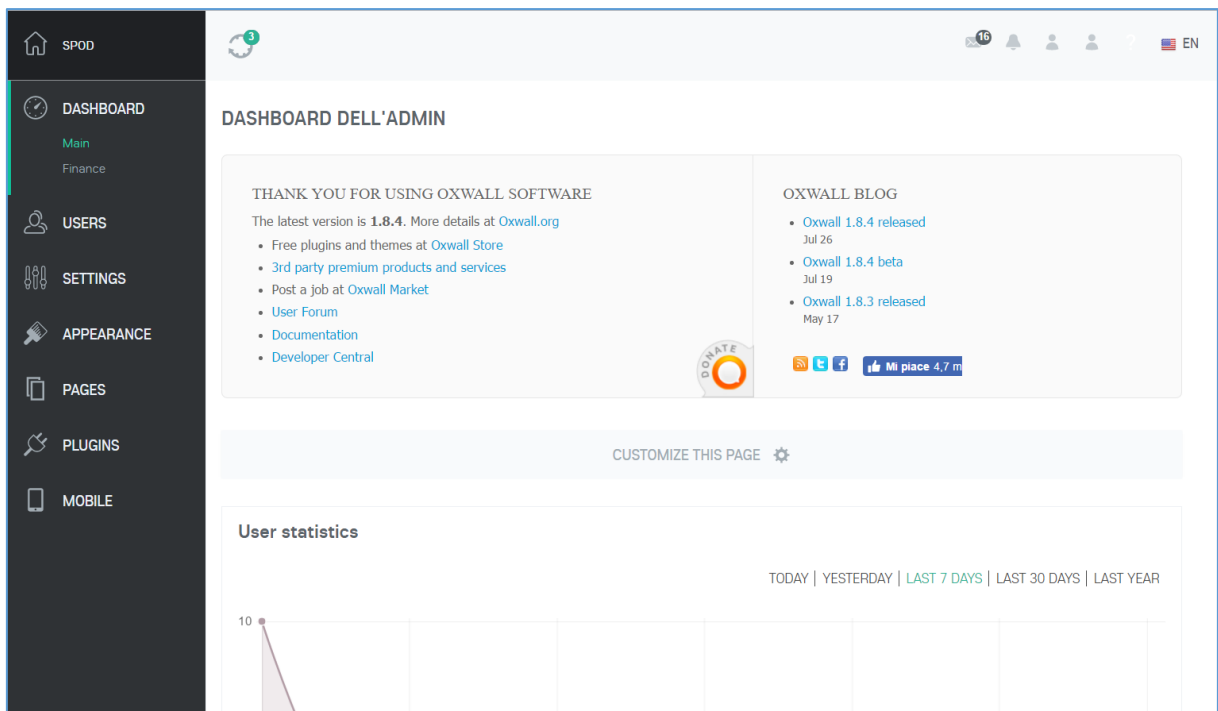


Figure 30: Administration UI dashboard

A tour through the administrator panel for exploring the back end of the SPOD platform, to describe the additional and advanced feature list introduced by UNISA team, some features have been explained in the previous sections:

- User roles management (section .5.2)
- Privacy (section 9.2)
- Manage rooms in the Agora (section 6.5.4)
- Customize the “What’s new” page (section 6.5.7)
- Customize the “Welcome page” (section 6.5.3)
- Define file formats for the attachments (section 6.5.2)

#### 6.4.11.1 DATA EXPORTER

To analyse data of each Agora and Cocreation, the admin area provides an export feature. It is possible to get discussions and other information in different formats such Excel, JSON or XML and, to have a snapshot related to the visualizations, in some case in HTML file. In this way, we provide a data collection to allow researchers to analyse them (below some screenshot of Agora and Cocreation exporter respectively).

AGORA EXPORTER						
ROOMs AVAILABLE						
	SUBJECT	BODY	# COMMENTS	# OPEN DATA	CREATED	SNAPSHOT(s)
TAKE SNAPSHOT	Itinerari storici	Città di Nocera Inferiore (Liceo Vico)	9	1	2017-01-11 16:26:45	
TAKE SNAPSHOT	Museo Open d'istituto	Città di Nocera Inferiore (Liceo Vico)	14	1	2017-01-11 16:24:56	
TAKE SNAPSHOT	Stanza di discussione	Liceo	156	2	2016-12-20 13:46:58	
TAKE SNAPSHOT	Casellario Politico Centrale	Città di Avellino	180	30	2016-11-22 09:10:10	2017-01-09 22:43:47 (#comments: 180, #opendata: 30) (delete) (download as JSON) (download as XLS)
TAKE SNAPSHOT	Marchi di fabbrica	Città di Avellino	180	9	2016-11-22 09:09:22	2016-12-13 01:18:47 (#comments: 180, #opendata: 9) (delete) (download as JSON) (download as XLS)
TAKE SNAPSHOT	Discussione e confronto	Città di Avellino	69	25	2016-11-21 11:20:42	
TAKE SNAPSHOT	Discussione e confronto	Città di Montoro	3	1	0000-00-00 00:00:00	

SETTINGS ANALYSIS						
	NAME	SUBJECT	DESCRIPTION	GOAL	TIMESTAMP	TYPE
DONWLOAD DISCUSSION	Demo - 1	demo	demo	goal	2016-07-26 18:22:11	data
DONWLOAD DISCUSSION	Prato Demo Wi-fi	Wi-fi hot spot	wi-fi hot spot in Prato	pppppp	2016-07-27 10:57:56	data
DONWLOAD DISCUSSION	Vini - Itinerari del Gusto	Vini della Campania	Vini della Campania	Censire i Vini della Campania	2016-09-22 11:04:31	data
DONWLOAD DISCUSSION	Musei in Campania	Musei in Campania	Musei in Campania	Musei in Campania	2016-06-22 13:00:24	data
DONWLOAD DISCUSSION	Nazionale Islandese	Nazionale Islandese	Nazionale Islandese	Nazionale Islandese	2016-06-28 16:19:41	data
DONWLOAD DISCUSSION	Parchi a Baronissi	xxx	xxx	xxx	2016-07-26 12:00:29	data
DONWLOAD DISCUSSION	Teatri e anfiteatri antichi in Campania	Teatri e anfiteatri antichi in Campania	Teatri e anfiteatri antichi in Campania	Teatri e anfiteatri antichi in Campania	2016-07-26 12:32:02	data

#### 6.4.11.2 SERVERS MANAGEMENT

From the Admin panel, in the plugins section, there are the setting pages to manage both notification and cocreation (text and spreadsheet editors) servers. In order to make easy the process of starting/stopping or status checking to the admin, we provide a simple interface with toggle buttons. For the cocreation we provide, furthermore, a way to select the kind of cocreation room the admin wants to be available for end-users.

SETTINGS

ANALYSIS

Document server status

Spreadsheet server status

Active room types

Knowledge room

Dataset room

SAVE

### 6.4.11.3 CACHE MANAGEMENT

From the back end, in the Open Data Enabler setting page, the admin can configure the associated providers and manage the cache (as described in the section 7.3.1) to avoid the overhead related to the requests for information about the datasets provided by providers.

SPOD

DASHBOARD

USERS

SETTINGS

APPEARANCE

PAGES

PLUGINS

MOBILE

3

ODE PROVIDERS

SETTINGS

PROVIDERS

NAME

URL

ADD

NAME

URL

×

ROUTE-TO-PA

http://ckan.routetopa.eu

×

TET-Prato

http://prato-tet.routetopa.eu:8080

1000

CREATE CACHE

Figure 31: Management of the associated provider and caching system

### 6.4.12 EMAIL NOTIFICATION STRATEGY

SPOD provides an email notification system to allow the users to know what activities are happening on platform.

At this time, we provide a way to ensure this service and we have implemented only a working proof of concept related to the activities in the cocreation room. We have planned to cover the all platform events in the third year.

Concerning the cocreation room: we define two kinds of notification: “Room creation”, only for moderator users, and “adding a comment to discussion”, for all users. In this way:

- When someone creates a new cocreation room, both knowledge and data, the system will send an email to notify this event to all platform moderator
- When someone adds a comment to discussion related to the cocreation data room, the system will send an email to all users who have subscribed to this feature: to avoid email bombing every time a user add a post in the discussion, an increasingly sparse mechanism has been adopted. The frequency of notifications follows Fibonacci series: after (13, 21, 34, 55, 89, 144, 233, 377 610,987, 1597, 2584, 4181, 6765) messages, a notification will be sent

All users, both moderator and regular, can subscribe to email notification service from profile setting. Only the available features will be shown, based on user role. It can also select the frequency of notifications in order to, if desired, to get a cumulative mail containing all activity happened.

CoCreation

☐ Someone creates a new cocreation room

☐ Someone comments in a data cocreation room

Events

☒ Someone invites me to an event

☒ Someone posts on event walls

Also send by email

☒ Immediately

☐ Automatically (if you don't show up on the site for 2 days)

☐ Never

SAVE

Here an example of email related to a room creation.

Dear Luigi Serra,

Here's the latest activity on [SPOD](#) related to you:

Jan 4



Platform Administrator has created a new room **Taiga**  
**-Issues -Statistic (From February-July)**

Talk soon,

SPOD, administration

<http://spod.routetopa.eu/>

[Unsubscribe from this type of notifications](#) [Change preferences and frequency of these notifications](#) | [Unsubscribe from all notifications](#)

#### 6.4.13 AGORA SETTINGS

By default, rooms in the agora are visible only to authenticated users. It is now possible to configure SPOD so that Agora rooms are visible (in read-only mode) by unauthenticated users. The settings is available in the Platform Administration panel:

	SUBJECT	BODY	# COMMENTS	# OPEN DATA	CREATED
X	new	new	1	0	2016-11-05 00:06:59
X	Prato: features	demo interactions	2	2	0000-00-00 00:00:00
X	Demo 1.9 - test da	Test data	2	3	0000-00-00 00:00:00
X	My room	Solo testo	0	1	0000-00-00 00:00:00

If a user tries to access an Agora room (e.g. by following a link from TET as shown in the pictures below), s/he will be forwarded to the login page (pictures below):

### Descrizione

Il dataset elenca tutti i punti di accesso alla rete Wi-Fi esistente sul territorio provinciale di Prato e gestiti sia dalla provincia di Prato che dal Comune di Prato. Il dataset contiene l'indirizzo dei punti di accesso gestiti da enti ed istituzioni, mentre i punti di accesso disponibili in luogo aperti (Piazze strade, etc..) non hanno indirizzo. Ogni punto è caratterizzato da coordinate geografiche relative o all'indirizzo della sede o al centroide dell'area aperta coperta dal servizio.

### Condividi questa pagina

Usa l'icona sotto per condividere questa pagina sul tuo Social Network favorito.

### Metadati

This Dataset was created at **21 October% 2016, 08:26** and last modified at **21 October% 2016, 08:26**.

This Dataset is published under **Creative Commons Attribution** license.

The data was published by **Comune di Prato, Italy**.

The data is available in 2 formats: **CSV XLS**

### Discussioni su SPOD

> Prato: features

Spodadmin  
Comment here...

partecipare alla discussione

### Dataset collegati

- Popolazione per eta' e sesso al 31/12/2013
- Popolazione straniera per eta' e cittadinanza al 31/12/2010

Qualità dataset

# Sign in

You are now logged in.

Please log in using your email and password.

E-mail:

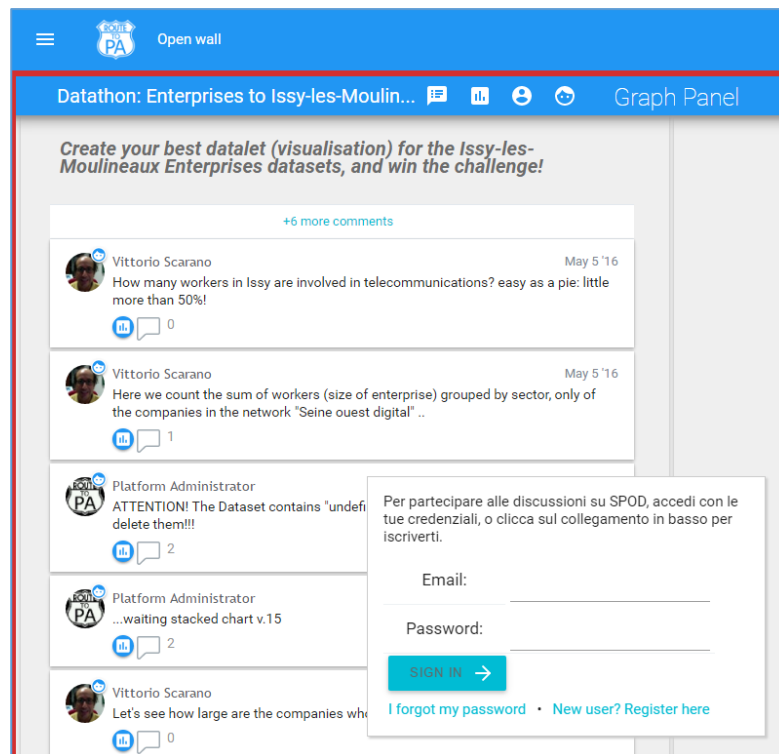
Password:

SIGN IN →

CANCEL ×

[I forgot my password](#)

If the privacy setting is disabled, guest users can read the agora room content, but they will need to authenticate in order to participate the discussion. The agora will show a small login form on the bottom right corner of the screen:



#### 6.4.14 PRIVACY PLUGIN

The Privacy plug-in allows the Platform Administrator to grant/restrict visibility of each feature of SPOD for guest users.

The plug-in works by intercepting requests made at Controller level (we mean the Controller of a Model-View-Controller architecture like Oxwall's). Each request is dispatched to a Controller that handles it through an Action.

The following table lists Controllers and Actions that handle main SPOD features:

Plugin	Controller:action
What's new	BASE_CTRL_ComponentPanel:index
Members	BASE_CTRL_UserList:index
My space	SPODPR_CTRL_Main:index
Agora	SPODPUBLIC_CTRL_Main:index
Public room	SPODPUBLIC_CTRL_PublicRoom:index
CoCreation	COCREATION_CTRL_Main:index

In order to hide other features, the Platform Administrator must identify the relevant Controller and Action. This can be done by enabling the Debug mode for the Privacy plugin (in its settings) and the normally browsing to the feature to disable: Controller and Action will be listed on top of the page (pictures below).



**SPOD PRIVACY SETTINGS**

Filter

BASE\_CTRL\_UserList:index,act2  
BASE\_CTRL\_UserListX:index

controller1:action1 controller2:action2,action3 controller3:\*

Enable debug ☒

SAVE SETTINGS

OW Debug - SPOD Privacy  
Controller: BASE\_CTRL\_UserList  
Action: index

What's new Members My space Agora CoCreation

**Browse Users**

ALL ONLINE SEARCH

Jan Pieter van de Innovaway Elena Santi

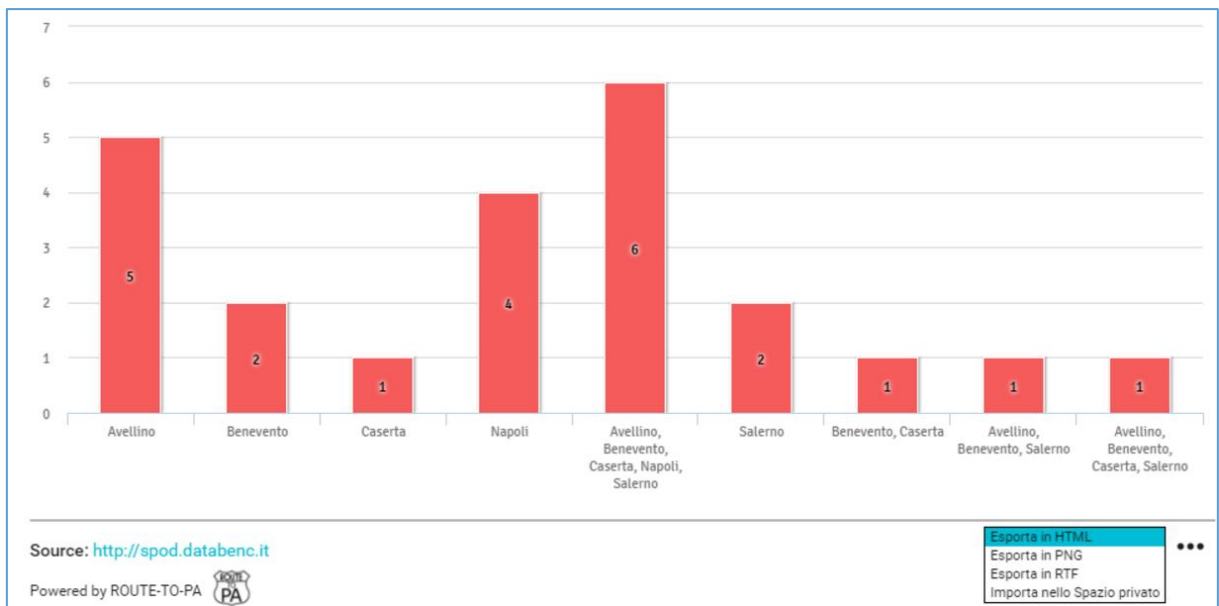
#### 6.4.15 DATALETS IN WORDPRESS PLUG-IN

The Wordpress plug-in<sup>20</sup> allows users to embed datalets in Wordpress pages or posts.

The plugin zipfile can be downloaded on the Releases section of the the GitHub repository <https://github.com/routetopa/wordpress-datalet>. The installation is standard: the Wordpress administrator installs the plugin by unzipping the archive into the Wordpress' plugin directory, or by submitting it in the Wordpress' admin section.

In order to use it, users need to create and export a datalet using the "Export as HTML" feature available under the three dots menu in the bottom right corner) of a datalet.

<sup>20</sup> <https://wordpress.org/plugins/>

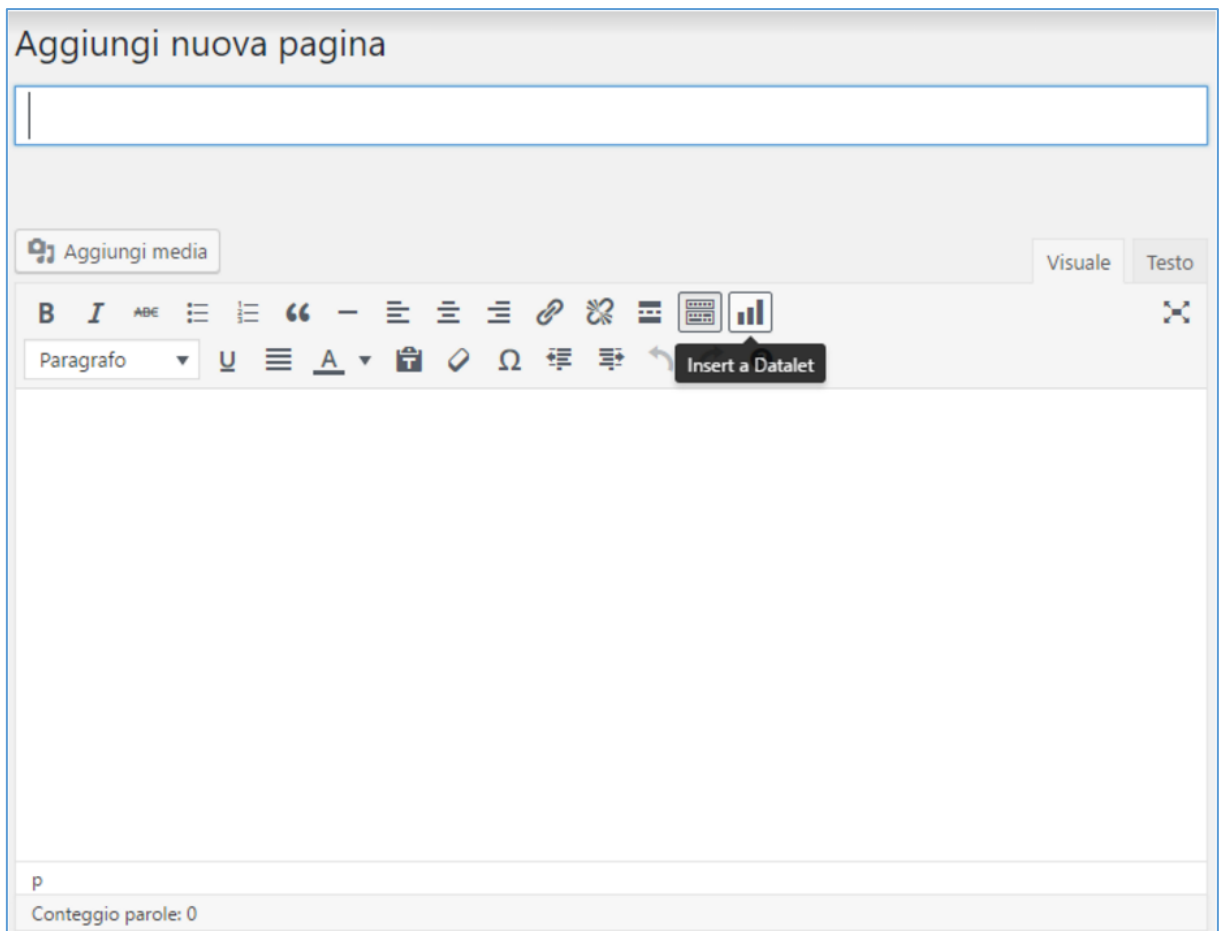


This will show the datalet code, that users need to copy to clipboard:

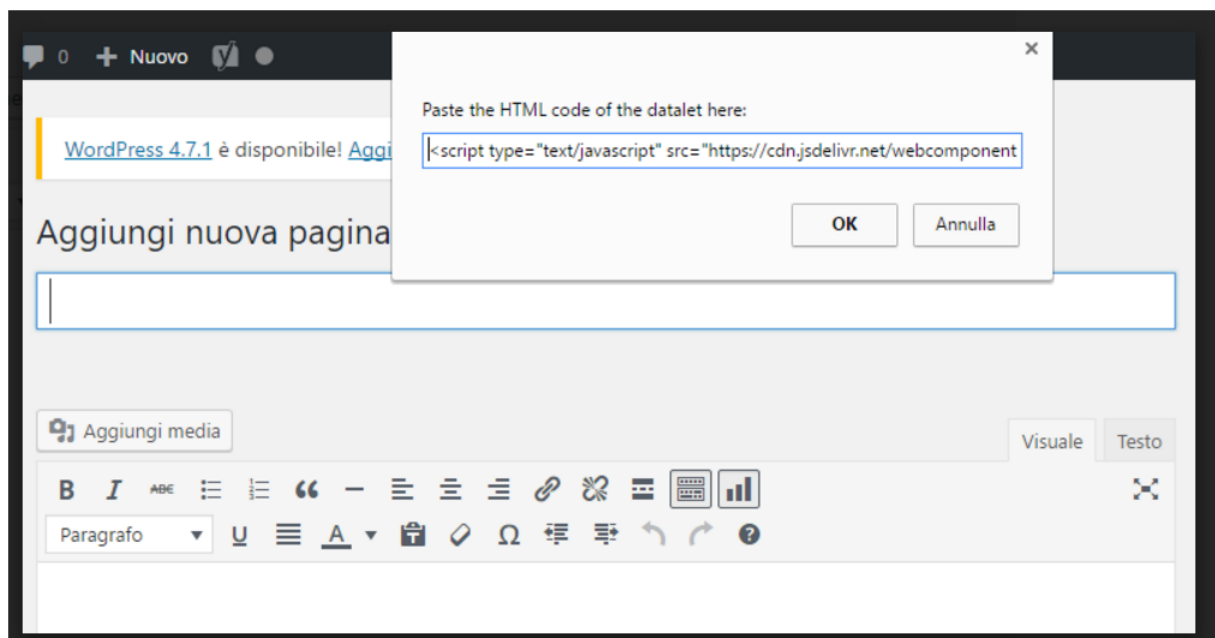
```
Copy to clipboard: Ctrl+C, Enter

<script type="text/javascript"
src="https://cdn.jsdelivr.net/webcomponentsjs/0.7.16/webcomponents-lite.min.js">
</script>
<script type="text/javascript" src="https://code.jquery.com/jquery-2.1.4.min.js">
</script>
<!-- REMOVE THE PREVIOUS SCRIPT TAGS IF YOUR PAGE ALREADY USES JQUERY AND
POLYFILL LIBRARIES -->
<link rel="import"
href="http://deep.routetopa.eu/deep_1_11/COMPONENTS/datalets/columnchart-
datalet/columnchart-datalet.html" />
<columnchart-datalet data=" fields=["Denominazione","Tipologia","Provincia"] datalet-
```

Once the code is copied, it can be pasted into a Wordpress article, using the "Insert datalet" command available on top of the Wordpress' editor:



Wordpress will ask to paste the datalet code previously copied from the datalet.



The editor will show a shortcode that will be rendered as a datalet once the page is published.

B I ABE

Paragrafo

[datalet import="[http://deep.routetopa.eu/deep\\_1\\_11/COMPONENTS/datalets/columnchart-datalet/columnchart-datalet.html](http://deep.routetopa.eu/deep_1_11/COMPONENTS/datalets/columnchart-datalet/columnchart-datalet.html)"][/datalet]

P

Conteggio parole: 0



## 7 USER INTERFACE AND ADVANCED SUPPORT

*“Focus on the needs of 80 percent of your users. When you do this, most people won’t have to supply any settings because the app is already set up to behave the way they expect.” Apple guidelines*

In SPOD, the Controllet is a main and fundamental component to visualize data and then for starting a discussion around the visual representation with charts, interactive maps. Many comments were received by the pilots to display data in a simple way with, once you choose the dataset (see table below). In Warsaw meeting<sup>21</sup>, held in May 25 2016, decision has been taken to work on support of creation of charts.

Requirement	Notes	DL 2.4	DL 3.1	DL 4.1	DoW
Open data access					
Advanced support for charts	When presented with the choice of the charts, the Controllet will present some suggestions on the right kind of charts to be used for the dataset selected, based on the popularity (how many users have used that visualization chart for that dataset) but also on similarity of charts and on heuristics on content, besides (if present) the use of metadata	UC20: View Chart and UC21: View Infographic (see in "Special Requirements")		"Usability" (sec. 2.2.1.3)	

Figure 32: Requirement for advanced support for charts highlighted in DL.4.1

Pilot	Feedback
Groningen (In SPOD v.0.3)	I found it easier to make a graph using the CKAN API, but when I finally managed to produce something, there was not a way to link it to publish my visualization on SPOD!
Groningen (In SPOD v.0.3)	In the MySpace step 1 to create a datalet is pretty straightforward for me as a ‘moderately experienced user’. Step 2 remains problematic in the sense of what records you should choose in order to be able to create a useful datalet. I know this is mainly a case of quality of datasets, however it could be useful to put a ? sign which displays what are useful records to choose, or referring the user to go back to an overview of the dataset he is using, in order to check which variable means what
Groningen (In SPOD v.0.3)	The Group By feature seems new but it is unclear to me what to do with it. Again a ? with some explanation would be useful. If I follow my instinct and use it to group the main variable on the Y-axis (TotaleBevolking/TotalPopulation) by a feature of this main variable, namely the part of the population aged 0-20 (k_Otot20jaar_2) I get an incomprehensible datalet preview: After playing around, I get some of the features, but fail to see the use of it in this particular dataset, for example the average amount of people aged 0 – 20 in Groningen from 2014-2040. However, I can see this would be a useful feature for datasets of bigger quality. Something I miss to find however is the possibility to add several variables in one visualization. Thus for example I want to see the visualization of population decline in Groningen from 2014 – 2040, with the variables: Total Population, population aged 0-20, population aged 20-60, population aged 60+, all shown in the same visualization, for example a line chart. This would be very useful and is also something that is already done in other data visualizations, e.g. Excel,

<sup>21</sup> Link available at <http://service.routetopa.eu:8000/d/481c8b5f9c/>

Groningen <sup>22</sup> (In SPOD v.1.4)	Loading a dataset can take minutes without anything happening. It seems to the student that the file may be too big, but this should be depicted in the tool. A statement saying this dataset is too big. Now you are just waiting and do not know what is happening.(#56)
Groningen <sup>22</sup> : (In SPOD v.1.4)	It is really hard to work with the datasets, because all the datasets are randomly described in a very long list. There are a lot of datasets that have nothing to do with the Groningen Project.
Groningen <sup>22</sup> : (In SPOD v.1.4)	If you want to filter on for example region when creating a datalet, the selection does not allow to see all the options because you cannot scroll down
Groningen <sup>22</sup> : (In SPOD v.1.4)	An important question of the students is how it is possible to create a graphic with two lines (e.g. different variables in the same graphic). For example: population total, population 18-30, population 31-45, population 46-60 all in one graphic.
Groningen <sup>22</sup> : (In SPOD v.1.4)	It is unclear to the students how they can change the labels in the graphs. From the database standard names are being used that are not always correct. The question is how these can be changed for the datalets.
Den Haag and Utrecht	Add data visualization - what is 'tree map view'? If I click it now I see a very small picture of explore data providers, hardly readable)

Table 6; Feedback from Pilot about Controllet

## 7.1 DATA VISUALIZATION: A REVIEW OF STATE OF THE ART

The purpose of this section is to review and analyze the state of the art in Visualization tools. First, we talk about metrologies for choosing the right chart and the taxonomy of interactive dynamics. Then we focus on two famous asynchronous collaborative visualization systems and other two minor tools. Finally, we list the best practices with data visualization.

### 7.1.1 METHODOLOGY FOR CHOOSING THE RIGHT GRAPH

Many books on graphing devote much attention to graphical integrity and readability, but little or none to graph selection (why?).

Professionals (engineers, scientist, manager, etc.) and others typically use the same few graph type for all their data sets, regardless of the amount and nature of their data. When asked how else they could graph the same data, they usually do not have a clue. Yet when shown a different graphical representation (new to them or not) of the same data, they recognize it as insightful; *“they just didn’t think of graphing it that way”*. (IEEE 2002)

[5] develops a methodology to help engineers, scientists and managers to enable to choose “the right graph”, and highlights the following three criteria:

- Structure of the data set (e.g., quantity (the number) and quality (the type of variable)
- Intended use of graph (e.g., analysis or answering question (personal) until conveying messages (publication) ): a graph should be not perfect but “optimal” for their intended use
- Research question: main generic categories are
  - Comparison (among individual data)
  - Distribution (data along scale)
  - Correlation (between variable)
  - Evolution (over time of a variable)

Each graph makes answering some questions easier and other questions harder.

<sup>22</sup> Usage for five weeks during Population Decline Challenge, April, May 2016

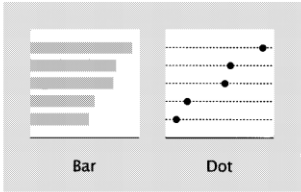
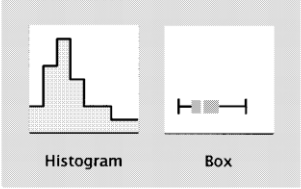
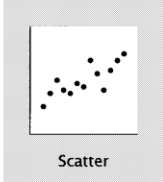
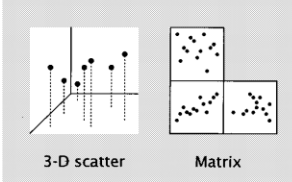
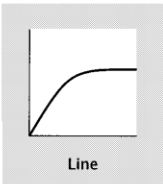
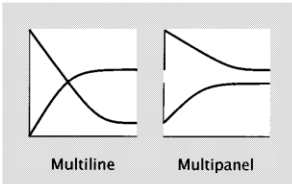
	One continuous variable	Two...	Three continuous variables
<b>Comparison</b>	 Bar      Dot		
<b>Distribution</b>	 Histogram      Box		
<b>Correlation</b>		 Scatter	 3-D scatter      Matrix
<b>Evolution</b>		 Line	 Multiline      Multipanel

Figure 33: An initial two-entry table where the column represents the structure of data set (e.g., one, two or three continuous variables) and the rows represent the research question grouped in the four categories (e.g., comparison, distribution, correlation and evolution)

### 7.1.2 TAXONOMY OF INTERACTIVE DYNAMICS

[6] The increasing scale and availability of digital data provides an extraordinary resource for informing public policy, scientific discovery, business strategy, and even our personal lives. To get the most out of such data, however, users must be able to make sense of it: to pursue questions, uncover patterns of interest, and identify (and potentially correct) errors. In concert with data-management systems and statistical algorithms, analysis requires contextualized human judgments regarding the domain-specific significance of the clusters, trends, and outliers discovered in data.

Visualization provides a powerful means of making sense of data. By mapping data attributes to visual properties such as position, size, shape, and color, visualization designers leverage perceptual skills to help users discern and interpret patterns within data. A single image, however, typically provides answers to, at best, a handful of questions. Instead, visual analysis typically progresses in an iterative process of view creation, exploration, and refinement. Meaningful analysis consists of repeated explorations as users develop insights about significant relationships, domain-specific contextual influences, and causal patterns. Confusing widgets, complex dialog boxes, hidden operations, incomprehensible displays, or slow response times can limit the range and depth of topics considered and may curtail thorough deliberation and introduce errors. To be most effective, visual analytics tools must support the fluent and flexible use of visualizations at rates resonant with the pace of human thought.

We show a taxonomy of interactive dynamics that contribute to successful analytic dialogues. The taxonomy consists of 12 task types grouped into three high-level categories, as shown in figure 23: (1) data and view

specification (visualize, filter, sort, and derive); (2) view manipulation (select, navigate, coordinate, and organize); and (3) analysis process and provenance (record, annotate, share, and guide). These categories incorporate the critical tasks that enable iterative visual analysis, including visualization creation, interactive querying, multiview coordination, history, and collaboration. Validating and evolving this taxonomy is a community project that proceeds through feedback, critique, and refinement.

<b>Data &amp; View Specification</b>	<b>Visualize</b> data by choosing visual encodings. <b>Filter</b> out data to focus on relevant items. <b>Sort</b> items to expose patterns. <b>Derive</b> values or models from source data.
<b>View Manipulation</b>	<b>Select</b> items to highlight, filter, or manipulate them. <b>Navigate</b> to examine high-level patterns and low-level detail. <b>Coordinate</b> views for linked, multi-dimensional exploration. <b>Organize</b> multiple windows and workspaces.
<b>Process &amp; Provenance</b>	<b>Record</b> analysis histories for revisitation, review and sharing. <b>Annotate</b> patterns to document findings. <b>Share</b> views and annotations to enable collaboration. <b>Guide</b> users through analysis tasks or stories.

Figure 34: Taxonomy of interactive dynamics for visual analysis

### 7.1.1.3 ASYNCHRONOUS COLLABORATIVE VISUALIZATION SYSTEMS

Many Eyes and Tableau are systems for creating visualization, data analysis, sharing and two popular Web-base, collaborative visual analytics systems [26].

#### 7.1.1.3.1 MANY EYES (IBM)

“Many Eyes” is a collaborative data visualization and a service public web site that provides an unique environment of open web-based collaboration, launched on January 23th, 2007 and Many Eyes is retired by IBM on June 2015 [7], [8].

##### **Overview**

The main activities on the site are to upload data, construct visualizations and add comments. Figure 36 shows a dataset browsing pages with a table displaying metadata about latest uploaded datasets, metadata about each dataset, the contributor’ name, size in byte links to existing visualizations with that dataset.

Three different areas in the design of in Many Eyes are been considered:

- Information visualization
- end-user data collection and manipulation
- and the social aspects of collaborative analysis

Types of visualization in Many Eyes includes standard business graphics (e.g., bar chart), academic technique (e.g., treemap) and experimental components (bubble chart and tag clouds)

##### **Overview**

- Explore (Visualizations, datasets, Comment and Topic center)
- Participate (Create a visualization, Upload a dataset, Create a topic center, register )





Figure 35: Many Eyes: an overview

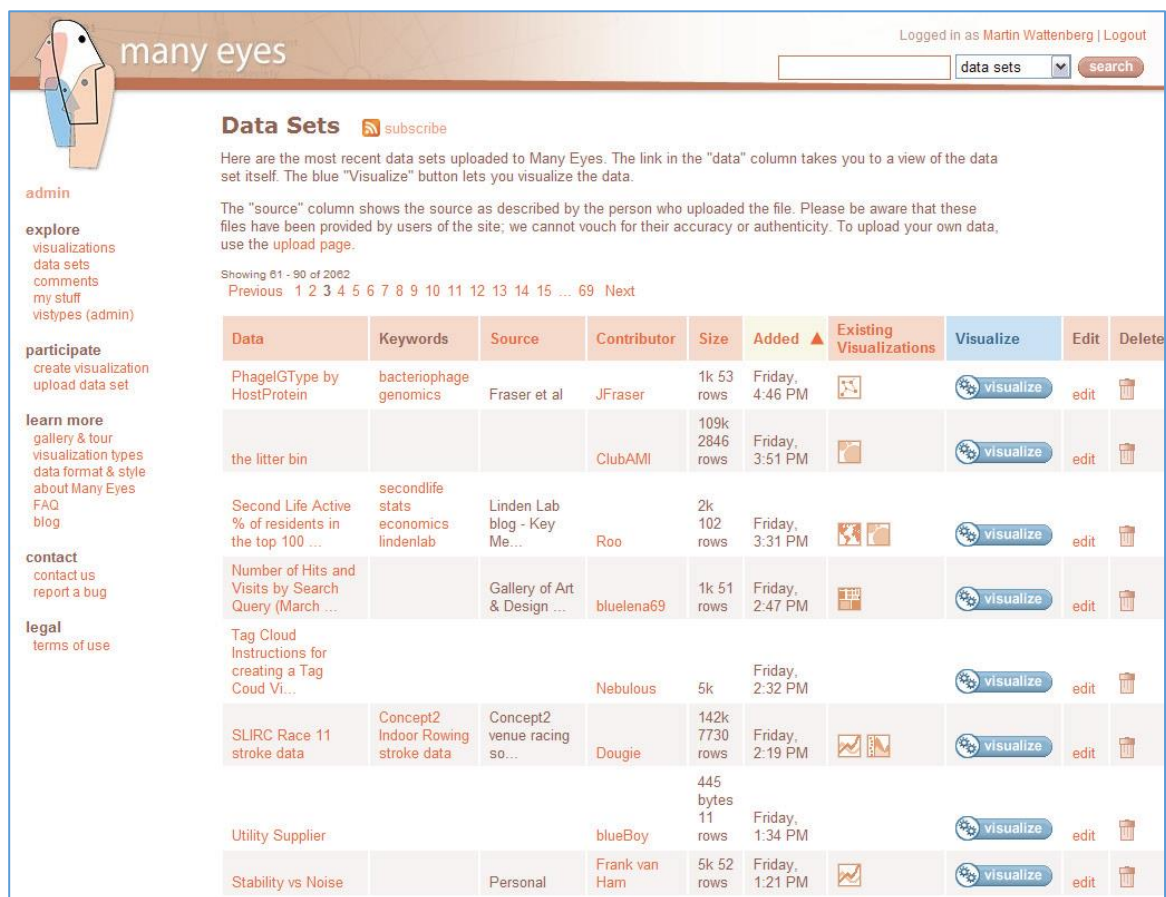


Figure 36: Many Eyes: an overview of all datasets uploaded to the site with existing visualizations for each dataset

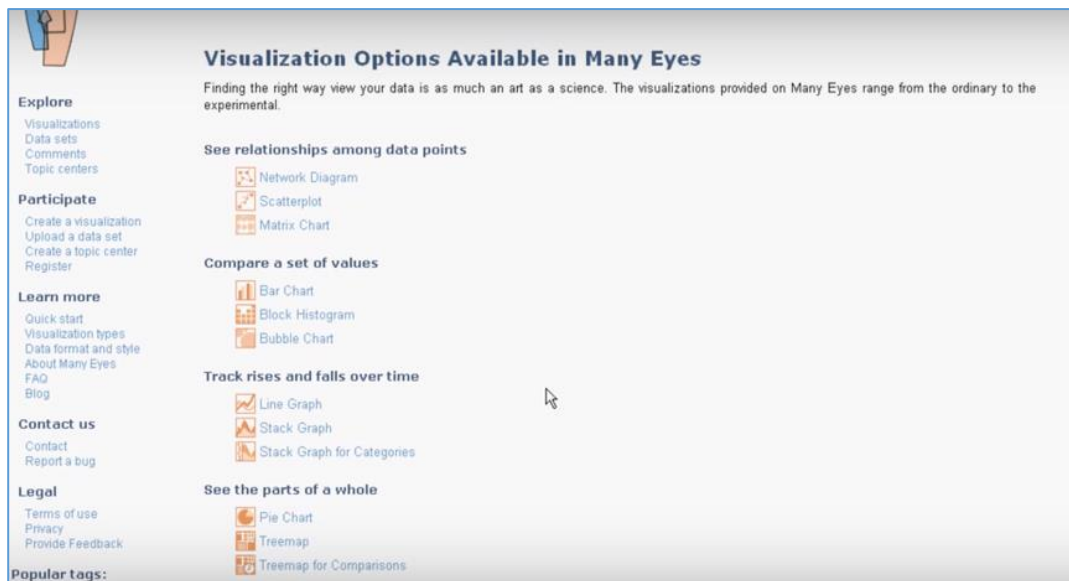


Figure 37: Grouping by “Relationship among data points” (chart types: Scatterplot, Network diagram, Matrix Chart), Compare a set of values (chart types: Bar Chart, Block Histogram and Bubble Chart), Track rises and falls over time (chart types: Line Graph Stack Graph and Stack Graph for Categories) and See the parts of whole (chart types: Pie Chart etc...)

### 7.1.3.2 TABLEAU

Tableau is a commercial business intelligence (BI) software tool to allow interactive, visual analysis of data. The end-user with no programming or Database experience can explore and visualize the dataset with few drag-and-drop operation; furthermore s/he can integrate large, heterogeneous external data source [7], [9]. The tool avoids the burden of writing queries and helps the task question of a user with a visual drag-and-drop operations without the experience/knowledge of queries or programming, that is the user creates her/his visualization by dragging and dropping fields from their databases onto a visual canvas, generating VizQL (Visual Query Language is a structured query language with support for rendering graphic) statements to reach the requested visualization.

The Cycle of the Visual Analysis” showed in the following figure:

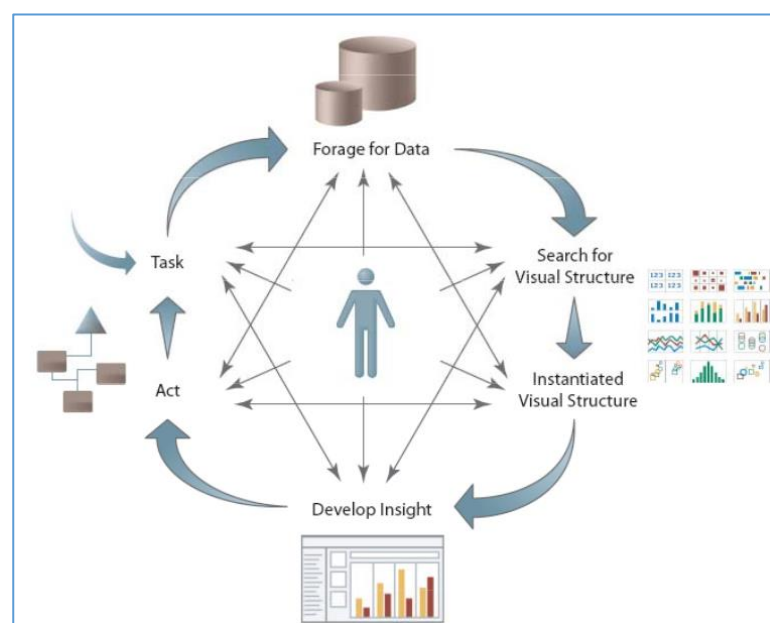


Figure 38: Sense-making model

This figure shows the relevance of the visualization in the sense-making model [10]: a theory where an user follows this cycle: search for, organize and generate new knowledge. The start point is the “knowledge worker” (shown at the center) with a task or question to address to. The user search for through databases with relevant data for the analysis task. Another step, s/he searches an appropriate visual structure and change visualization. Once the necessary insight is obtained, the user can then make an informed decision and take action

This cycle is driven by the user and aims to visualization flexible system to support user feedback and alternative path can help the user’s exploratory tasks.

Criticism are addressed to most visualization tool (e.g., Fusion Table [11], [9], M any Eyes [12] because they treat this cycle as a single pipeline and offer a few interaction with the user. The need of the user to analytic questions over multiple data source. A new feature, “*data blending*”, introduced in the cycle of visual Analysis” to allow to the user to combine and visualize data from multiple different data source on-the-fly [11].

The tools integrates rich public dataset (e.g. data.gov), data from external source (e.g., Web or corporate data, Excel spreadsheets) into enterprise data warehouse to analysis.

In Tableau version 6.0 the new data blending feature to simply the integration data from heterogeneous sources filtering data with several options, application of filtering based on aggregate condition data.

[9] shows challenges to visualization solution with the growth of the data volumes. Caching and prefetching techniques were applicated because the query processing and it illustrates key data processing components in Tableau (the system uses several level of caching).

## Overview of Tableau System

How it works:

- First, a type of connection file (Excel, Text file, access other files) , or a server (Tableau Server, Mysql, Oracle, more Server) or saved data source is chosen
- More table with join operation on the data allows to create a dataset (see Figure 39)
- “Show me” dialog of available charts is available to select a chart types (see Figure 40)
- Place data variable among “Data” tab on “shelves” corresponding to visual encodings such as spatial position, size, shape, and color. The visual specification is then translated into an underlying formal grammar that determines both the visualization design and corresponding queries to a database. This approach leverages the expressiveness of formal grammars while avoiding the need for programming. Another advantage is that formal grammars can be augmented with automated design facilities (see Figure 40).

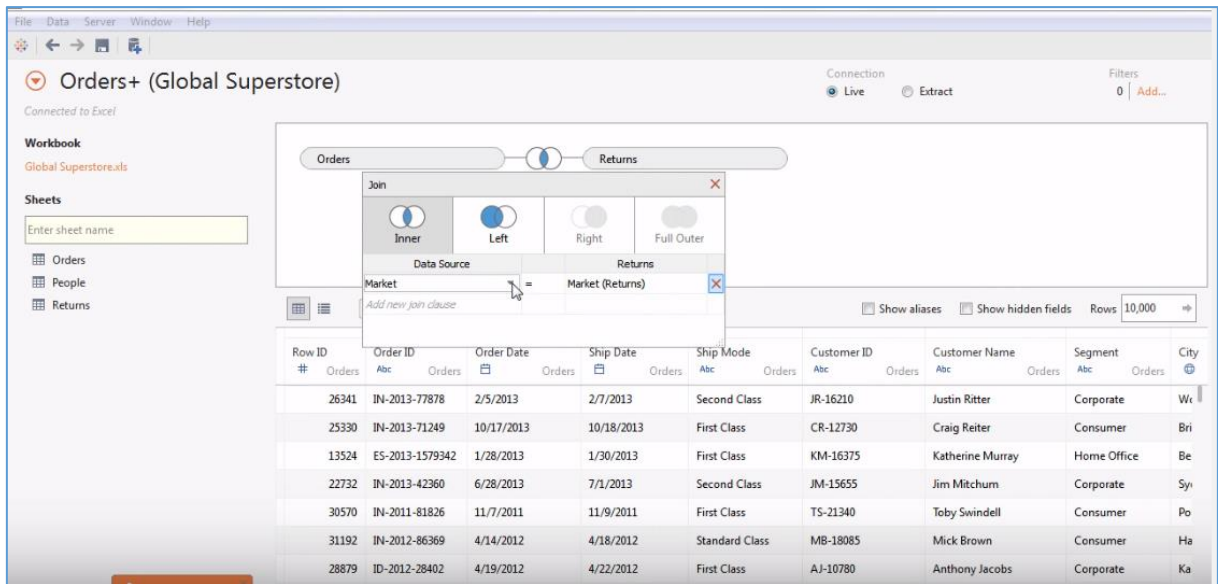


Figure 39: On the left side, show the data in more table and a join operation combines fields from two or more tables in a database.

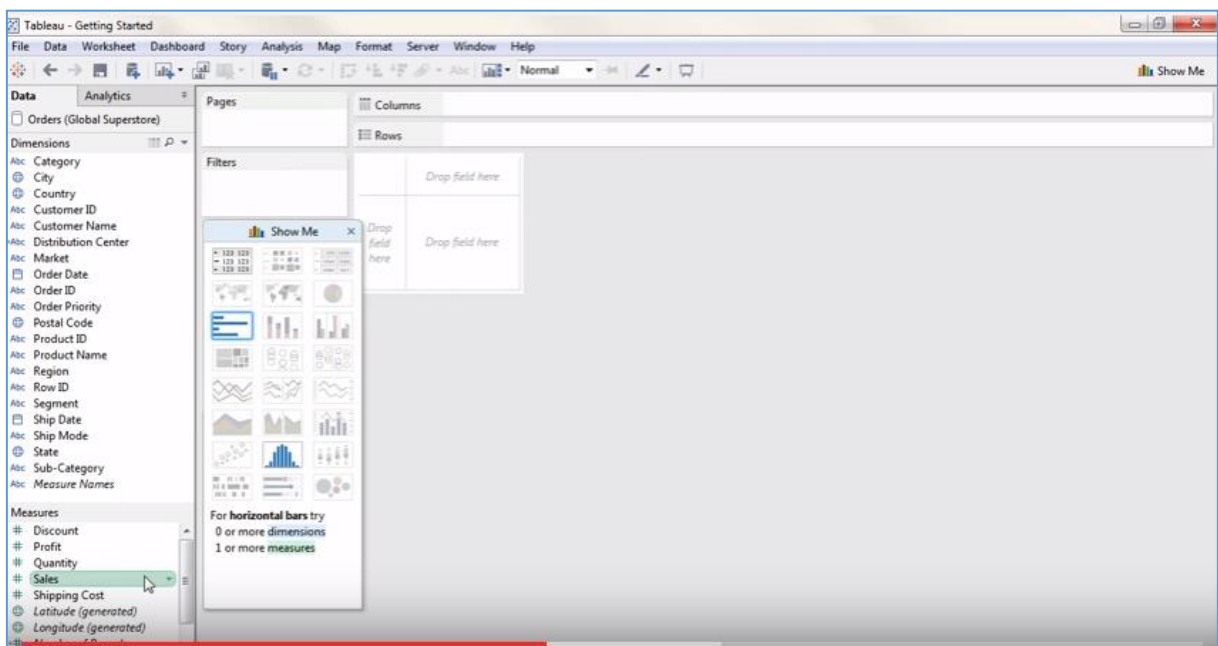


Figure 40: Drag and drop operation to visualize a chart: The visual specification is then translated into an underlying formal grammar that determines both the visualization design and corresponding queries to a database.

## 7.1.4 OTHER DATA VISUALIZATION TOOLS

### 7.1.4.1 CHART CHOOSER TOOL

Chart Chooser, based on Andrew Abela's flowcharts ("Juice Analytics"<sup>23</sup>)

- InfoVis tool from Juice Analytics
- Use filter (All, Comparison, Distribution, Trend, relationship and table) to find a right chart (e.g., filter with comparison, distribution and composition => bar chart)

<sup>23</sup> <http://labs.juiceanalytics.com/chartchooser/index.html>

- Download the template for Excel or PowerPoint

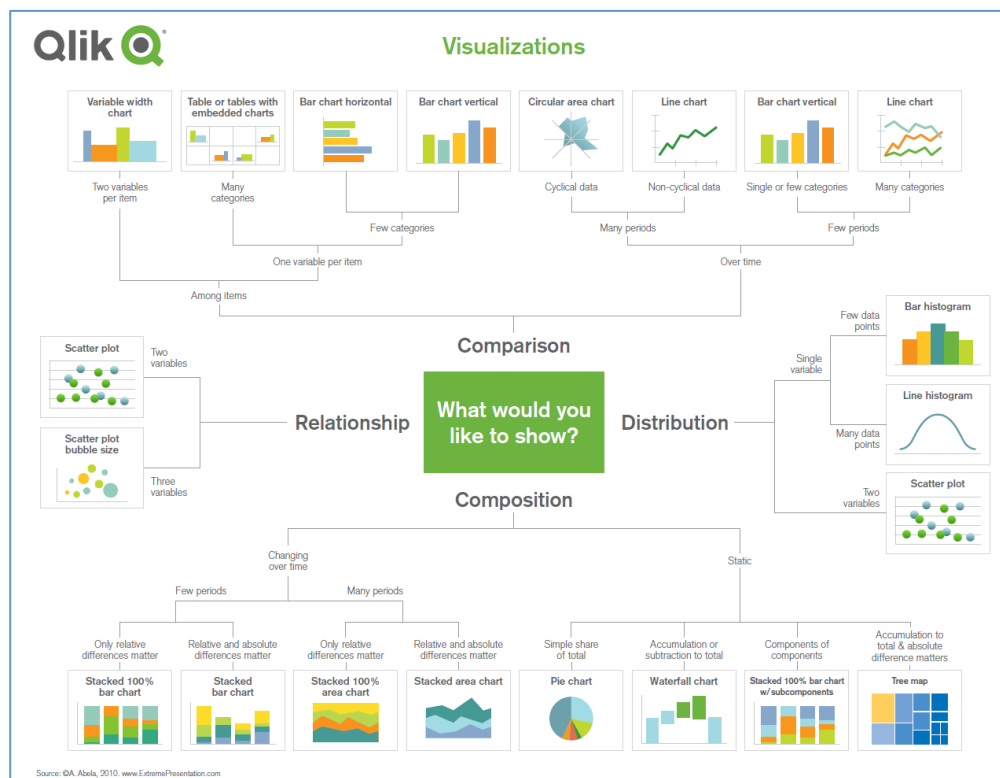


Figure 41: Visualizations grouping for categories (Comparison, Relationships, Distribution and Composition)

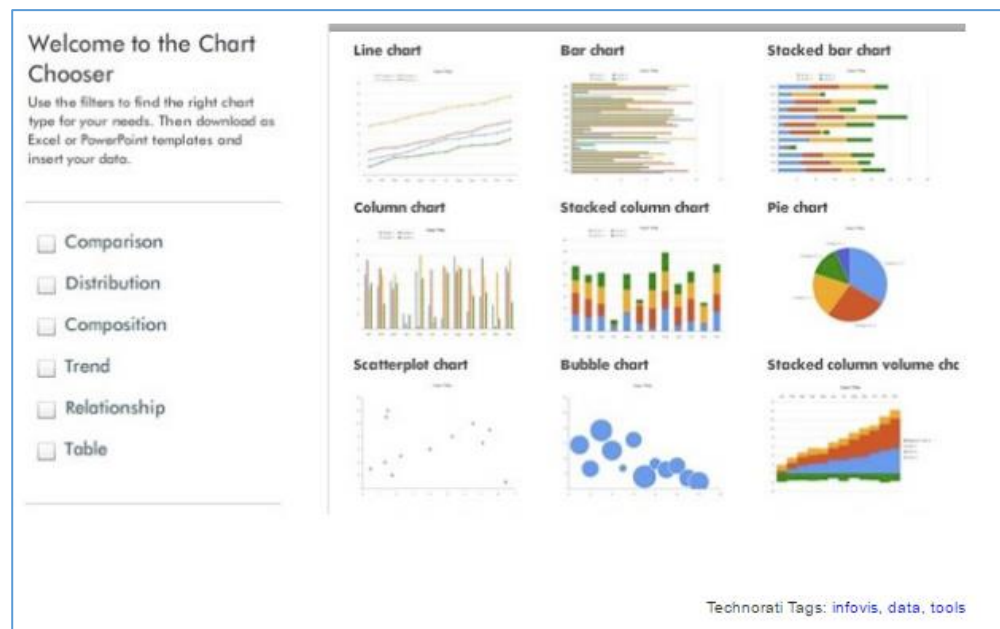


Figure 42: Chart Chooser

#### 7.1.4.2 CHART TAMER: ADD-IN IN EXCEL

Chart Tamer 1.0 ("From Theory to Practice Workshop Proposal" by Stephen Few), developers: BonaVista System

24

<sup>24</sup> <http://bonavista-systems.software.informer.com/>

Align Excel's chart with the best practices of data visualization [13].

- Limit the list of available charts to the few that are most useful
- It provides a simple new interface for selecting an appropriate chart, which guides us to the right selection
- New three useful chart in Excel (dot plots, strip plot and box plot)

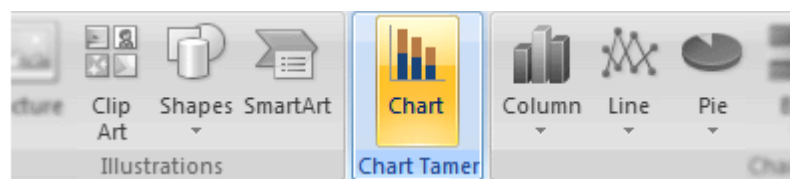
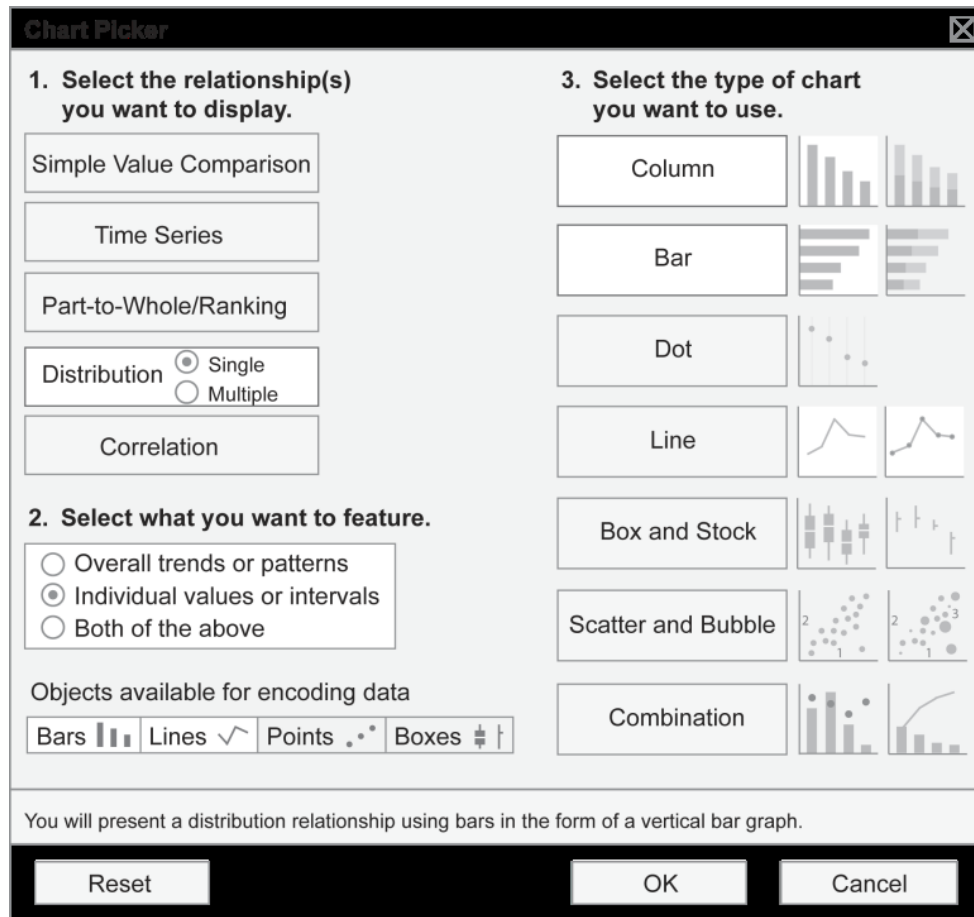


Figure 43: Chart Tamer: mock-ups for an Excel add-in (Workshop by Stephen Few, in 2007- 2009) and chart tamer icon (below) of the excel menu



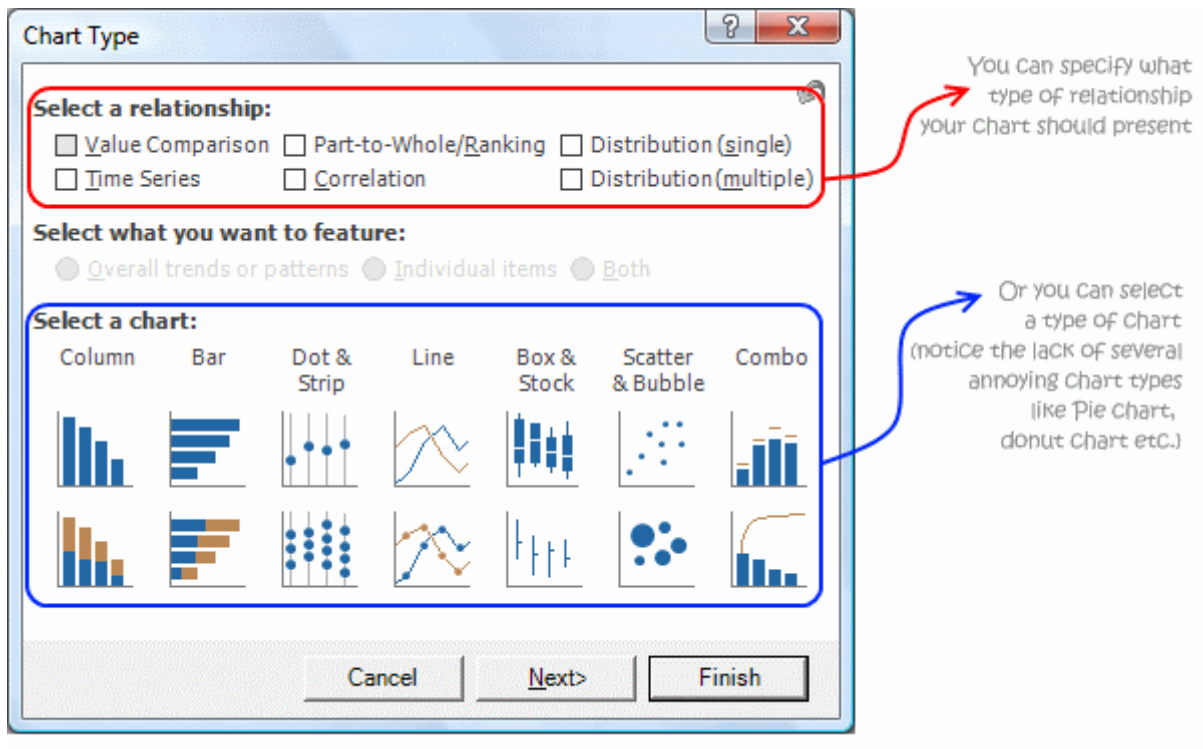


Figure 44: Two ways to get started in Chart Tamer: (1) specify the type of relationship of chart and (2) recommended charting options <sup>25</sup>

Two ways to get started (see Figure 44: Two ways to get started in Chart Tamer: (1) specify the type of relationship of chart and (2) recommended charting options Figure 44) and also color picker tool that provides contrast and comparison (see Figure 45):

1. Specify the type of relationship of the chart and the Chart Tamer recommends the chart to use (for example for distribution the chartings options are Column, Bar, Line, Stacked Column, Stacked Bar and Dotted Line)
2. Select the type chart directly

<sup>25</sup> <http://chandoo.org/wp/2009/05/04/review-chart-tamer/> (Posted on may 4<sup>th</sup>, 2009 )

The default colors of Chart tamer are well chosen to provide excellent contrast and comparison

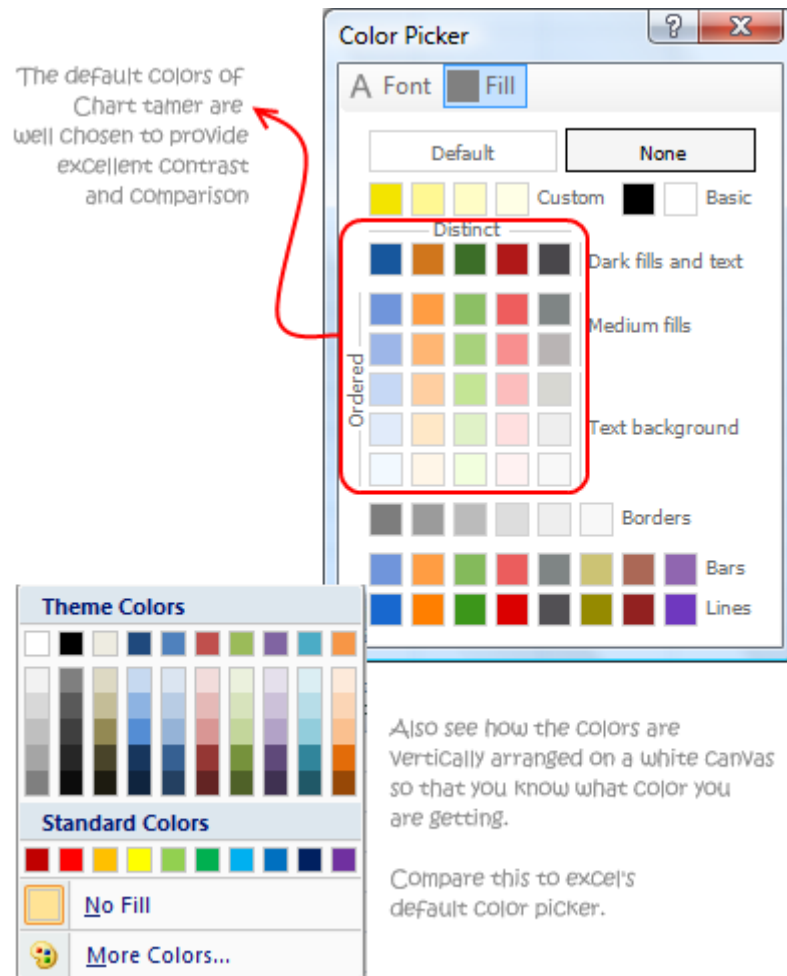


Figure 45: Chart tamer: Color Picker tool to provide contrast and comparison:



### 7.1.5 WHICH CHART OR GRAPH IS RIGHT FOR YOU?

[14] is a white paper listing the best practice recommendations to transform the data into an effective/visualization:

Also consider...	Charts	When to use /Categories
<ul style="list-style-type: none"> <li>• Include multiple bar charts on a dashboard.</li> <li>• Add color to bars for more impact.</li> <li>• Use stacked bars or side-by-side bars</li> <li>• Combine bar charts with maps</li> <li>• Put bars on both sides of an axis.</li> </ul>	Bar chart	Comparison data across categories Effective with numerical data to split into different category to see the trend within the data:
<ul style="list-style-type: none"> <li>• Combine a line graph with bar charts.</li> <li>• Shade the area under lines.</li> </ul>	Line chart	<i>Trend:</i> Viewing trends in data over time.
<p>Show relative proportions – or percentages – of information</p> <p>Other: • Limit pie wedges to six. If you have more than six proportions to communicate, consider a bar chart. It becomes too hard to meaningfully interpret the pie pieces when the number of wedges gets too high.</p>	Pie chart	Showing proportions.
<ul style="list-style-type: none"> <li>• Use maps as a filter for other types of charts, graphs, and tables. Combine a map with other relevant data then use it as a filter to drill into your data for robust investigation and discussion of data.</li> <li>• Layer bubble charts on top of maps. Bubble charts represent the concentration of data and their varied size is a quick way to understand relative data. By layering bubbles on top of a map it is easy to interpret the geographical impact of different data points quickly.</li> </ul>	Map	Showing geocoded data.
<p>Also consider:</p> <ul style="list-style-type: none"> <li>• Add a trend line/line of best fit. By adding a trend line the correlation among your data becomes more clearly defined.</li> <li>• Incorporate filters. By adding filters to your scatter plots, you can drill down into different views and details quickly to identify patterns in your data.</li> <li>• Use informative mark types. The story behind some data can be enhanced with a relevant shape</li> </ul>	Scatter plot	Investigating the relationship between different variables.
<ul style="list-style-type: none"> <li>• Accentuate data on scatter plots: By varying the size and color of data points, a scatterplot can be transformed into a rich visualization that answers many questions at once.</li> </ul>	Bubble chart	<ul style="list-style-type: none"> <li>• Showing the concentration of data along two axes. Examples: sales concentration by product and geography, class attendance by department and time of day.</li> </ul>

<ul style="list-style-type: none"> <li>• Overlay on maps: Bubbles quickly inform a viewer about relative concentration of data. Using these as an overlay on map puts geographically-related data in context quickly and effectively for a viewer.</li> </ul>		
<p>Test different groupings of data. When you are exploring your data and looking for groupings or “bins” that make sense, creating a variety of histograms can help you determine the most useful sets of data.</p> <ul style="list-style-type: none"> <li>• Add a filter. By offering a way for the viewer to drill down into different categories of data, the histogram becomes a useful tool to explore a lot of data views quickly</li> </ul>	Histogram	<ul style="list-style-type: none"> <li>• Understanding the distribution of your data. Examples: Number of customers by company size, student performance on an exam, frequency of a product defect.</li> </ul>
<p>Vary the size of squares. By adding a size variation for your squares, heat maps let you know the concentration of two intersecting factors, but add a third element. For example, a heat map could reveal a survey respondent’s sports activity preference and the frequency with which they attend the event based on color, and the size of the square could reflect the number of respondents in that category.</p> <ul style="list-style-type: none"> <li>• Using something other than squares. There are times when other types of marks help convey your data in a more impactful way.</li> </ul>	Heat maps	<p>Showing the relationship between two factors. Examples: segmentation analysis of target market, product adoption across regions, sales leads by individual rep.</p>

#### 7.1.6 BEST PRACTICES WITH DATA VISUALIZATIONS

[15] is a guide thought the most common charts and visualizations to help the right presentation for data with design tips to make sure to avoid mistake [16], [17].

Analyse the type of data to visualize and their relationships before understand visualizations and the most common type of data illustrated in the Figure 46 and the most common data types and relationships and different ways of data visualization in the Figure 47.

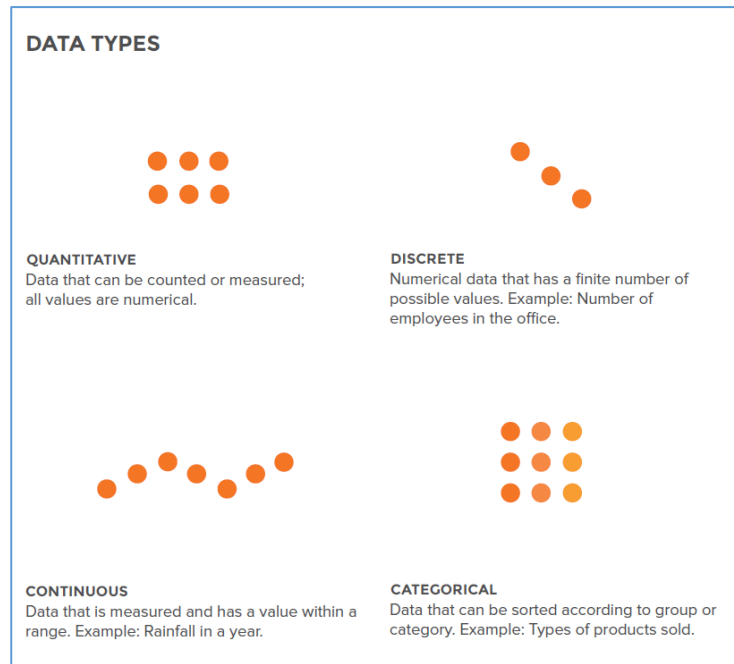


Figure 46: Data type and their relationship

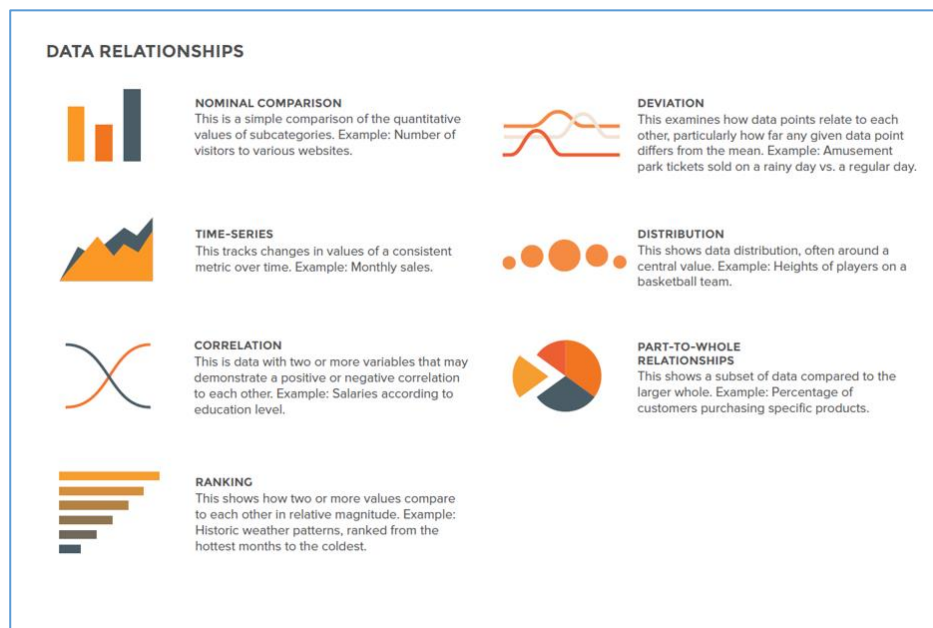


Figure 47: The most common data types and relationships and different ways of data visualization

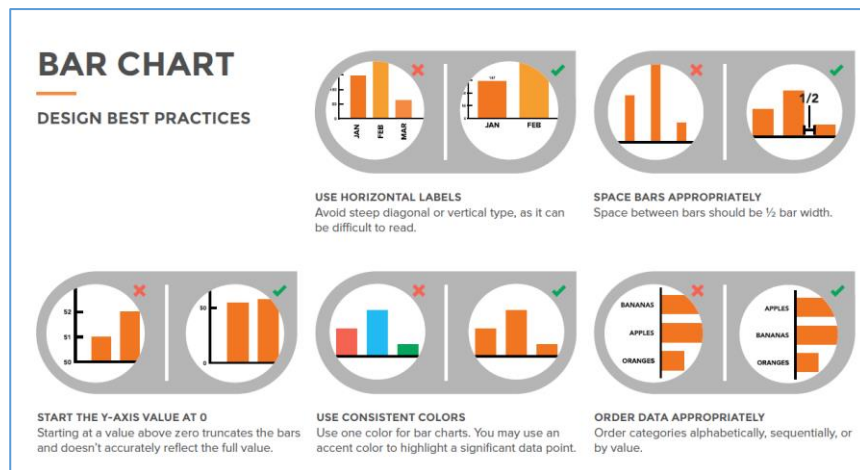


Figure 48: Design best practises: an example bar chart

### 7.1.7 SELECT THE RIGHT GRAPH FOR YOUR MESSAGE

[18] describes seven quantitative relationships and also identifies the best ways to encode their values . Knowing what you do now about the relative strengths and weaknesses of points, lines, and bars for encoding values, try determining which objects would work best before looking at the encoding method guidelines that I provide.

Type/Description	Encoding Methods	Example
<b>Nominal Comparison</b> A simple comparison of the categorical subdivisions of one or more measures in no particular order	<ul style="list-style-type: none"> <li>Bars only (horizontal or vertical)</li> </ul>	
<b>Time Series</b> Multiple instances of one or more measures taken at equidistant points in time	<ul style="list-style-type: none"> <li>Lines to emphasize overall pattern</li> <li>Bars to emphasize individual values</li> <li>Points connected by lines to slightly emphasize individual values while still highlighting the overall pattern</li> <li>Always place time on the horizontal axis</li> </ul>	
<b>Ranking</b> Categorical subdivisions of a measure ordered by size (either descending or ascending)	<ul style="list-style-type: none"> <li>Bars only (horizontal or vertical)</li> <li>To highlight high values, sort in descending order</li> <li>To highlight low values, sort in ascending order</li> </ul>	
<b>Part-to-Whole</b> Measures of individual categorical subdivisions as ratios to the whole	<ul style="list-style-type: none"> <li>Bars only (horizontal or vertical)</li> <li>Use stacked bars only when you must display measures of the whole as well as the parts</li> </ul>	

<b>Deviation</b> Categorical subdivisions of a measure compared to a reference measure, expressed as the differences between them	<ul style="list-style-type: none"> <li>• Lines to emphasize the overall pattern only when displaying deviation and time-series relationships together</li> <li>• Points connected by lines to slightly emphasize individual data points while also highlighting the overall pattern when displaying deviation and time-series relationships together</li> <li>• Bars to emphasize individual values, but limit to vertical bars when a time-series relationship is included</li> <li>• Always include a reference line to compare the measures of deviation against</li> </ul>	
<b>Frequency Distribution</b> Counts of something per categorical subdivisions (intervals) of a quantitative range	<ul style="list-style-type: none"> <li>• Vertical bars to emphasize individual values (called a <i>histogram</i>)</li> <li>• Lines to emphasize the overall pattern (called a <i>frequency polygon</i>)</li> </ul>	
<b>Correlation</b> Comparisons of two paired sets of measures to determine if as one set goes up the other set goes either up or down in a corresponding manner, and if so, how strongly	<ul style="list-style-type: none"> <li>• Points and a trend line in the form of a scatter plot</li> <li>• Bars may be used, arranged as a <i>paired bar graph</i> or a <i>correlation bar graph</i>, if scatter plots are unfamiliar</li> <li>• (Note: For descriptions of these graphs, see my book <i>Show Me the Numbers</i>.)</li> </ul>	

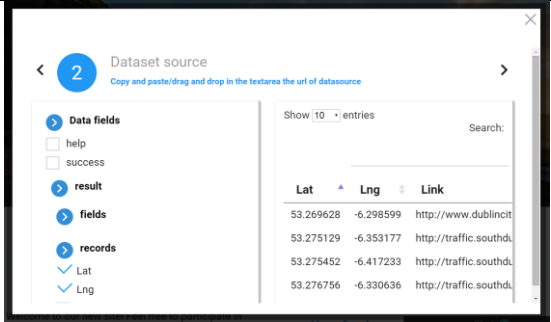
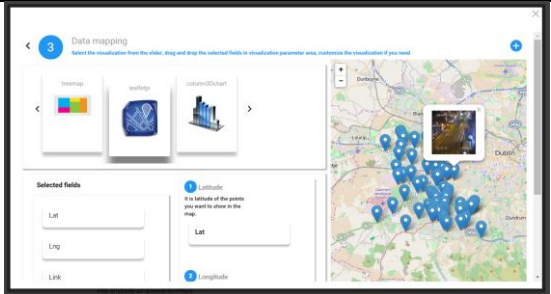
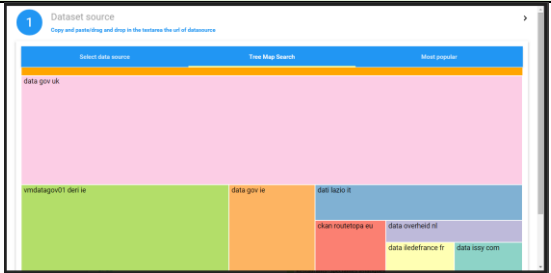
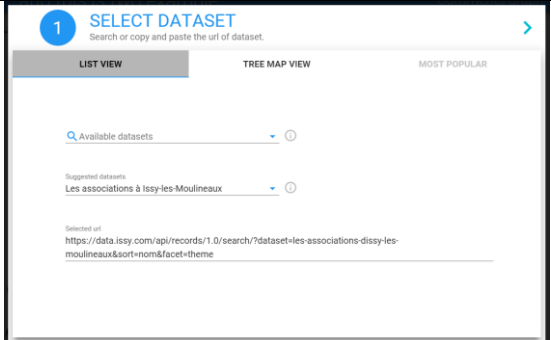
Figure 49: Seven quantitative relationships and the correspondent encode

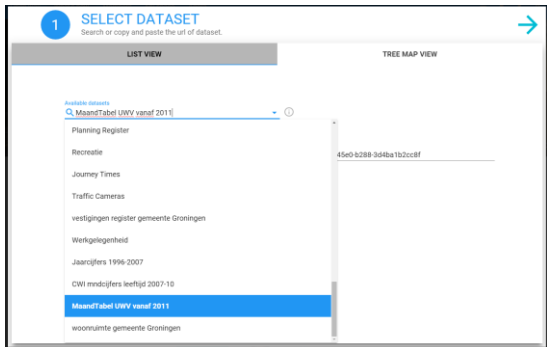
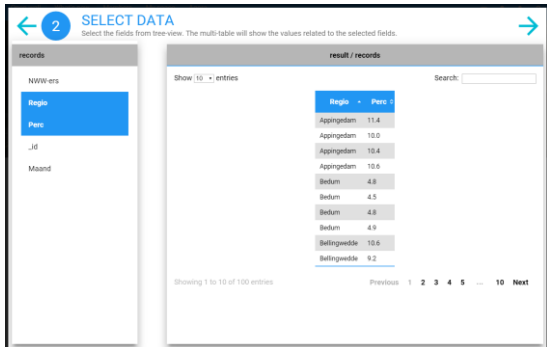
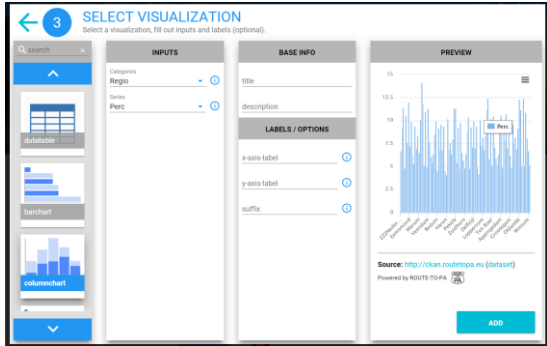

## 7.2 A NEW DESIGN OF USER INTERFACE FOR THE CONTROLLET

The new design of the User Interface with new features of the Controllet takes in account of the previous experiences, feedback received from each Pilot.

An overview of lifecycle of the Controllet in the SPOD platform from version 0.1 until version v.1.12 follows (see Figure 50):

SPOD “Antares” (0.1)		
Step 1	Initially to select a dataset only a simple drop down list was available, in addition to the name and the url no other information about the dataset was displayed.	

Step 2	In the second step, to select dataset fields was implemented a tree view, where all section of the dataset, included metadata and extra fields were present. It was difficult to understand and could easily lead to error the average user.	
Step 3	In the third step was only possible to select the chart and fields, without checks and support information. It was used drag and drop technique revealed later not user friendly in this specific case.	
<b>SPOD "Diadem" (0.4)</b>		
Step 1	A tree map was developed to simplify the search of the dataset, then revealed not useful and difficult to understand.	
<b>SPOD "Release Candidate" (0.6)</b>		
Step 1	Suggested dataset was introduced.	

SPOD "Alpha" (1.0)		
Step 1	List view and Treemap	
Step 2	The component tree-view on the left used to field selection has been replaced with a simple list, because thanks to an upstream work we can see the dataset like a single table.	
Step 3	Several changes have been introduced. The style has been improved, more information is available, and new features have been added.	
SPOD "Izar" (1.4)		
Step 3	Expert section for grouping by; personalize the charts with several themes (Base, DarkUnika, SandSignika, GridLight and SPOD)	

*SPOD "Jabbah" (1.5)*

Step 1	<p>A new interface in Step 1) to allow :</p> <ul style="list-style-type: none"> <li>(1) When a dataset is select, all its metadata are shown</li> <li>(2) Search in Treemap and in List view</li> </ul> <p>User interface improvements: (1) each color of a datasets identifies the correspondent provider (same color as the treemap)</p> <ul style="list-style-type: none"> <li>(2) Choice of datasets by Open Data provider (i.e.: CKAN, Issy-les-Moulineaux)</li> <li>(3) Search also the content of datasets, through the UltraClarity search portal for CKAN, by using, n Step 1) a new tab, named "Extended search"</li> </ul>
--------	---

Step 1	<p>A new interface in Step 1) to allow :</p> <ul style="list-style-type: none"> <li>(1) When a dataset is select, all its metadata are shown</li> <li>(2) Search in Treemap and in List view</li> </ul> <p>User interface improvements: (1) each color of a datasets identifies the correspondent provider (same color as the treemap)</p> <p>(2) Choice of datasets by Open Data provider (i.e.: CKAN, Issy-les-Moulineaux)</p> <p>(3) Search also the content of datasets, through the UltraClarity search portal for CKAN, by using, n Step 1) a new tab, named "Extended search"</p>
--------	--



Step 3	New datalets: stacked column, bar and area charts with several themes
--------	---

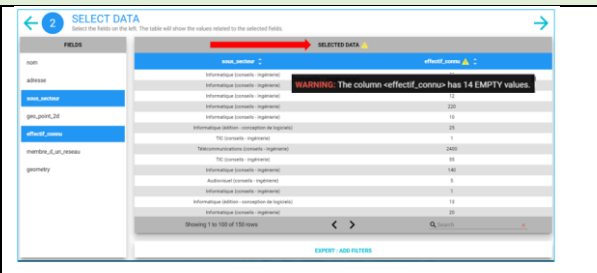
Step 3	New datalets: stacked column, bar and area charts with several themes
--------	---



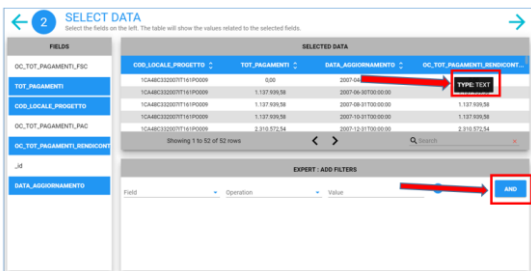
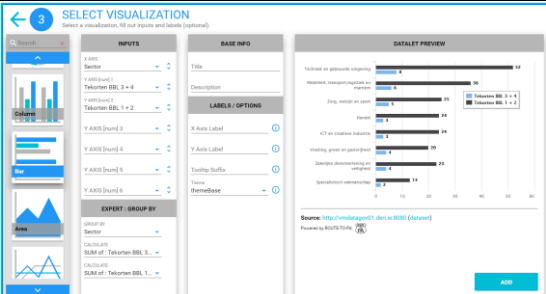
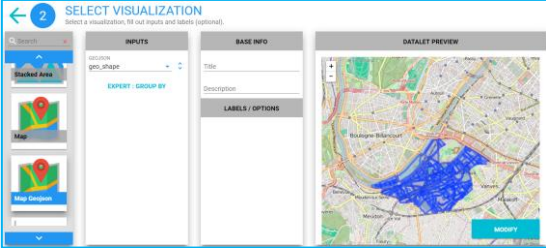
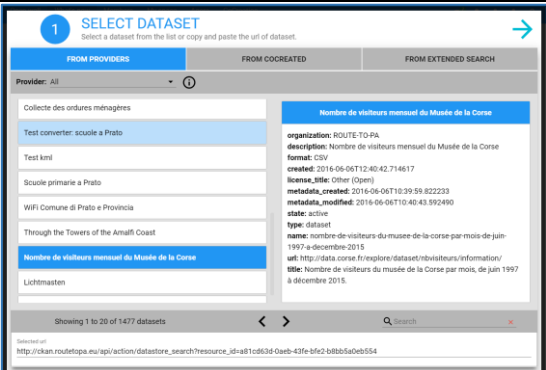
*SPOD "Kuma" (1.6)*

Step 2	<p>User Interface cleanup (possible work on support of creation of charts)</p> <p>(1) Data quality of a dataset: in the step 2) a further check on the data quality of a uploaded dataset highlighting error types in each field of the dataset and “?” white</p>
--------	---

Step 2	<p>User Interface cleanup (possible work on support of creation of charts)</p> <p>(1) Data quality of a dataset: in the step 2) a further check on the data quality of a uploaded dataset highlighting error types in each field of the dataset and “?” white</p>
--------	---





	<p>question mark appears over the fields of a dataset suggestions field type.</p> <p>(3) Multiseries: In step 3, multiseries charts are easier to create (this is a first step toward further improvements available in the next release.</p> <p>A new library (JSDataChecker<sup>26</sup>) has been included for recognise data types and analyse data quality.</p>	
Step 3	<p>A new important feature is the possibility to create chart with multi-series and then optionally stacked.</p>	
Step 3	<p>A new important feature is the possibility to create chart with multi-series and then optionally stacked.</p>	
<p><b>SPOD "Lesath" (1.7)</b></p>		
Step 1	<p>User Interface cleanup</p>	

<sup>26</sup> See paper "Support citizens in Visualising Open Data" in Sec. 0

**Step 2**

In addition to the selection of the dataset fields (vertical filter / filter columns), it's given the possibility for users to filter rows (horizontal filter) thanks to a new panel that allow to add SQL like conditions (tab 'filters'). Furthermore in the same panel but in a different tab is possible to use the SQL-like group by statement (tab 'group by'). In the third tab 'query' the current query generated by field selection, filters condition and group by statement is shown. We decided to hide by default these options in a dedicated section "expert mode" so as not to make the interface too complicated or confusing for inexperienced users.



Step 3

(1) Suggest right chart for selected data, coordinates types (e.g., number etc) with “Datalet Info” to explain each chart, “Datalet preview” shows a preview of the chart;

(2) Fill in “INPUTS” section with suggestion of the fields to choose

(3) Comparison charts

(4) Chart labelling Options (e.g., chart title, Legend Options, Axis Title, Stacked chart)

For each chart more options are available now depending on chart, for example ‘legend position’, ‘theme’, ‘show/hide data labels’ and so on.

On the right side the preview div has been replaced by a new panel with two tabs. The first one shows the preview of the datalet as before when it’s ready, while the second tab shows all the information about the selected chart that is the image of the chart, a brief description of the chart and when it’s recommended, the list of all inputs and options available with related descriptions.

SPOD “Regulus” (1.12)

Step 1

If a dataset have more than one resource the tab associated to the dataset become a drop down menu with a tab for each resource.

Figure 50: Controllet life cycle from SPOD v.0.1 to SPOD 1.12

### 7.3 THE DATALET WIZARD CONTROLLET

Starting from Alpha version many of the efforts to improve the interface and the user experience was focused on the *Datalet* wizard controllet (from now on *Controllet*) as a crucial and critical feature of SPOD and probably the most complex. The development was driven by two main factors: the study of the literature on data visualization and of other existing systems, and the user feedbacks which should never be underestimated. During this phase of development of the duration of about 10 months a lot of changes and improvements were introduced, from small graphical corrections up to completely new features.

We decided to keep the choice of the three steps because we can watch them as three independent operations and it's not too challenging or annoying for the users:

- dataset selection
- data selection and manipulation (select, filter and group by)
- data visualization

An effort was made to maintain the interface homogenous through the steps in regards using the same style, colours, positions and other graphical elements in order to not mislead the users. Indeed the users feedback were increasingly better until the usability was perceived really good for most of the users, we must do not forget that these steps are for nature complexes for most people and we can't image that all users can simply use the *Controllet* without a minimal knowledge of the context, and moreover that the quality of the dataset is fundamental and impacts strongly on the difficulty of create a significant visualization or just the chart that we would like.

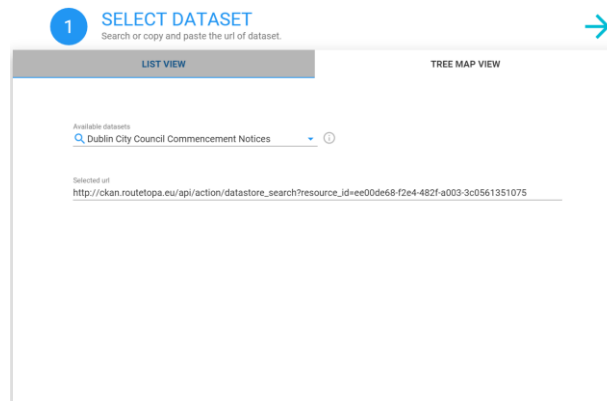
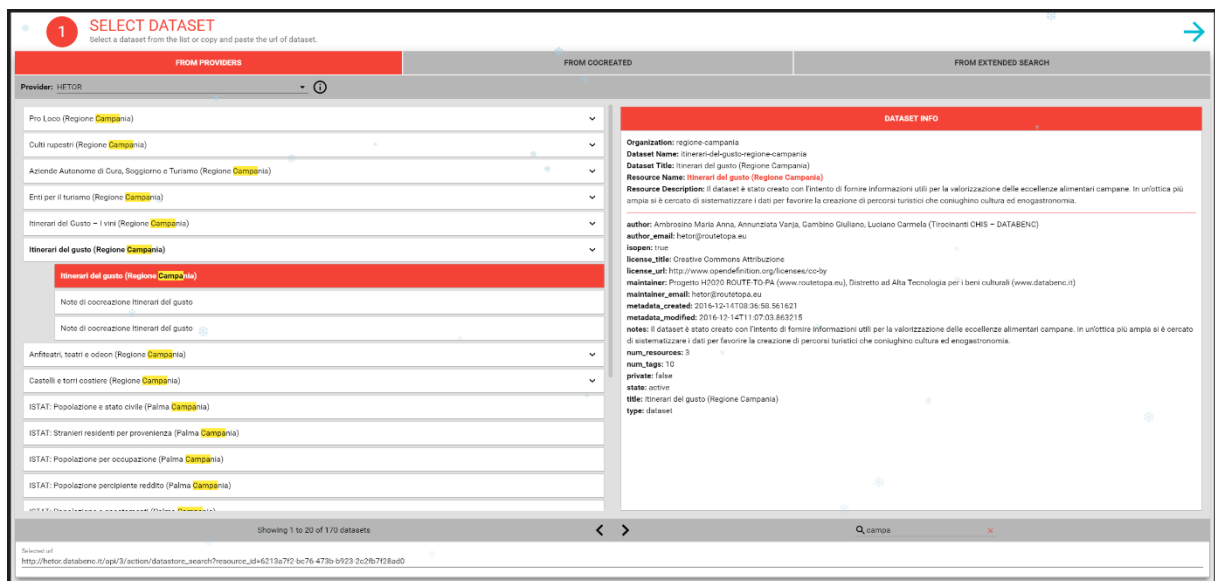
The following table summarizes how the revision of the state on the taxonomy of interactive dynamics, visualization and comparison of existing graphics system (such as Tableau, choose Chart Tamer and instruments, see section 7.1), previously described, led to a new design, an improvement in the interface and adding new features of controllet guided datalet.

	2nd Step		3rd Step		
	Filter, Group By	Type Inferencing, Data Quality	Recommended Charts	Multi Series	Default Values
<b>Taxonomy of interactive Dynamics (7.1.2)</b>	According the taxonomy in the second step is given the possibility for users to select, filter, group and order data thanks to a new panel 'expert mode' that allow to add SQL-like statement.				
<b>Chart Chooser Tool, Chart Tamer (7.1.1)</b>		A new library 'JSDataChecker' has been included for recognise data types and analyse data quality (determines if there are inconsistencies or warnings)	Based on the data types it is possible to reduce the space of proposed visualisations, removing specific charts unsuitable for the selected columns of the		

			dataset. For instance, a map visualisation can be created only if the dataset has numerical latitude and longitude fields.		
<b>Many Eyes, Tableau (7.1.3)</b>				Inspired by other data visualization tools we introduced a new important feature that is the possibility to create chart with multi-series.	
<b>Best Practices with Data Visualizations (7.1.6)</b>					Following the guidelines suggested by the literature we chose carefully our defaults values, colors and options for all charts.

Table 7: A state of art contribution for the design of the datalet wizard controllet (Controllet)

### 7.3.1 STEP 1: SELECT DATASET



Step one is significantly changed due to meet the needs of users and make it simpler, but all the same more functionality and more information to the interface have been added

- Dataset List

To show all the available datasets of the associated providers we choose to replace the simple drop down list with a new component that lists all the datasets as tab grouped twenty at a time, two arrows in the bottom center are available to show the next or previous twenty datasets. If a dataset have more than one resource the tab associated to the dataset become a drop down menu with a tab for each resource. At bottom left is displayed the number of available datasets instead at bottom right is available a search input box.

- Search input box

Now is available a search input box to filter the datasets which may be many to manage. The text that match the searched input text is highlighted with a yellow colour.

- Provider drop down list

Since datasets can be from different providers a drop down list was added to the top left in order to select the provider you want to use. By default all dataset of all associated providers are shown.

- Metadata

In order to give more information about the dataset, in addition to a simple name that can sometimes be meaningless, when a dataset is selected we show, properly embedded, in the div on the right side, all the information, specifically metadata, available for that dataset.

- **Loader**  
When a dataset is selected, you have to wait a variable time, needed to contact the provider and download the whole dataset and the metadata associated with it. The user is often confused and do not realize what the system is doing and if it is working properly. To avoid this reaction, a loader was added in the div dedicated to the metadata, so as to help the user to understand that all is still working fine and that it's a normal waiting time.
- **Cache improvement**  
The available datasets in the *Controllet* can be thousands and associated with many different providers. To make sure you always have the complete list available and to quick search to have a quick search feature, the datasets are cached. Over time the cache mechanism has been improved more and more up to obtain a small size and easy to update cache.
- **Co-created dataset**  
In addition to providers' datasets now is possible to use the datasets created/co-created by SPOD users, clicking on the tab 'FROM COCREATED' the list of all the datasets published on SPOD by users is shown. Also for this datasets information are available, that are the list of creators, the name of the *room* where the dataset has been created and the metadata, if inserted by creators.
- **Tree map**  
The tree map has been removed. The users unanimously evaluate it not useful and difficult to understand, moreover there aren't many systems that use this paradigm and, also for this, it's not suitable for the SPOD's audience.
- **Suggested / most used datasets**  
Differently from what had been planned, there aren't suggested datasets (most used, most important and so on), that's for lack of statistical data.

### 7.3.2 STEP 2: SELECT DATA

**SELECT DATA**  
Select the fields on the left. The table will show the values related to the selected fields.

Pagina Facebook	Provincia	Indirizzo
https://www.facebook.com/epivallano/	Avellino	Via Due Pitegati, 22 A
https://www.facebook.com/turismo/benemerito/	Benemerito	Via N. Sals, 31
https://www.facebook.com/cptcoasta/	Capota	Palazzo Reale
https://www.facebook.com/epi.napoli/	Napoli	Piazza del Martiri, 18
https://www.facebook.com/ETT Salerno 5000824008953/	Salerno	Via Vella, 15

Showing 1 to 6 of 6 rows

**FILTERS** **GROUP BY** **QUERY**

GROUP BY:   
 CALCULATE:   
 COUNT of   
 SUM of   
 MIN of   
 MAX of   
 AVG of   
 FIRST of   
 LAST of

**DISABLE GROUP BY**

**SELECT DATA**  
Select the fields from tree-view. The multi-table will show the values related to the selected fields.

**records**

- BC\_Ref
- Decision
- LONG**
- App\_Date
- LAT**
- Y\_ITM
- Location**
- Application\_Type
- X\_ITM
- Proposal
- \_id

**result / records**

Show 10 entries Search:

LONG	LAT	Location
53.3041801453	-6.2923431396	47, Rathdown Drive, Terenure, Dublin 6w
53.3083152771	-6.2847619057	28, Fergus Road, Terenure, Dublin 6w
53.3089771271	-6.2969498634	55, Parkmore Drive, Terenure, Dublin 6w
53.3104095459	-6.2803769112	4,5,6,7,8 & 9, The Townhouses, Terenure Road East, Terenure, Dublin 6
53.3106536865	-6.2650709152	44, Saint Kevin's Park, Dartry, Dublin 6
53.3106899261	-6.2989220619	3, Lavama Road, Terenure, Dublin 6w
53.3139419556	-6.2732594013	85, Rathgar Road, Rathgar, Dublin 6
53.3143386841	-6.2802495956	71, Brighton Square, Terenure, Dublin 6
53.3151397705	-6.2810525894	44, Brighton Square, Terenure, Dublin 6
53.3160171509	-6.2941257954	282, Kimmage Road Lower, Kimmage, Dublin 6w

Showing 1 to 10 of 100 entries Previous 1 2 3 4 5 ... 10 Next

The step two has been simplified in structure, thanks to an upstream work on the datasets to remove unnecessary and non-informative data (e.g. \_id) and the metadata. In addition was added an 'Expert Mode' for manipulation of the data that allows you to filter and/or group by. In the end it has been included a new library (JSDataChecker<sup>27</sup>) for recognise data types and analyse data quality.

- Tree View Menu - Multi Table / List – Table

The component tree-view on the left used to field selection has been replaced with a simple list, because thanks to an upstream work we can see the dataset like a single table. For the same reason the component multi-table on the right has been replaced with a single table. These changes simplifies the understanding of the dataset to the users, making the data manipulation tool much more clear and simple.

- Expert mode

<sup>27</sup> JSDataChecker is an open source (GPL License) JavaScript library described in the paper "Support citizens in Visualising Open Data" (see Sec. 0)

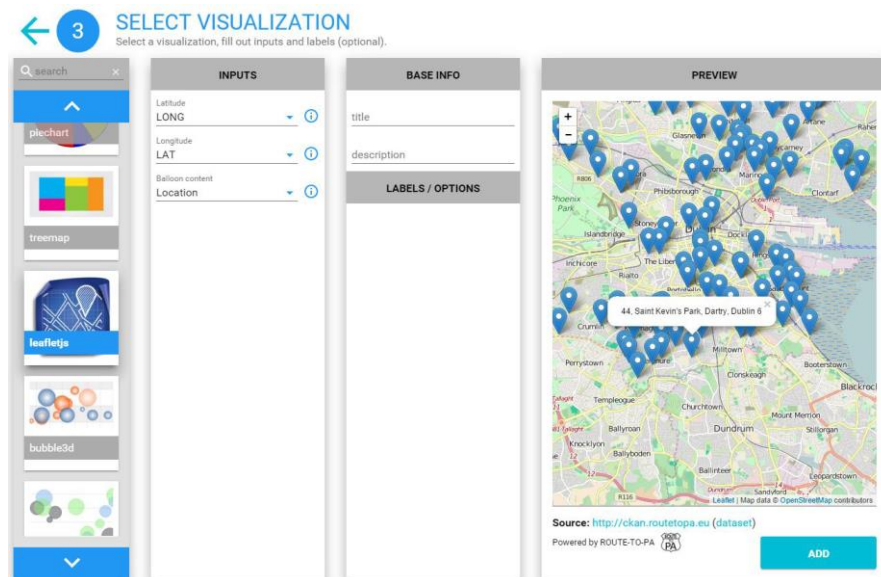
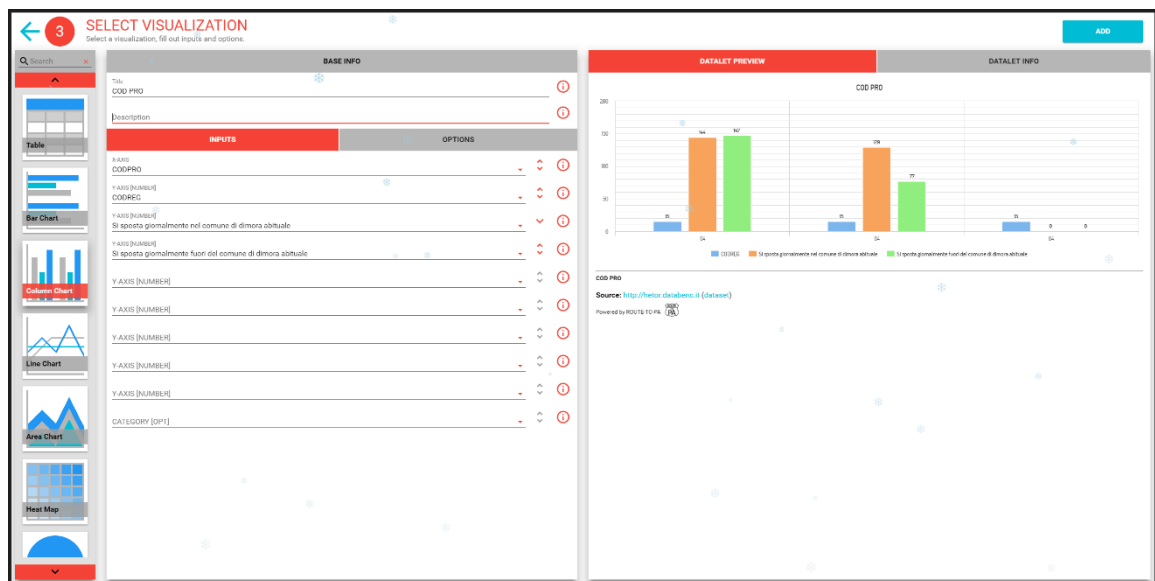


In addition to the selection of the dataset fields (vertical filter / filter columns), it's given the possibility for users to filter rows (horizontal filter) thanks to a new panel that allow to add SQL like conditions (tab 'filters'). Furthermore, in the same panel but in a different tab is possible to use the SQL-like group by statement (tab 'group by'). In the third tab 'query' the current query generated by field selection, filters condition and group by statement is shown. We decided to hide by default these options in a dedicated section "*expert mode*" so as not to make the interface too complicated or confusing for inexperienced users.

- Type inferencing and data quality

From citizens' practical point of view, most of the Open Data are available in csv formats, which of course does not have explicit indication of the data type, essential to choose a suitable visualisation. Some platforms, for instance Linked Open Data (LOD) platforms, store datasets with additional metadata about the data types, however, many deployed institutional OD platforms are not LOD. In order to help to select the right visualisation starting from a tabular dataset, it is essential to know the data type of each stored value. Then, based on the data types it is possible to reduce the space of proposed visualisations (in the third step), removing specific charts unsuitable for the selected columns of the dataset. For instance, a map visualisation can be created only if the dataset has numerical latitude and longitude fields; so, if the dataset does not contain any numerical field in form of latitude and longitude, the system will not show the map visualisation. To know the type of a column and to determine if there are inconsistencies (warnings) just move your mouse over the table header that contains the field in question.

### 7.3.3 STEP 3: SELECT VISUALIZATION



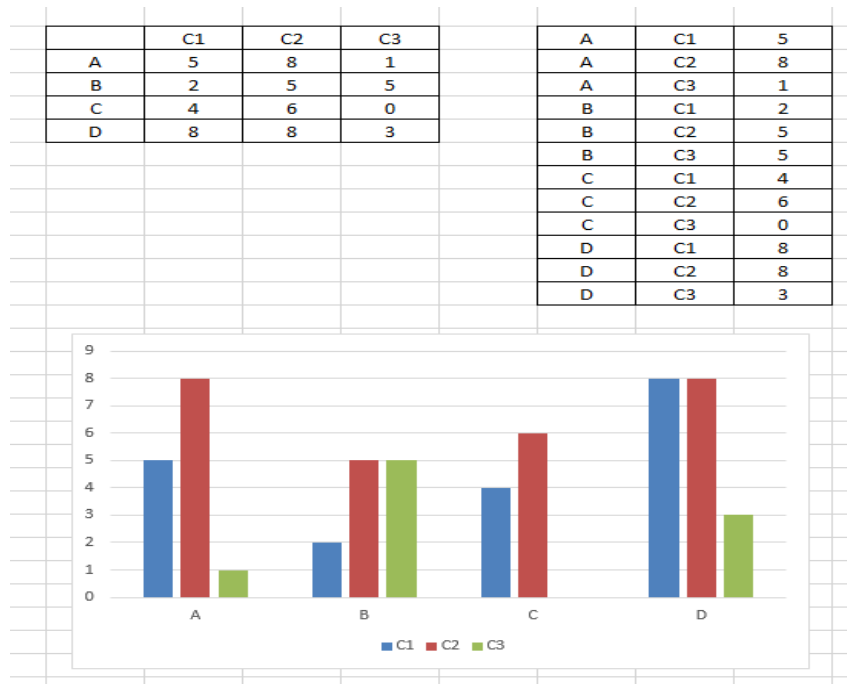
In the same way as the previous two steps, several changes have been introduced. The style has been improved, more information is available, and new features have been added.

- Chat images  
The images of the charts have been updated, new images were created with colors and styles homogeneous and consistent with the theme of SPOD.
- More options  
For each chart more options are available now depending on chart, for example 'legend position', 'theme', 'show/hide data labels' and so on.
- Datalets info  
On the right side the preview div has been replaced by a new panel with two tabs. The first one shows the preview of the datalet as before when it's ready, while the second tab shows all the information about the selected chart that is the image of the chart, a brief description of the chart and when it's recommended, the list of all inputs and options available with related descriptions.
- Order by

Is now possible to sort, in ascending or descending order, alphabetically or numerically depending on the type, the values of charts. The sorting is very easy to obtain, just click on the arrows next to the selected input and the effect will be immediately reported on the previewed chart to the right side (if the chart has been already generated).

- **Multiseries – Stacked**

A new important feature is the possibility to create chart with multi-series and then optionally stacked. There are two ways in which a multi-series can be represented by a table / dataset. There is no right and wrong.



We can manage both situation. In the first case we have to use the first columns as X AXIS and the other three column as three different Y AXIS. In the second case we have to pick again the first column as X AXIS, but the second column as CATEGORY and the third column as Y AXIS.

## 8 CROSS-CUTTING DESIGN CONCERNS

### 8.1 PRIVACY

This section describes how Privacy has been addressed in two different plug-ins, that is, Agora and Data Cocreation. All taken decisions have been included in our Data Policy (i.e., <http://spod.routetopa.eu/Data-policy>).

### 8.1.1 PRIVACY IN THE AGORA

By default, rooms in the Agora are visible only to authenticated users. It is now possible to configure SPOD so that Agora rooms are visible (in read-only mode) by unauthenticated users. The settings is available in the Platform Administration panel (more details in Manual at <http://service.routetopa.eu:8000/d/481c8b5f9c/>).

The option of opening Agora rooms for external access has been envisioned to respond to a specific request of the Prato Pilot. In detail, discussions made inside Agora rooms are visible outside SPOD in a read-only mode. Comments are allowed only after providing credential information. External access means that the URL of a room can be published and made accessible on social networks, blogs, forum, and any type of Web site.

As shown in Figure 51, when privacy settings are disabled, discussions about the expansion of the Wi-Fi network in Prato (for which a specific room has been created in the Agora) are visible just for reading, any access in writing mode, in order to participate the discussion, requires the entering of credential (login form shown in the red rectangle in the bottom right corner of the screen). Moreover, for privacy reasons, we show through a pop-up that the current room is accessible outside SPOD (bottom left).

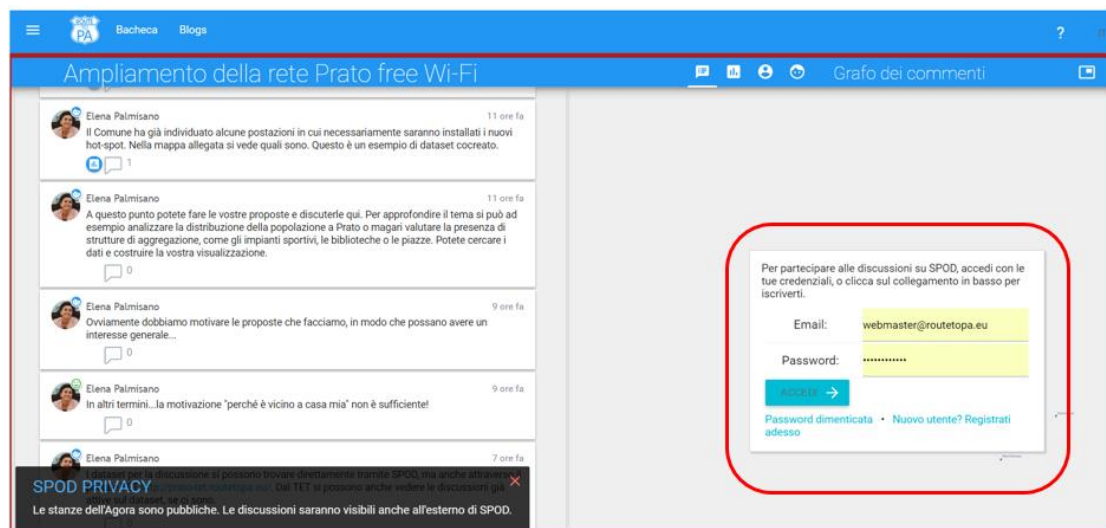


Figure 51: External access to Agora rooms with credentials.

Currently this option is available for the Prato Pilot, only. The option that allows different entities to adopt a local policy has been discussed with Dr. Balachander Krishnamurthy, privacy advisor of Route-To-PA project. Such a resolution preserves local cultural norms without the entire system devolving to the least common multiple (more details in Appendix – A Feedback from the privacy advisor Dr. Balachander krishnamurthy).

### 8.1.2 PRIVACY IN THE DATA COCREATION

As discussed in Section 6.1.2, the Data Co-Creation rooms are virtual places where small groups of participants meet together to collaboratively create new datasets. Citizens become active players, who do not merely access to open data, but contribute to their creation.

An issue related with this functionality is the violation of privacy whether personal and sensitive information are unconsciously added in the data set.

A possible solution is to make the creator of the dataset aware of this possibility, by showing him/her a pop-up (see Figure 52). This pop-up recalls the creator that, according to the current privacy regulation (the Italian privacy legislation n. 196/2003), the data set must not contain any personal or sensitive information. Once the creator accepts, he/she becomes the only responsible for the data entered in the dataset.

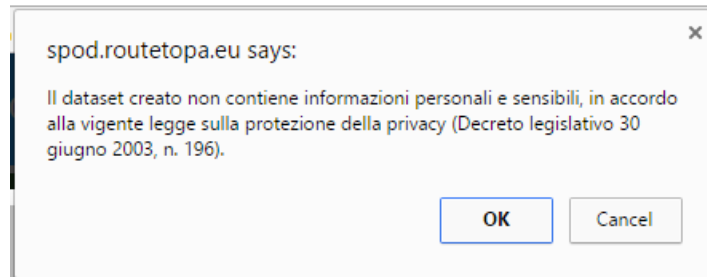


Figure 52: Full control of the owner of the data about their privacy.

## 8.2 RRI-ICT

According to the thematic elements highlighted in the section 7.2 “RRI-ICT” of the Deliverable D.4.1 “Alpha Version of SPOD” in conformity with the EU framework program for R&I, Horizon 2020 [19], the Responsible Research and Innovation (RRI) criteria about the work described in this document addressed are:

- Privacy in the Agora (see sections 6.4.13, 8.1.1)
- Privacy in the Data Cocreation room (see section 8.1.2 )
- Privacy in each page of SPOD platform (see section 6.4.14 )

# 9 PLANNING FOR THE THIRD YEAR OF SPOD

---

In the first year of the project, the design and realization of the alpha version of SPOD (WP4) were achieved considering as input the analysis of “User and system Requirement”s Workpackage (WP2) through the D2.4, the “Models and Methods” Workpackage (WP3) through their deliverable D3.1, the requirements within this document and DoW (Description of Work, the Annex 2 of the Grant Agreement). In the second year of the project, the realization of the Beta version of SPOD (WP4) was achieved considering as input also the analysis of requirements of the Deliverable D.4.1 “Alpha Version of SPOD”, Pilot Feedback, Den Haag Meeting in February and Warsaw meetings in May 2016.

From a preliminary analysis of the requirement, we will be placing particular attention on some issues that were signalled by researchers, users and developers as well, as follows:

- Notification of activities
- Improve discussion room efficiency with a process redesign

- Performance of the platform
- Re design of the co-creation data mechanism to improve efficiency in case of intense use

In addition, we have gathered some general Pilot recommendations for the third year during the last meeting in Salerno (January 10-11 2017) that will be further discussed to possibly be included into the third year activities:

- Improvement of documentation and support:
  - Get rid of language as a barrier, ensure platform is translated in full
  - Create one comprehensive tutorial full of illustrations in all languages
  - Have sectional tutorials available on each section on SPOD
  - Create short videos in your local language and share on social media
- Email notifications by SPOD of relevant events
- Tuning of visibility rules in SPOD (Agora and Co-Creation rooms) – wide privacy/visibility configuration options
- Improving usability
- Tighter integration with existing Social Networks to attract users
- Introduce defense against Spoiler – Troll

A summary table below illustrates candidate improvements based on requirements of all previous releases of SPOD (Alpha and Beta releases):

	Requirement	Notes	DL 4.1	DL 4.2	DoW
	<b>Collaborative space</b>				
1	Enhancement for small-group collaboration	Strategy for persistent layer for synchronous collaboration	(DI 3.1) 6.2. Organizational Level. SPOD, in the initial level (Collaborative Space) facilitates discussions around the meaning of data, and then, in the public room enables participation and collaboration	Sec. 6.1	
	<b>Social capabilities</b>				
2	Attach a document/image	To complete uploading of documents to any comment. At the moment it is available in discussion space of Cocreation room.	(DL2.4) UC4(A): Attach a File as Support. UC5: Share a link s1.2: Insert file on SPOD  (DL3.1) 6.1 Three democratic traditions, Implication for TET/SPOD: SPOD facilitates also collaboration e.g., being able to share documents, share visualizations and work together in these documents or visualizations.	In discussion space of Cocreation data room, an user can upload documents (supported type files are pdf, image doc ) . (Sec. 6.4.1)	“Usability” (sec. 2.2.1.3)

3	Notifications of activities	Notify all the activities to everybody or only the friends	(DL 3.1): 6.1 Three democratic traditions, Implication for TET/SPOD: SPOD facilitates participation. It facilitates open communication, a dialogue, a two-way flow of information in which citizens can make their voices heard (DoW): Section 2.1.3.1 "Social Platform for Open Data": supporting collaboration and social interaction	E-mail notification strategy (Sec. 6.4.12)	
	<b>Public Discussion (Agora)</b>				
5	Enhancement for discussion room	A new re-design for discussion rooms in Agora		Pilots Feedback	
	<b>Dataset visualisation (Controllet)</b>				
6	Clustering options for maps of large quantities of data.	A study for visualization in map of large quantities of data introducing kind of optimization like grouping of data etc. Reducing the amount of points on the map can be accomplished by grouping points by certain criteria and displaying with different colors and sizes, icons, placement rules, and more		Pilots Feedback	
	<b>Support to user</b>				
7	Manual easily accessible	Manual is easily accessible within SPOD and Tutorial in all languages	(DL 2.4) : s5.1: Learn TET/SPOD		
8	Video instructions on YouTube easily accessible	Videos is easily accessible within SPOD in local language	(DL 2.4) : s5.1: Learn TET/SPOD		
9	Nationalisation in each language (EN, FR, IT, NL)	Platform available in each language	(DL 2.4) : s5.1: Learn TET/SPOD		
10	Context-sensitive help	Tooltip or brief explanation to define page or widget of UI (e.g., button etc...), precisely how to interact with the control/page in question	(DL 2.4) : s5.1: Learn TET/SPOD		

Table 8: A summary of the planning of the third year of SPOD

## RELEVANT PUBLICATIONS

---

The activities that led to this deliverable have been partially reported in the following publications:

- Gennaro Cordasco, Renato De Donato, Delfina Malandrino, Giuseppina Palmieri, Andrea Petta, Donato Pirozzi, Vittorio Scarano, Luigi Serra, Carmine Spagnuolo and Luca Viciomini,  
*“DataEt-Ecosystem Provider (DEEP): Scalable Architecture for Reusable, Portable and User-friendly Visualizations of Open Data”* in International Conference for E-Democracy and Open Government 2017 (CeDEM17), 17- 19 May 2017 (Submitted)
- Gennaro Cordasco, Renato De Donato, Delfina Malandrino, P. Palmieri, Andrea Petta, Donato Pirozzi, Vittorio Scarano, Luigi Serra, Carmine Spagnuolo and Luca Viciomini,  
*“A Scalable Data Web Visualization Architecture”* in Conference “25th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP 2017)”, - March 6th-8th, 2017, St. Petersburg, Russia, 2017
- Jerry Andriessen, Michael Baker, Gennaro Cordasco, Renato De Donato, Delfina Malandrino, Giuseppina Palmieri, Mirjam Pardijs, Andrea Petta, Donato Pirozzi, Vittorio Scarano, Luigi Serra, Carmine Spagnuolo and Luca Viciomini  
*“Increasing Public Value through Co-Creation of Open Knowledge”*  
In International Conference on eDemocracy & eGovernment (ICEDEG 2017), 19-21 April, 2017, Quito-Ecuador
- Paolo Boscolo, Gennaro Cordasco, Renato De Donato, Delfina Malandrino, Giuseppina Palmieri, Elena Palmisano, Andrea Petta, Donato Pirozzi, Vittorio Scarano, Luigi Serra, Carmine Spagnuolo, Luca Viciomini *“Engagement of citizens through the co-creation of Open Data: the Prato Wi-Fi coverage extension Use Case”*  
in Conference GARR 2016, November 30 – 2 December 2016, Firenze Polo Universitario di Novoli, Italy.
- Gennaro Cordasco, Renato De Donato, Delfina Malandrino, Giuseppina Palmieri, Andrea Petta, Donato Pirozzi, Vittorio Scarano, Luigi Serra, Carmine Spagnuolo, Luca Viciomini  
*“Data-Driven Discussions: A Social Platform for Open Data to discuss and visualize from heterogeneous data sources”*  
in Conference Data for Policy 2016, “Frontier of Data Science for Government: Ideas, Practices and Projections” September 15 – 16 2016, University of Cambridge.
- Donato Pirozzi, Vittorio Scarano  
*“Support citizens in Visualising Open Data”*  
In Information Visualisation (IV), 2016 20th International Conference, July 19-22, Lisbon, Portugal
- G. Cordasco, R. De Donato, D. Malandrino, G. Palmieri, A. Petta, D. Pirozzi, V. Scarano, L. Serra, C. Spagnuolo and L. Viciomini  
*“How to Engage Citizens in Discussing Around Open Data”*  
In 2nd CINI Annual Conference on ICT for Smart Cities & Communities, September 29-30, 2016 Benevento, Italy.



## 10 CONCLUSIONS

---

The purpose of SPOD Beta release was to meet the specific needs of our Pilots and in fact, we have been mostly taking into account feedback received from Pilot during your activities of the first and second year of the ROUTE-TO-PA project and requirements received from project meetings to plan additional enhancements for SPOD platform.

At the end of the second year of the ROUTE-TO-PA project, a SPOD platform (Beta version ) for each Pilot is ready introducing mainly new tools (such as Cocreation knowledge and data room, Events, Blogs ) and a new re-designed wizard (i.e., the Controllet) to provide an easy mode to visualize data with charts, maps.

The difficulties and barriers towards already existing datasets of Open Data portal (e.g., missing metadata and data, name of fields not significant), met by our Pilots, can be addressed through the process of the Co-creation, that is groups of citizens have the opportunity privately to create a room and meet in small groups to co-create a dataset (Co-creation data room). At some point, they can decide to publish the co-created dataset within the platform, enlarging the audience, allowing other users to have look to and exploit them. In addition, the citizens can co-create information or knowledge using the Co-Creation knowledge room. They can meet in small group to analyse, interpret, and transform data producing collaboratively a document to stimulate discussions or become part of a post blog within the SPOD platform.

The work will continue in the third year as anticipated roughly in Section 9, and the entire team of the project is ready to gather additional feedback from the end-users and the pilots.

## 11 BIBLIOGRAPHY

---

- [1] A. H. M. T. V. Hudson Borges, «Predicting the Popularity of GitHub Repositories,» *Proceedings of the The 12th International Conference on Predictive Models and Data Analytics in Software Engineering*, n. ISBN: 978-1-4503-4772-3, ACM New York, NY, USA ©2016.
- [2] M. H. Moore, *Creating public value: Strategic management in government.*, Harvard university press, 1995., 1995.
- [3] R. L. Ackoff, From data to wisdom, *Journal of applied systems analysis*, vol. 16, no. 1, pp. 3–9, 1989., 1989..
- [4] O. K. International, The Open Definition, available online at <http://opendefinition.org/>.
- [5] J. L. D. a. P. Vandenbroeck, «Choosing the right graph,» *IEEE Transactions on Professional Communication* , Vol. %1 di %2 (Volume:45 , Issue: 1 ), pp. 1-6, Mar 2002.
- [6] J. a. S. B. Heer, «Interactive Dynamics for Visual Analysis,» *ACM's annual Applicative conference is June 1-2 in New York City.* , February 20, 2012.
- [7] K. a. B. M. a. G. D. a. K. R. a. M. J. Morton, «Public Data and Visualizations: How Are Many Eyes and Tableau Public Used for Collaborative Analytics?,» *SIGMOD Rec*, 2014.
- [8] F. B. V. a. M. W. a. F. v. H. a. J. K. a. M. McKeon, «ManyEyes: a Site for Visualization at Internet Scale,» *IEEE Transactions on Visualization and Computer Graphics* , Vol. %1 di %2 (Volume:13 , Issue: 6 ), pp. 1121 - 1128, 2007.
- [9] P. a. X. F. a. S. M. a. K. V. a. W. R. Terlecki, «On Improving User Response Times in Tableau,» *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 2015.
- [10] J. D. M. B. S. Stuart K. Card, «Readings in Information Visualization: Using Vision To Think,» *Academic Press*, 1999.
- [11] H. a. H. A. a. J. C. S. a. L. A. a. M. J. a. S. R. a. S. W. Gonzalez, «Google Fusion Tables: Data Management, Integration and Collaboration in the Cloud,» *Proceedings of the 1st ACM Symposium on Cloud Computing*, 2010.
- [12] R. B. J. M. R. M. a. C. S. Kristi Morton, «Dynamic Workload Driven Data Integration in Tableau,» *SIGMOD '12* , May 20–24, 2012, Scottsdale, Arizona, USA. Copyright 2012 ACM 978-1-4503-1247-9/12/05.
- [13] S. Few, *Show me the numbers*, Analytic Press, 2004.
- [14] D. H. R. P. \. L. W. Maila Hardin, *Which chart or graph is right for you?*, Copyright Tableau Software, January 2012.
- [15] H. a. Visage, *Data visualization 101: How to design charts and graphs*.
- [16] J. L. J. R. Ross Crooks, *The Power of Visual Storytelling*, Wiley 2012, 2012.
- [17] D. M. Wong, *The Wall Street Journal Guide to Information Graphics*, 2010.
- [18] S. Few, *Selecting the Best Means and Manner of Display Effectively Communicating Numbers*, Principal, Perceptual Edge, November 2005..
- [19] M. W. Bauer, *Indicators for promoting and monitoring Responsible Research and Innovation*, Report from the Expert Group on Policy Indicators for "Responsible Research and Innovation", 2015.
- [20] P. Vajjhala, «Understanding Data: A Comparison of Information Visualization Tools and Techniques».

## 12 APPENDIX - A: FEEDBACK FROM THE PRIVACY ADVISOR DR. BALACHANDER KRISHNAMURTHY

---

*What follows is the transcript of the feedback sent by the Privacy Advisor after the visits to the University of Salerno and the meetings that were held with the entire team about the privacy issues of SPOD.*

During the fourth week of November 2016, I visited University of Salerno to discuss the updates on the ROUTE-TO-PA project. I had meetings with Professor Vittorio Scarano (coordinator) and Doctor Delfina Malandrino. I went over the Administrator Guide, the data policies for developing SPOD, and the testing manual (version 1.8 of Mira). I also caught up on the questions raised by Prato, Dublin, and Paris groups about various issues and presented my views on them. I went over the multiple videos including the various SPOD tutorials on map co-creation, datalets, Agora etc. The visualization of the discussion in different rooms with indications of number of participants, depth of discussions etc. are all interesting UI features.

Below I'll cover some of the issues discussed and I'll present my justifications. None of the issues have binary answers and given the subjective nature of privacy, the participating entities may well have different views.

### 12.1 COLLECTING PERSONAL AND SENSITIVE INFORMATION

First, the views about collecting or not collecting sensitive information may be colored due to differing local cultural expectations. In such matters the traditional way to make progress is to provide an option that would allow different entities to adopt a local policy. Such a resolution preserves local cultural norms without the entire system devolving to the least common multiple. For example, one city may want strong protection on who can see messages without registration, while another more 'liberal' city may allow open access. By localizing decisions both cities can retain their points of view without one policy overriding the other.

### 12.2 AWARENESS ABOUT DISCUSSIONS OF SENSITIVE NATURE

Next, the clear enunciation of the data policy, ethical guidelines, supporting videos, are all a sign of progress in preserving privacy. Rather than attempt to add privacy later, the project is integrating privacy considerations from the ground up. The visual presentations provide a strong contextual feedback. Having a constant visual feedback allows for users to be always aware. As an example, for decades, at the National Security Agency in the US, when visitors without top secret credentials were on the floor, a light would continually blink above the desks of all employees on that floor so they could be aware of their presence. Such a visual feedback prevents

users from forgetting that their discussions could be overheard by others. The principal embodied here is the awareness of the difference between on-line and off-line privacy. Although many users are quite savvy when it comes to privacy in their off-line (i.e., real world interactions), they are a lot less savvy in the online world. By using cues they are familiar with in the off-line world, we may be able to bring about the same level of awareness in online interactions. Thus, in the Agora, when discussions of sensitive nature may occur, users could be given continual visual cues to raise their awareness.

### 12.3 AUTHENTICATION

The issue of fake accounts was discussed. There are extreme ways to resolve authentication: we could disallow posting by all, which would render SPOD useless. We could insist on very strong authentication whereby users have to present highly sensitive personally identifiable information before they could post anything. Gathering of such information may be perceived as a violation of privacy guidelines. So we need to come up with a via media. The underlying concern is that of gaming. It is well known that astro-turfing is widely used in community interactions to game discussions. Astro-turfing implies an attempt to create support for one particular view via fake accounts.

Well known sites such as tripadvisor.com have had to deal with users maliciously posting false information about hotels, restaurants etc. Often the resolution has been to insist that users create an account in the hope that the effort required to create one would reduce the amount of fake postings. However, it is not hard to create a large number of accounts with different easily generated email addresses.

A way to mitigate the problem is to increase the amount of work that a user has to do the first time they create an account or post a message. Since most users are likely to post only a few messages they would be able to handle the added work. However, an astroturfer intent on creating hundreds of fake accounts would have to do significantly higher amount of work.

The work can be in the form of complicated CAPTCHA puzzles, identifying rivers present in only a subset of a collection of images correctly etc. A single mistake would result in a new puzzle slowing down malicious users. Yet another alternative is to ask for some unique information about a user such as a cellphone number. Almost no malicious user is likely to obtain multiple valid cellphone numbers to thwart the scheme. A single one-time text message sent to the cellphone would need to be entered to gain access. A cellphone number could not be reused. While this would solve the problem completely, there is the new concern of the potentially private information of user (cellphone number) being gathered. Here, we need to decide if there is enough trust in the administrator of SPOD to handle such information. For example, the data could be kept scrambled (one-way hash) for comparisons alone.

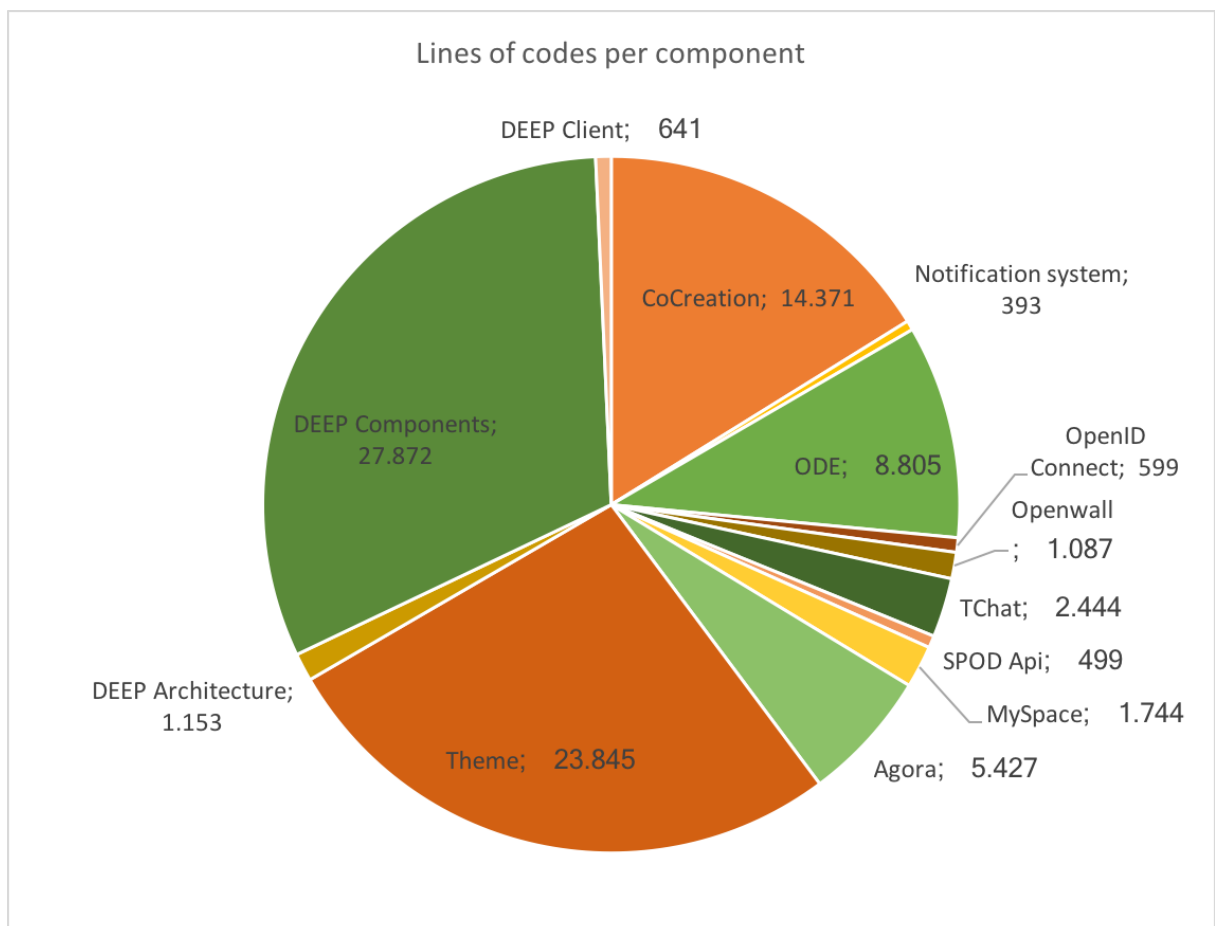
Balachander Krishnamurthy

## 13 APPENDIX – B: METRICS

### 13.1 SPOD BETA METRICS

For each component (SPOD plugin or DEEP component) the number of lines of code of each kind (HTML, PHP, XML, JavaScript, etc.) are as follows:

Module	Apache config	Bourne Again Shell	CSS	HTML	JavaScript	JSON	Markdown	PHP	Smarty Config	SQL Files	Text	XML	Total
CoCreation	3	433	2.616	4.618	3.284	328	32	2.850	43	5	145	14	14.371
Notification system	-	112	-	48	-	-	20	173	21	-	5	14	393
ODE	6	-	-	5.034	535	-	13	3.173	-	-	30	14	8.805
OpenID Connect	-	-	37	36	-	-	-	507	-	-	5	14	599
Openwall	-	-	-	163	-	-	-	791	-	-	110	23	1.087
TChat	-	-	-	449	649	-	11	1.316	-	-	5	14	2.444
SPOD Api	-	-	-	-	-	-	-	480	-	-	5	14	499
MySpace	-	-	48	1.038	142	-	21	476	-	-	5	14	1.744
Agora	-	-	240	2.467	802	-	34	1.864	-	-	6	14	5.427
Theme	-	-	11.954	511	-	-	-	-	-	-	11.364	16	23.845
DEEP Architecture	6	-	-	-	-	-	33	256	-	-	8	850	1.153
DEEP Components	-	-	1.269	21.721	4.469	130	-	-	-	-	201	82	27.872
DEEP Client	-	-	-	497	74	-	70	-	-	-	-	-	641
SPOD BETA	15	545	16.164	36.582	9.955	458	234	11.886	64	5	11.889	1.083	88.880



## 14 APPENDIX – C: NON-FUNCTIONAL REQUIREMENTS FROM D2.1

---

The process of non-functional requirements derived from the state of art of the D2.1 “Review of the State of the Art and Evaluation of Existing Open data Platforms” is dealt in the section 2 “Methodology” of D2.1 by describing a conceptual model of transparency presented in Cappelli 2009 <sup>28</sup> highlighting a catalogue of Non-Functional requirement (NFR) (e.g., quality attributes as usability, accessibility etc.).

---

<sup>28</sup> Cappelli C et al, Managing Transparency Guided by a Maturity Model, 3rd Global Conference on Transparency Research HEC PARIS, October 24th – 26th, 201