



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



Project number 645860
H2020-INSO-2014

D4.5 Beta version of TET

(Final, version 1.0, 31/01/2017)



WISE&MUNRO



Document produced by

Organisation: Insight, NUIG

Author / email:

Adegboyega Ojo, adegboyega.ojo@insight-centre.org

Arkadiusz Stasiewicz, arkadiusz.stasiewicz@insight-centre.org

Edobor Osagie, edo.osagie@insight-centre.org

Lukasz Porwol, lukasz.porwol@insight-centre.org

Mohamamd Waqar, mohammad.waqar@insight-centre.org

Samuel Adebayo, samuel.adebayo@insight-centre.org

Subject: Beta version of TET

Due date: 31 January 2017

Dissemination level: Public (PU)

Reviewed and approved by

Date	Name	Organization
21-01-2017	Albert Meijer	UU
25-01-2017	Vittorio Scarano	UNISA

Revision History

Version	Date	Authors	Status	Description of Changes
0.1	2016/12/07	A. Stasiewicz	Draft	Concept
0.2	2016/12/19	M. Waqar	Draft	Section 5 draft
0.3	2017/01/06	M. Waqar	Draft	Section 6 & 7 draft
0.4	2017/01/09	A. Stasiewicz	Draft	Section 2
0.5	2017/01/09	A. Stasiewicz	Draft	Update all sections
0.6	2017/01/17	E. Osagie	Draft	Section 3 & 7 draft
0.7	2017/01/18	E. Osagie	Draft	Addition section 7 and related appendices
0.8	2017/01/20	M. Waqar	Draft	Added more content
0.8	2017/01/23	E. Osagie	draft	Correction of errors/omissions
0.8	2017/01/23	E. Osagie	draft	Completion, proof-reading
0.9	2017/01/30	A. Stasiewicz	Draft	General updaets
0.91	2017/01/30	A. Stasiewicz	Draft	Completion, proof-reading
1.0	2017/01/31	L. Porwol	Final	Final Version

TABLE OF CONTENTS

List of Figures.....	4
List of tables.....	5
List of abbreviations	6
1 Executive Summary	7
2 Introduction.....	9
2.1 The development process	9
2.2 From Alpha to Beta: a roadmap of the major changes	10
2.3 The Development Process: Design Implementation and Release	11
2.4 Release notes	11
2.5 Consolidated Expert Review Report - Addressing issues.....	13
3 Requirements	14
3.1 Functional Requirements	14
3.2 New and Update to Functional or System Requirements	16
3.3 System Requirements.....	23
4 TET Architecture and Design	27
4.1 Architecture principles	27
4.2 TET Reference Architecture.....	27
4.3 Architectural Layers.....	27
4.4 Implementation Architecture	28
5 Implementation.....	30
5.1 Approach.....	30
5.2 Implementation Technology.....	31
5.3 Implementation details	32
5.4 Implemented features.....	32
5.5 Integration between TET and SPOD	40
6 Deployment	42
6.1 Deployment requirements	42
6.2 Deployment process.....	43
6.3 Deployment issues	44
7 Testing and Evaluation	45
7.1 Verification and Validation Concepts	45
7.2 Feedback from Testing and Evaluation Processes	46
8 Cross-cutting Design Concerns	54
8.1 Privacy	54
8.2 RESPONSIBLE RESEARCH AND INNOVATION COMPLIANCE	54
9 Future Development Plans	56
10 Conclusion	58
11 Appendices.....	59
Appendix 1A: Pre-usability evaluation testing – 2 Samples of manual with user comments	59
Appendix 1B: Detailed Results from Pre-Usability Evaluation Testing Mar, 2016	86
Appendix 2: Usability Evaluation – Instruction Manual	92
Appendix 3: Usability Evaluation – Google Survey Report.....	97
Appendix 4: Github Issues List	105
Bibliography.....	106

LIST OF FIGURES

Figure 1: Relationship between deliverables and tasks	7
Figure 2: TET development cycle	9
Figure 3: Scrum Process.....	11
Figure 4: ROUTE-TO-PA Requirements upgrade from Evaluation Feedback.....	17
Figure 5 TET Reference Architecture	27
Figure 6 TET implemented architecture	29
Figure 7: Requirements gathering	31
Figure 8: User interface - Homepage of TET	33
Figure 9: Search experience improved with automatic keyword suggestion	33
Figure 10: Screenshots showing search result filtering tools	34
Figure 11: Screenshot of a selected dataset showing various boxes of information on the dataset	35
Figure 12: Tabular view of a selected dataset	36
Figure 13: Summary View screen offers further information on a dataset	37
Figure 14: Autogenerated chart on a dataset.....	37
Figure 15: Dataset quality metrics from the metadata	38
Figure 16: Metadata information in natural language.....	38
Figure 17: A screenshot from SQL query on a dataset	39
Figure 18: System recommendation of related dataset with a slider to adjust relatedness of items on list.....	39
Figure 19: Screenshot showing compatible datasets that can be merged and analysed together & incompatible dataset that cannot be merged	40
Figure 20: Screenshot showing user discussions on a dataset imported from SPOD.....	41
Figure 21: TET deployment process.....	44
Figure 22: <i>Validation and Verification Processes in Software Development (Baresi, L & Pezze, M. 2006)</i>	45
Figure 23: <i>A) V&V Scope, B) V&V Interconnections. (Easterbrook, S. (2010))</i>	46
Figure 24: Evaluators' rating across functional areas of ROUTE-TO-PA system	50

LIST OF TABLES

Table 1: TET releases notes.....	12
Table 2: Recommendations concerning future work.....	13
Table 3: Overview of the Societal Activity model of Open Data use	14
Table 4: Complete System (Solution) Requirements - TET (implemented components only)	15
Table 5: Earlier System Requirements not in Alpha but in Beta Version of TET.....	16
Table 6: Pre-Evaluation Testing Feedbacks and Implementations thereof	18
Table 7: Review Results from Other Testing and Pilot Sites.....	18
Table 8: Usability Evaluation Comments and Implementations	21
Table 9: Requirements for CKAN with TET extensions.....	42
Table 10: Hardware Requirements	42
Table 11: Links to the repositories.....	43
Table 12: Development Processes in Categories	45
Table 13: ROUTE-TO-PA Tech Evaluation Survey – System Features and Usability statements	48
Table 14: QUIM usability factors and criteria (excerpt)	49
Table 15: Computation of average score from usability rating.....	49
Table 16 Candidate Feature for TET final release	56

LIST OF ABBREVIATIONS

The following table presents the acronyms used in the deliverable in alphabetical order.

Term	Definition
DEEP	DataEt-Ecosystem Provider
DoW	Description of Work (i.e. Annex 2 of the Grant Agreement)”
ICT	Information Communication Technology
OD	Open Data
PA	Public Administration
SPOD	Social Platform of Open Data
TET	Transparency-Enhancing Tools
UC	Use Case
WP	Work Package

1 EXECUTIVE SUMMARY

ROUTE-TO-PA, which is derived from Raising Open and User-friendly Transparency-Enabling Technologies for Public Administration, is an innovation project focused on improving Transparency through prototyping and piloting the integration of Open Data platforms with social networking technologies. The combined system is enhanced with tools to facilitate a better understanding of Open Data, better information discovery, improved linking of metadata and enhanced personalization of data usage. The two main objectives of the ROUTE-TO-PA project can be sum up as:

1. To enable the transition to the next generation of Open Data portals by creating tools that will enable citizens to engage themselves socially over Open Data resources. This is termed *Social Platform for Open Data (SPOD)*.
2. To provide tools that could be integrated into existing Open Data platforms to deliver greater data transparency, quality and understandability. This component is termed the *Transparency Enhancing Toolset (TET)*, and it provides basic analytical tools for reducing datasets into a more understandable form for users.

The report entitled “D4.5 TET beta” presents the results of development activities performed as part of Task 4.2 between months M12-M24 to release the beta version of Transparency Enhancing Toolkit (TET). TET beta is a continuation of work done as part of Task 4.2 between months M6-M12 and documented in Deliverable D4.2 TET Alpha. The background information and documentation related to the alpha version of TET is available in Deliverable 4.2 Alpha version of TET and can be accessed through the project website: <http://routetopa.eu>. Therefore, to avoid redundancy and for better clarity in this document we do not include the content already put in D4.2. Deliverable D2.4 requirements specification and Deliverable D2.3 user case models User Stories on and User Stories produced in Task 2.3 serve as input to this task (see Figure 1).

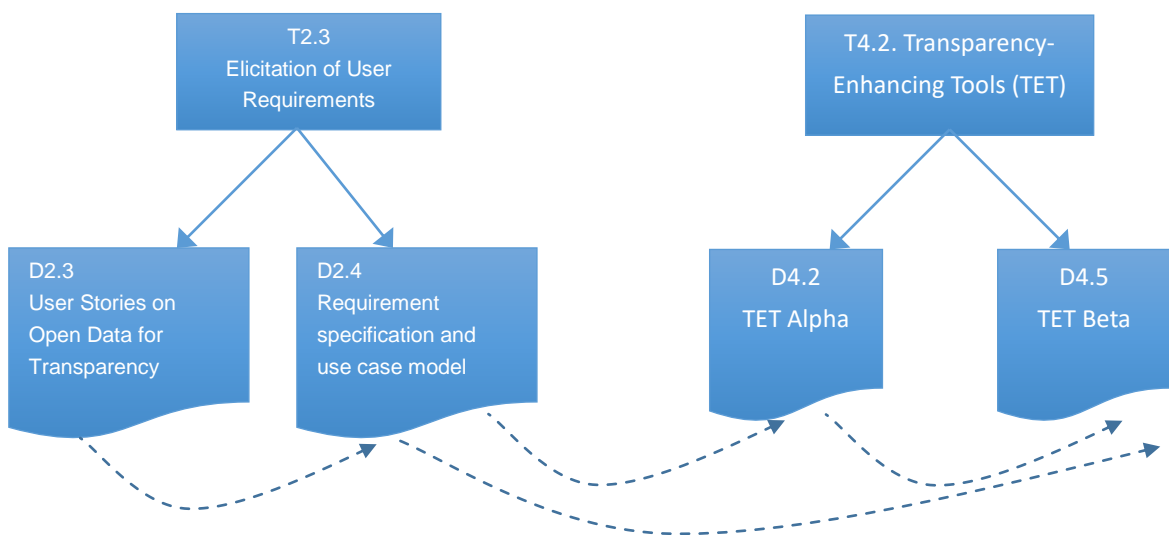


Figure 1: Relationship between deliverables and tasks

This document presents current state of the software development after Year 2 and the future development plans. It also presents the results of the various platform tools testing tasks – unit tests, system test and usability evaluation tasks. The feedback from the testing tasks was used as a base for updating functional requirements

and supplied as part of the requirements to upgrade the platform from alpha level to beta level. Summary of results from evaluations mentioned indicate significant improvements achieved by TET beta in respect to TET alpha (according to users who evaluated the usability of the tools). In particular, the TET subsystem now has a simpler, clearer user interface which enables the use of keyword search and advanced filtering and sorting tools that improve search result accuracy. Better presentation and clarity increased user satisfaction. Additionally, improved data analytics and visualisation, data artefact sharing tools, table views and other options to view related discussions on datasets are imported from the SPOD interface as a result of integration efforts between TET and SPOD. Nevertheless, users were still demanding more learnability and informative qualities on the platform, such as tooltips, help documentation and demo videos which were insufficient or absent at the time of system evaluation, affecting usability quality.

The focus of the TET beta release is to improve on the shortcomings of TET alpha release by addressing the challenges identified during the evaluation process and to introduce new features that will make open data more accessible to a wider audience. Usability was identified as a key barrier to adoption of TET alpha, which was mainly caused due to its inherent dependence on CKAN. Usability of TET beta has been significantly improved by redesigning the entire user experience, as a result the new user interface is neat, elegant and easy to use for exploration of data. Users can search, explore, integrate, analyse and visualize data in a few intuitive steps without requiring deep technical knowhow. As TET and SPOD are set to be two complementary components for enabling transparency (TET covers the human data interaction part whereas SPOD handles the social interaction over data), both TET and SPOD beta implementations have focused on streamlining the interaction between the two components by having joint authentication system and share the capabilities to deliver value to the end users, explained in detail in section 5.5. For technical users, TET beta offers Querying, Data Integration and API that can be used for getting more deep insights from data. Section 8 provides details related to our vision for future developments activities.

The target audience for this document includes platform designers, developers as well as various platform stakeholders, including the external testers (non-consortium members) who will participate in the development and testing phase of the project.

2 INTRODUCTION

The purpose of the ROUTE-TO-PA platform is to develop ICT tools for social and open data that will enable citizens to engage over open data and to understand open data for achieving better transparency. There are two main tools being developed as part of ROUTE-TO-PA, SPOD that is the social platform for interaction over open data and TET that helps open data users to discover and understanding open data.

2.1 THE DEVELOPMENT PROCESS

In order to manage the product development process for TET, we adopted SCRUM a popular agile framework that allows development of usable software with incremental changes based on the feedback. We collect feedback from pilots and the evaluation workshops to plan tasks and organize the development activities. Development tasks are done in a series of sprints, each sprint ends with release of useable software. Planning meetings are organized at the start of each sprint to clearly define the objectives of each sprint and to plan development tasks. To track the project progress daily short stand-up meetings are held. A retrospective session is conducted at the end of each sprint to learn from the sprint experience and to use the results for improving future work. The software components are extensively tested before the release. Feedback from the pilots and evaluation workshops are used to prioritize new feature development, issues fixing and enhancement of existing features. The TET development cycle is presented in Figure 2 TET development cycle and consists of four stages.

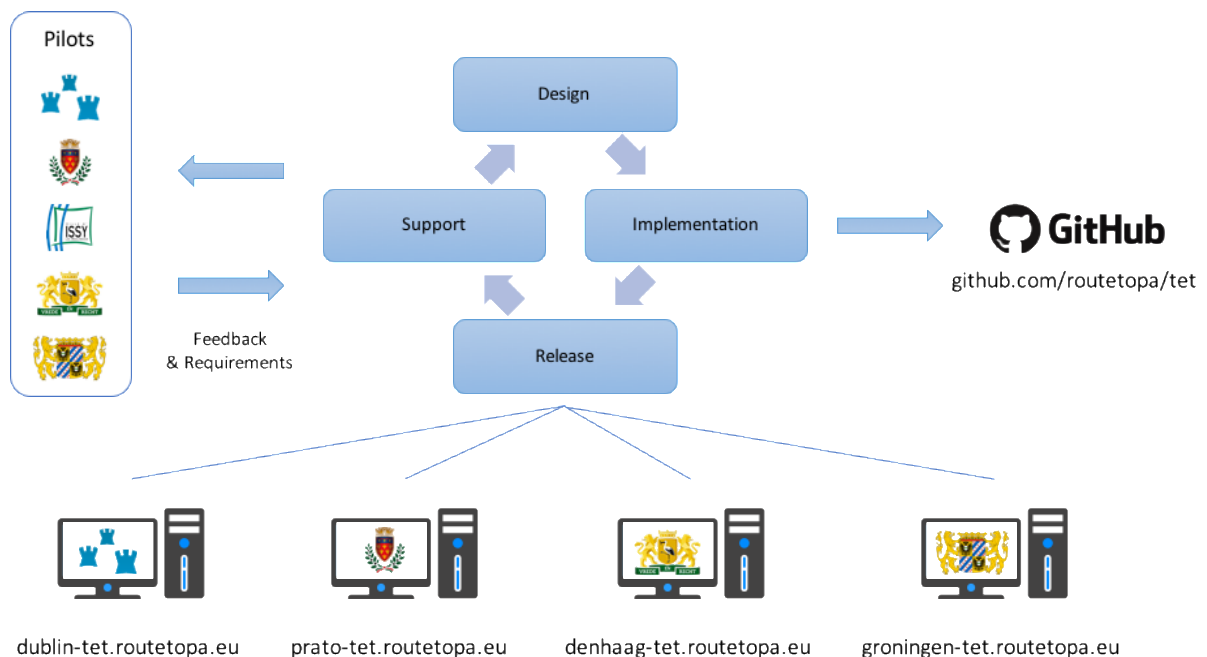


Figure 2: TET development cycle

Feedback gathered from pilots is used as a basis for new features design, bug fixes and enhancement to the platform at the design stage. Software implementation and internal testing take place at the implementation stage where developers are assigned to implement new features enhance existing ones and fix bugs reported by the pilots or discovered during the evaluation workshops. After the implementation phase is complete, the software is released and deployed on four pilot sites for evaluation. Pilots are supported throughout testing and the evaluation phase after each release; once the testing and evaluation is complete feedback is collected from pilots and provided to the development team. This feedback is used by the development team as input to the next iteration of development activities. This process ensures that all pilots get usable software at the end of each development iteration. Improvements and bug fixing are done as software evolves and the overall risk for project is reduced. Moreover, this approach guarantees, that the produced software is aligned with the requirements of all pilot partners at the end of every iteration cycle.

2.2 FROM ALPHA TO BETA: A ROADMAP OF THE MAJOR CHANGES

Several usability issues were identified during the evaluation process of TET alpha release, which were mostly related to the inherent complexity of the CKAN platform. To rectify this, a new user interface for CKAN was designed from scratch with strong focus on user experience and ease of use. The TET user interface was revamped to meet the requirements of the ordinary citizen. Search and filtering options were improved by taking inspirations from popular search engine and ecommerce sites. A search feature was highlighted and made accessible throughout the user interface, which helps users flexibly explore new datasets. Autosuggestions and personalized search results are provided wherever possible. A range of search facets like location, data, file formats etc. with counts are provided for filtering the results space, in addition new options for ordering are added to assist users in finding only relevant results. Users can directly view graphs, summary and the resources attached with the datasets from the search results page with a single click.

The Datasets page is designed and crafted according to needs of common citizens. Metadata is displayed in the form of readable text. Graphs are automatically generated from resources associated with datasets and are presented to the user on the dataset page. Additional panels are added to enhance the users understanding of the dataset. The dataset quality panel provides information about key quality indicators related to the dataset and associated resources. The files panel lists all the files resources attached to the dataset. The recommendation feature automatically suggests a list of datasets for further exploration. The SPOD discussion panel provides discussion related to the dataset from SPOD. The share panel can be used to share datasets on popular social media outlets.

The resource view provides multiple views over the tabular data as also display the associated metadata, discussion and sharing options. A link to API is provided with copy options so that they can be copied and used to create meaningful visualizations in SPOD. Data Explorer and PivotTable view are imported from a previous version. The Data Explorer shows data in the form of a data grid with options for searching and filtering. Data can also be viewed as a graph or map if coordinates are provided in a table in the form of longitude and latitude. Raw data could be hard to understand, therefore a PivotTable view presents data in aggregated form and can be used to visualize data as a Table, Heatmap, Line Chart, Bar Chart, Scattered Chart, TreeMap and Area Chart etc. For more technical users, an SQL query console is added, which can be used to query one or more datasets using the SQL querying language. To better understand the dataset and its distribution, a new feature of the dataset summary is added which provides descriptive statistics using tables and charts. The feature also describes the correlation between different datasets fields in the form of a correlation matrix, which is useful in understanding relationships between the fields.

Advanced features such as text analytics, analysis and integration of related datasets and smart recommendation are added to the beta release of TET to enhance the analytics and discovery related capabilities of the platform. Text Analytics uses online a text analytics service to extract entities, keywords, concepts and relations from the text document associated with the dataset and visualize the relevance of these features using SPOD charting widgets, this helps users to understand the document without browsing it. Analysis and integration of the related dataset feature allows users to combine datasets with the same schema for analysis and visualization purposes. This is a very useful feature for analysing datasets that are published periodically and which have to be combined together for analysis needs. The smart recommendation feature uses an advanced Self Organizing Maps (SOM) based algorithm to recommended related datasets that helps users in exploring related content without much effort.

The TET beta version is designed to improve the shortcoming of the previous version, by mainly focusing of usability, human data interaction and performance. Language support is improved and all critical bugs are

fixed. Interaction between TET and SPOD is enhanced and is made as seamless as possible. The next phase will focus on improvement of these features and advanced analytics capabilities.

2.3 THE DEVELOPMENT PROCESS: DESIGN IMPLEMENTATION AND RELEASE

Source code, issue tracking, documentation and project management activities of TET are managed using the online service GitHub, which is the world's largest code repository and acts as a hub of all open source activities. Github apart from source code management provides several rich features for issue tracking, documentation and project management related tasks. GitHub is an ideal choice for agile open source projects and hosting large-scale successful open source projects.

All the development activities are managed according to the SCRUM framework. Issues and features requested by pilots are added to a product backlog and prioritized. For each sprint a list of issues are selected and sized in the sprint-planning meeting. Once the sprint is complete, the product is released after testing all the features. A sprint retrospective session is organized at the end of each sprint to learn from sprint experience, which can be used to improve future work. Figure 3 illustrates the process visually.

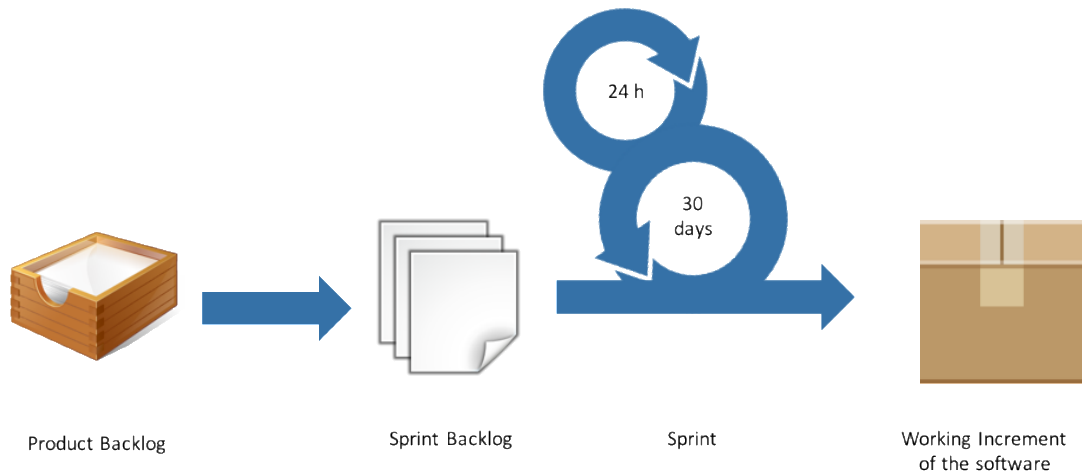


Figure 3: Scrum Process

Feedback from pilots is gathered using the feedback forms. The filled-out documents are processed and converted to the issues in GitHub and labelled by category by the development team. Issues are selected for each sprint are added and tracked using the online service Trello¹ which is a Kanban based method of project management where a project is represented as a board with a collection task list. Each list contains issues which progress from one list to another, this provides a snapshot of the project at any stage and helps management in tracking the progress of the development activities. All the development tasks and management activities are improved based on feedback collected from team members and industry best practices.

2.4 RELEASE NOTES

A list of release notes related to the beta release are given in Table 1, the release notes list details and features implemented in TET version wise.

¹ <https://trello.com/>

Table 1: TET releases notes

Name of Release	Release Notes
TET 2.0	<ul style="list-style-type: none"> Redesigned the user interface A new user centric interface was implemented. Usability of all features was enhanced to meet the requirements of ordinary citizens' needs. Distracting elements of the user interface were removed. The navigation was simplified. Enhanced search and filtering experience The search and filtering experience was improved by taking inspiration from popular search engines and ecommerce sites that have been very effective in offering intuitive user interfaces to their users. Auto-generated charts using SPOD charts components Charts with proper labelling are automatically created from the data associated with datasets and displayed as a series of slides. SPOD discussions inside the TET user interface Discussions from SPOD related to datasets are displayed in a panel on the dataset page. Fixed bugs and removed performance issues Improved permanence and critical bugs are fixed.
TET 2.1	<ul style="list-style-type: none"> Text Analytics using online service Text analytics provides a summary of the text, key entities mentioned in the text, concept related to the document and keywords shown as charts with respective relevance of each of these elements to the document. Autosuggestion for search queries Users are provided automatic suggestions when they try to enter category or role name on which they are interested, the personalised search feature is used to retrieve most relevant results wherever possible. Improved translation and bug fixes Italian and Dutch translations were improved and the critical bugs were fixed.
TET 2.2	<ul style="list-style-type: none"> Dataset summary The feature provides useful descriptive statistics to the user in the form of tables and charts. Correlation between different fields is shown in the form of an easy to understand correlation matrix. Dataset quality indicators This feature provides feedback to users about key quality indicators related to the datasets both as numbers as well as visuals.
TET 2.3	<ul style="list-style-type: none"> SQL querying An SQL query console is added for power users, which can be used to query one or more datasets for detailed analysis of tabular data. Data Integration and Analysis Automatically detects compatible datasets and allows users to combine them together. Users can view, analyse, chart or download the merged datasets. Smart Recommendation

	The features provide recommendations for related datasets using a smart Self Organizing Maps (SOM) based algorithm. The user interface lets users adjust the distance to get more or less results depending on their needs.
TET 2.4 (Beta)	<ul style="list-style-type: none"> • Dashboard for key indicator Configurable dashboard that extracts and combines key indicators from datasets and displays them visually. • Bug Fixes Fixed major bugs and issues.

2.5 CONSOLIDATED EXPERT REVIEW REPORT - ADDRESSING ISSUES

The review report did not include specific issues concerning D4.2. Therefore, in this document we draw from the general review comments provided.

Table 2: Recommendations concerning future work

#	Comment	Addressed in
R1	The use of methodology is not addressed at the necessary level and depth and the application of agile techniques in the development of SPOD and TET is not obvious. The engagement of the users is not evident in most of the deliverables and systematic information on the process and the relevant activities is largely missing across all reports. Sufficient evidence is also lacking for if and how the information from the engagement process is embedded in the deliverables from WP2, WP3 and WP4 as well in the alpha versions of SPOD and TET. Thus it is not clear if engagement is enabling/empowering the stakeholders to more effectively use SPOD and TET. Involving the pilots in the projects activities on a systematic basis is not evident, due to the lack of information and reporting of the activities and the processes involved.	<p>In the process leading to the development and release of the beta version of TET, we seriously engaged various groups of users. We carried out product testing and usability evaluation using stakeholders as follows:</p> <ul style="list-style-type: none"> • Pre-usability Evaluation testing: <ul style="list-style-type: none"> ○ Open data user category: students, researchers ○ Data Publisher category: Dublinked staff • Scenario based Usability Evaluation workshop: <ul style="list-style-type: none"> ○ Open data user category: students, researchers ○ Data Publisher category: Dublinked staff ○ PA category: Dun Laoghaire Rathdown County Council staff <p>Please refer to section 7 for details of stakeholder composition used in testing and evaluation exercises</p>
R2	Deliverables 2.1, 2.2, 2.3 and 2.4 should be revised and resubmitted to report on the used methodologies, the linkages with the other tasks and workpackages, as well as on the knowledge co-creation workflow and its capture and use in the relevant research and innovation tasks.	The deliverables mentioned were reviewed, recomposed and re-submitted as demanded in the month of May, 2016
R3	Full pilot reports should be included in one of the deliverables to inform on the processes and the activities involved with regard to the different tasks and workflows.	<p>In this deliverable, we include a full report of the testing and evaluations done within the period leading to release of beta in:</p> <ul style="list-style-type: none"> • Appendix 1A: Pre-usability evaluation testing – 2 Samples of manual with user comments • Appendix 1B: Detailed Results from Pre-Usability Evaluation Testing Mar, 2016 • Appendix 2: Usability Evaluation – Instruction Manual and • Appendix 3: Usability Evaluation – Google Survey Report

3 REQUIREMENTS

3.1 FUNCTIONAL REQUIREMENTS

3.1.1 OVERVIEW OF REQUIREMENTS

Initially, prior to commencement of implementation of system components, we carried extensive requirement gathering tasks and reported the results in deliverable **D2.4 Requirement Specification and Use Case Models** along with the Use Case tables and models. The document also showed the use case diagrams and tables and every part was fine tuned in Deliverable **D4.2 Alpha Version of TET** to support the release of alpha version of TET. The same deliverable D2.4 also explained how we achieved requirements traceability back to the earlier tasks of the project including requirements elicitation and analysis, and their respective documentation were planned and executed to align with ROUTE-TO-PA objectives and the democratic values they demonstrate or exemplify. These objectives include:

- Integration of related datasets using Linked Data principles
- To provide basic analytics on datasets to detect violations of rules (e.g. transparency related rules), that is, where metadata supply for the purpose of enhancing data searchability and linking is lacking through the metadata completeness measurement.
- Personalisation – exploit of social media profiles of users, e.g. demographic features and interests, the possibility of analysing contents contributed on social media (e.g. topic detection).
- Dataset Quality – generate reports on quality of metadata and provenance information (e.g. based on specific standards) – through view metadata functionality
- Recommendation – cluster of datasets based on categories, and usage patterns (e.g. datasets downloaded or requested together) as popular searches and consumption pattern or as natural categories e.g. budget datasets, population datasets, economic datasets, etc.

These objectives were mapped to see how they demonstrate democratic values: Monitorial Democracy, Deliberative Democracy and Participatory Democracy in the earlier documentation referred to above and modified in Table 3.

Table 3: Overview of the Societal Activity model of Open Data use

Use of open data	Monitorial democracy Monitoring government behaviour	Deliberative democracy Feeding public debates	Participatory democracy Enabling collective action
Object	Policy issue	Policy issue	Policy issue
Subject Citizen	Watchdog	Partner in dialogue	Partner in Action
Tool	- TET ensures transparency by access to data, data search and tools to visualize the data - SPOD facilitates data sharing	- TET ensures access and enables data analysis, personalization, comparing and combining datasets and feedback options - SPOD facilitates participation with debate and feedback options	- TET ensures access, decision-making tools, the ability to add data - SPOD facilitates collaboration with government and other stakeholders e.g. refining solutions and working together in documents
Rules/Mechanism	Transparency, checks and balances, e.g. foia laws, privacy laws, but also rules regarding the virtual community	Deliberation, open communication e.g. procedures regarding participation	Collaboration or co-creation e.g. procedures regarding who is invited (based on expertise)?
Community	Government, journalist	Government and citizens	Government, citizens, businesses, researchers
Role e.g. government	Provider of information	Facilitating dialogue	Partner in Action

Outcome	Critical view on government behaviour	Contribution to debate about policy issue	(Collective) action to produce public value
----------------	---------------------------------------	---	---

In this section of the document, we present the new and the updated versions of system requirements with reference to the earlier documentation on the ROUTE-TO-PA system requirements deliverable **D4.2 Alpha Version of TET** previously submitted. As often and necessary, we refer to this deliverable D4.2 as the main documentation covering system requirements being updated by this particular deliverable **D4.5 Beta Version of TET** while keeping track of the requirements traceability through use of Use Case and User Story IDs. It suffices to note that in D4.2, we informed users of the report (D4.2) that the specific TET requirements then under implementation included those in table 4 of D4.2, hereunder reproduced (as Table 4) for quick reference purpose:

Table 4: Complete System (Solution) Requirements - TET (implemented components only)

Use Case	U/Story ID	Functional Requirements	TET Objectives	System Quality Requirements
UC23		User Login	The system provides login form data fields for you to complete and save to user account and to enable the user to enjoy a better experience with the platform.	Availability, Operability, Portability, Data Security and Privacy, Simplicity
UC1	S29.1	Check Metadata Completeness	Visualise metadata completeness quality for a dataset by displaying the measurement bar, showing percentage rating for the completed metadata fields for the dataset to enable users to decide on the suitability for use or fit-for-purpose	Usability, Simplicity, Integrity, Accessibility, Accuracy
UC2	S29.3	Add/Update/View Provenance and metadata-related to a dataset	System support for provenance records: values entered for the – source, versions (current & previous) and frequency of publication fields in the metadata screen of the dataset.	Usability, Operability, Auditability, Integrity and Auditability
UC10	S1.4	Analyse a Dataset	System provides simple analysis, analytics and visualisation tools to enable users (including non-tech savvy users) to analyse datasets & visualise results in common graphs to enhance comprehension of the data meaning	Usability, Supportability, Simplicity, Clarity, Correctness, Intuitiveness, Performance, Accuracy
UC14	S8.1	Link a Dataset	The part of this objective so far implemented is the part that enables dataset linking for searchability and recommendation purposes as support for dataset accessibility. Dataset linking for the purpose of dataset integration has not been implemented.	Supportability (file formats), Usability, Operability, Extensibility, Traceability and Accessibility.
UC17	S28.2	Personalize Search (User not logged in)	The system provides options e.g. role, location, dataset category, etc., for the user to personalise search by default. This supports dataset accessibility and improved search experience for users	Operability, Usability, Reliability, Accuracy, Data

				security & Privacy, Performance
UC1 7	S28.2	Personalize Search (User logged in)	The system provides options e.g. interest, role, location, dataset category, etc., for the user to personalise search as system support for dataset accessibility and improved search experience for users.	Operability, Usability, Reliability, Accuracy, Data security & Privacy, Simplicity, Performance
UC1 9	S26.5	Recommendations for Datasets	Provide the user with dataset recommendation options in the form of Dataset category or subject or by popular searches to enable user search within recommended options while not logged in.	Usability, Personalisation, Simplicity / Interface friendliness, Accuracy and Performance
UC1 9	S26.5	Request Recommendations for Datasets	Provide the user with options (location, role, age, email, sex, marital status, disability, interest, etc.) to enable user personalise account & receive content suggestions based on account data. <i>(Not implemented yet: option for user to opt in for data recommendation alert via email)</i>	Usability, Personalisation, Simplicity / Interface friendliness, Data Privacy & Security, Integrity and Performance
UC2 4	S38.1	Enrich Profile	System provides more personal data options e.g. location, age, marital status, sex to enable users adds more data to profile so that user can search & receive more relevant resources based on his/her enriched profile details	Operability, Accuracy, Accessibility, Dependability, Data Privacy & Security

3.2 NEW AND UPDATE TO FUNCTIONAL OR SYSTEM REQUIREMENTS

Some new use cases shown in Table 5 were implemented for the first time in the beta release of TET. The listed use cases are among those elicited and analysed prior to the commencement of development work and although were shown in deliverable D4.2 as Table 5 (List of ROUTE-TO-PA Use Cases showing the groups of the various cases), they were not implemented in alpha release of TET.

Table 5: Earlier System Requirements not in Alpha but in Beta Version of TET

UCase ID	U/Story ID	Functional Requirements	TET Objectives	System Qualities Requirements
UC4(A)	S2.3	Provide Supporting Content from SPOD Post	System should import posted text contents as related discussion on the dataset posted by user on SPOD	Usability (operability simplicity) Understandability (dependability, accuracy & clarity)
UC4(B)	S2.1	Share/ Export (visualised) dataset from TET to SPOD	System should enable file export and import, file sharing capabilities from TET to SPOD post	Usability (operability simplicity, performance, support for file formats) Understandability (dependability)
UC11	S32.2	Query Dataset	System provides SQL query builder should execute queries to display queries results to user.	Accessibility (Portability, availability, publicity); Usability (Simplicity, User-friendliness); Understandability (Conciseness); Informativeness (correctness)
UC12	S8.1	Integrate Datasets (within OD platform)	System allows user to select multiple datasets, combine them and analyse as a single dataset. The part implemented is combining	Usability (inter-operability, supportability, performance, adaptability, simplicity); Accessibility (availability); Understandability

			datasets within a platform and not across platforms.	(extensibility, dependability); Informativeness (integrity, accuracy)
UC20	S2.2	View Charts	System provide user with tools to select from and use to visualise dataset in common graphical format. View as infographic is not yet implemented	Usability (support for analysis & visualisation tools, operability, intuitiveness, performance); Understandability (decomposability, composability); Informativeness (accuracy, correctness, clarity)

In order, to meet up with improving demands of the users towards a better system usability quality, the system requirements were updated with feedbacks from various technical and common (citizen) user consultations carried out within the period from March to November, 2016, Figure 4. The relevant user consultation sessions and a workshop as well as the breakdown of the compositions of the user categories are detailed in section 7.2 of this document. In addition to the above updates, some of the earlier requirements (Figure 4) – elicited and analysed prior to the commencement of development works and which were shown in Table 5: List of ROUTE-TO-PA Use Cases showing the groups of the various cases of D4.2 – but which were not implemented in alpha version have been developed in the Beta version.

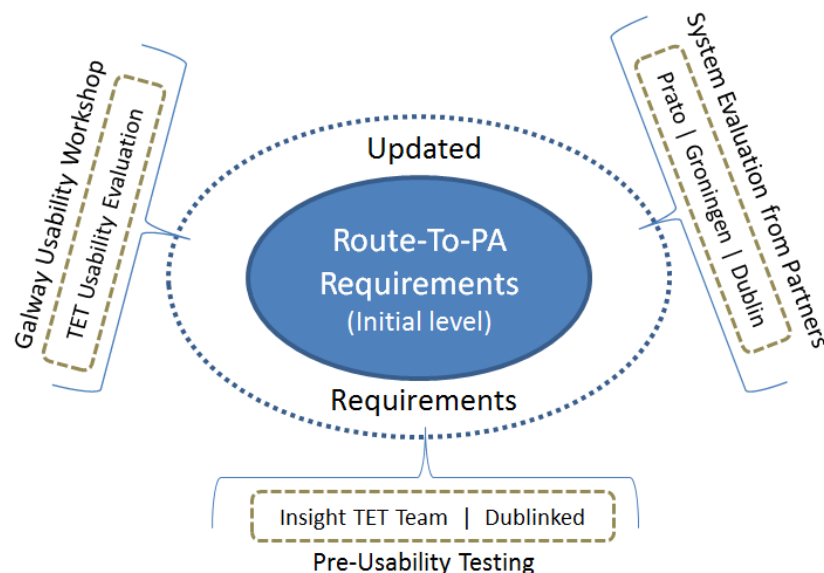


Figure 4: ROUTE-TO-PA Requirements upgrade from Evaluation Feedback.

3.2.1 COMMENTS FROM PRE-USABILITY EVALUATION TESTING

For TET, the first (pre-usability) evaluation testing that took place was organised by Insight (NUIG) whereby some staff and student members of Insight Centre NUIG were recruited along with some staff of Dublinked open data portal as well as staff of Dun Laoghaire Rathdown County Council to test the functionality and usability of TET tools. In total, there were eleven testers made up of 7 males and 4 females of which were

- Potential open data platform users – represented by 2 computer students, 2 open data experts, 3 public administration representative, and
- Open data publishers – represented by 4 open data portal staff.

These rough testing exercises produced important feedbacks (Table 6) that were used in improving the alpha versions iteratively – [alpha v1](#) and then [alpha v2](#). Details of the evaluation manuals and the report are show in section 7 and as Appendix 1A: Pre-usability evaluation testing – 2 Samples of manual with user comments and Appendix 1B: Detailed Results from Pre-Usability Evaluation Testing Mar, 2016.

Table 6: Pre-Evaluation Testing Feedbacks and Implementations thereof

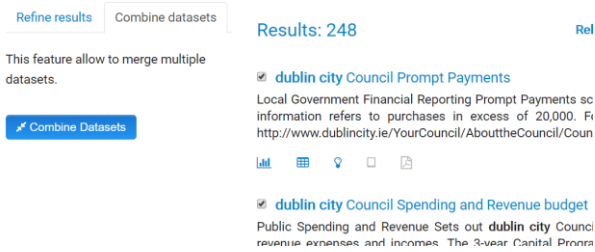
Testers' Feedbacks as update to requirements	Interpretations and Implementations of updated requirements (comments)	Update to D4.2: UC IDs and U/Story IDs														
1. If you search with category search word, you have a broad result, some of which might not be very relevant to your need.	Interpretation: Seeking keyword search capability Implementation: Keyword query capability provided in later release of alpha version	Dataset searching issues, link to: U/Case ID: UC17 U/Story ID: S28.2														
2. There need for date field that could be auto-completed for dynamic data being uploaded on the site to shows also the frequency of data updates.	Interpretation: Seeking date: created or uploaded and date of last update for datasets Implementation: The Implementation which features the display of metadata completeness table for each dataset including relevant dates. Further enhancement of this feature was carried out after the main usability evaluation workshop which took place August 2016. The feature is shown in the beta version – see item 4 & 5 in Table 8 below	Metadata issues link to: U/Case ID: UC2, UC1 U/Story ID: S29.3, S29.1														
3. Manner of Implementation of the compulsory metadata form should not become a hindrance for data suppliers. Include help prompts for data publisher	Interpretation: Data suppliers should not be restricted from uploading dataset due to inability to supply all fields of metadata table. Provide reminders for relevant fields to be supplied Implementation: Offer of suggestion of what field of metadata table has not been supplied instead of restricting data upload – see item 4 & 5 in Table 8 below <table><tr><th>Field</th><th>Value</th></tr><tr><td>Language</td><td>English</td></tr><tr><td>Target Audience</td><td>Citizen</td></tr><tr><td>Category</td><td>Transport</td></tr><tr><td>Maintainer</td><td>data@dublincity.ie</td></tr><tr><td>Maintainer Email</td><td>data@dublincity.ie</td></tr><tr><td>Completeness</td><td>65 %</td></tr></table>	Field	Value	Language	English	Target Audience	Citizen	Category	Transport	Maintainer	data@dublincity.ie	Maintainer Email	data@dublincity.ie	Completeness	65 %	Metadata issues link to: U/Case ID: UC2, UC1 U/Story ID: S29.3, S29.1
Field	Value															
Language	English															
Target Audience	Citizen															
Category	Transport															
Maintainer	data@dublincity.ie															
Maintainer Email	data@dublincity.ie															
Completeness	65 %															
4. On interface friendliness and learnability, there should be auto-suggestions while searching. Users would have to go through all datasets to be sure datasets displayed are relevant ones	Interpretation: Include refining tools to improve search result relevance Implementation: Inclusion of general filter options on the right panel of the screen: Location, Period, etc. and specific filtering options: Relevance, Name (ascending/descending orders) and Dates in a drop-down menu. <table><tr><td><div><div></div><div></div><div>Order by: <div>Relevance <div></div></div><div>Relevance</div><div>Name Ascending</div><div>Name Descending</div><div>Last Modified</div></div></div></td><td><div>Organizations</div><div>Dublin City Council (84)</div><div>Groningen (16)</div><div>City of the Hague M... (12)</div><div>Insight (4)</div></td></tr></table>	<div><div></div><div></div><div>Order by: <div>Relevance <div></div></div><div>Relevance</div><div>Name Ascending</div><div>Name Descending</div><div>Last Modified</div></div></div>	<div>Organizations</div> <div>Dublin City Council (84)</div> <div>Groningen (16)</div> <div>City of the Hague M... (12)</div> <div>Insight (4)</div>	Dataset search & UI issues. Link to: U/Case ID: UC17 U/Story ID: S28.2												
<div><div></div><div></div><div>Order by: <div>Relevance <div></div></div><div>Relevance</div><div>Name Ascending</div><div>Name Descending</div><div>Last Modified</div></div></div>	<div>Organizations</div> <div>Dublin City Council (84)</div> <div>Groningen (16)</div> <div>City of the Hague M... (12)</div> <div>Insight (4)</div>															

3.2.2 FEEDBACKS FROM OTHER PILOT SITES

The table below (Table 7) shows some of the results of the system reviews and evaluations done in other sites. These were not based on scenarios testing.

Table 7: Review Results from Other Testing and Pilot Sites

Testers' Feedbacks as update to requirements	Interpretation and Implementation of updated requirements (comments)	Update to D4.2: UC IDs and U/Story IDs
	Groningen	
1. User profile: it seems just one topic of interest can be selected, would it be useful to have more?	Interpretation: User seeks enrichment of user profile features in respect being able to select more than one	Enrich Profile. Link to: U/Case ID: UC24 U/Story ID: S38.1

	Implementation: This issue has not been fixed in the current beta release. However, it has been marked down to be fixed in next release.	
2. From a search on "Environment" I got also this dataset, it seems it is not related with such keyword, since "Category" is "Local Government"	Interpretation: User seeks improved search result accuracy based on keyword. Implementation: This requirement was implemented via introduction of more filtering options both on the right (e.g. Location, Period, Themes, etc.) & sorting options (Relevance, Name & Data) are tools to improve accuracy of searching results. We also improved metadata records which is used in searching functions also	Dataset searching issues. Link to: U/Case ID: UC1, UC2, UC17 U/Story ID: S29.1, S29.3, S28.2 respectively
3. How can I create groups of datasets? It seems no functionality is active in the "Group" page	Interpretation: The user seeks functionality to select and combine dataset so as to analyse them together as single dataset. Implementation: This use case was implemented in form combining datasets within TET in ROUTE-TO-PA platform. In the search results view of TET interface, there is a 'Combine Datasets' tab in which you have a 'combine datasets' button. In the tab, user can select dataset of interest and click on combine button. The system will then indicate if the selected datasets are compatible or incompatible for combination and analysis purposes 	Integrate/combine a dataset. Link to: U/Case ID: UC12, U/Story ID: S8.1
4. How can I become a member of an organization? Should it be written in the profile?	Interpretation: This question was relevant to old TET interface which was previously build direct on CKAN infrastructure Implementation: The problem has been fixed by building a complete level over CKAN infrastructure	Dataset search & UI issues. Link to: U/Case ID: UC17 U/Story ID: S28.2
5. API visualization is still wrong (screen 1280x800, Chrome, Windows 7)	Interpretation: This was overlapping of two screen views (Data explorer view was laying over API screen) Implementation: The problem has been fixed and these view are now separate with overlapping.	N/A
6. When clicking on a dataset there is no Auto-generated chart like there is in the General TET. There are just the two pointers, which when you click on them do not do anything. See BUG.1	Interpretation: This was a temporary problem caused by changes in SPOD. Implementation: Problem is now fixed	Analyse a dataset. Link to: U/Case ID: UC10 U/Story ID: S1.4
7. When you click 'Export to SPOD' when creating a chart in the pivot table, it gives an Internal Server Error. See BUG.2	Interpretation: Integration error. Integration issues for navigability. Implementation: Fixed	System usability for main operations. Link to: Analyse/visualise a dataset: U/Case ID: UC10 U/Story ID: S1.4
1. A lot of the text is still in English. Actually only a few words are in Dutch. We need to translate the site before it is usable for others. See ENHANC.1	Interpretation: Language translation issues Implementation: Problem now fixed	N/A
2. If you for example plot a Bar Graph it does not display the unique values per variable. Displays only when you hover over it.	Interpretation: CKAN issues that could not be fixed at the moment Implementation: The problem has been marked down for fixing in the future by redesigning the data visualisation mechanism	N/A

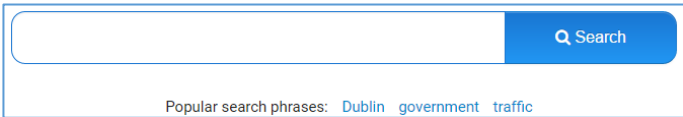

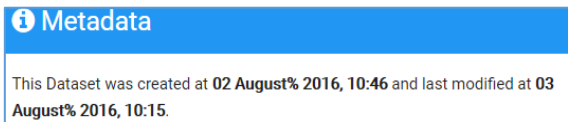
	<u>Prato</u>	
1. On the "Trust this site" page the link to the EU flag is broken (see screenshot2) and Bar on the top covers the interface	Implementation: fixed	N/A
2. In the Category list the word "Ambiente" is doubled. We see that while in the CKAN version of the site also Italian language is supported (probably due to the CKAN feature) the TET parts will be translated into Italian. If you send us the "strings" in English of the site to be translated, we can translate it into Italian a.s.a.p. (this is a pre-requisite for the usage of TET in the Prato pilot)	Interpretation: Translation problems Implementation: All old translation issues has been fixed. New ones will arise and they will be fixed accordingly	N/A
	<u>Dublin</u>	
1. Remember me on the login screen doesn't work – it doesn't remember you when you log out and return – check box is ticked but no details are retained.	Interpretation: This was due to the user' browser issue not TET Implementation:	
2. See Appendix A, Example 3 -Export to SPOD and Embed buttons – not working	See Groningen #7 above	See Groningen #7 above
3. System still requires 2 separate logins for TET and SPOD	Interpretation: Integration problem Implementation: TET and SPOD now better integrated	Login and integration issues: Linked to: UC23
4. Include testimonials to promote tools	Interpretation: This could mean enhance adoption through learnability Implementation: We implemented tooltips and demo videos and better interface friendliness	
5. Social media widget to spread the word about TET and SPOD	Interpretation: Include sharing tools and discussion tools Implementation: Data and related visualisation sharing tools implemented in addition to dedicated SPOD interface functionalities	No specific use case but many use cases.
6. Why is the metadata on the dataset not completed? Please see table in Appendix – A, Example 4 for elaboration and mapping of metadata tags.	Interpretation: Incomplete fields in the metadata table of the dataset Implementation: This is meant for data publishers and uploaders to complete. Not an issue from TET implementation	N/A

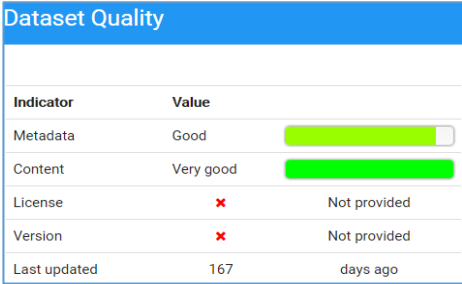
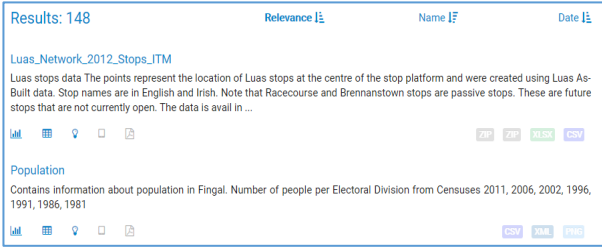
3.2.3 FEEDBACK FROM USABILITY EVALUATIONS

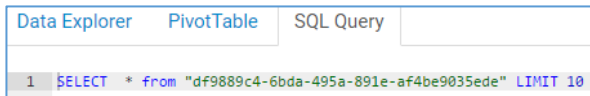
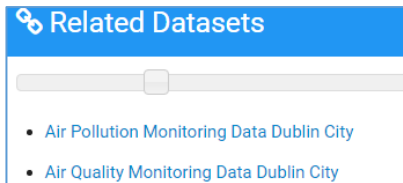
From the main usability evaluation workshop (Galway, 03 August, 2016) we obtained a number of user feedbacks useful for further upgrade. While details of usability evaluation feedback are presented in section 7, below is a list of comments (Table 8) from a wide range of stakeholder evaluators which included the age groups represented by the 19 users (10 males and 8 females, 1 unknown) who participated in this evaluation workshop range from 18 to 54 years old. The most represented of these age brackets was the group of 35 – 44 years old (i.e. 7 or 38.9%) followed by the group of 25-34 years old (i.e. 6 or 33.3%). Evaluators who have no significant experience in the use of open data were highest in number (5 or 27.8%) thus accounting for the ordinary citizen stakeholder group, followed by those with 2-3 years of experience (4 or 22.2%). In terms of functions, Student evaluators were more in number (8 or 44.4%), followed by Researchers (5 or 27.8%) and next were the Public

Servants (4 or 22.2%). Software Developers were the least represented in number (1 out of 19 or 5.6%). The sample of users who participated in evaluating the ROUTE-TO-PA tools in Galway workshop and in the survey that followed the evaluation task is quite representative of a typical open data community because we had representatives from the sectors such as education, public services, research and the software applications development. However, important sectors not represented include business, health and transport sectors; however, these sectors will be included in the subsequent evaluation exercises for later releases of ROUTE-TO-PA system. Details of the results from the usability exercise are presented in section 0; as well as in Appendix 2: Usability Evaluation – Instruction Manual and Appendix 3: Usability Evaluation – Google Survey Report.

Table 8: Usability Evaluation Comments and Implementations

Evaluators' Feedbacks (requirements update)	Interpretation and implementation of updated requirements (feedbacks)	Update to D4.2: UC IDs and U/Story IDs
1. When adding new features, consider optimisation not to miss the view of user interest. Optimise for Google's style in mind	<p>Interpretation: Seeking keyword search capability and seeking clutter-free, Google style landing page.</p> <p>Implementation: A complete re-design of the user interface with inclusion of keyword query capability and the introduction of 'Google' style landing page by removing screen clutters and adding search bar to improve user friendliness, quick search by keywords and provision of popular search topics.</p> 	<p>Searching and UI friendliness functions.</p> <p>Link to: U/Case ID: UC17 U/Story ID: S28.2</p>
2. Use advance options only when needed	<p>Interpretation: Asking for the removal of very technical tools that can only be used by tech savvy people.</p> <p>Implementation: We implemented this requirement by putting all advanced features for advanced users into the CKAN classic view (on the Top menu) instead of totally removing these features. These features are now preserved for the advanced users</p> 	<p>New feature supporting UI friendliness and system learnability linked to Analyse/visualise a dataset:</p> <p>U/Case ID: UC10 U/Story ID: S1.4</p>
3. Search results should be provided with statistics	<p>Interpretation: This is not very clear to our understanding. Perhaps the user want data search results to be displayed with better statistics e.g. in categories and showing number of dataset in each category,</p> <p>Implementation: This was implemented by providing the facets on the filter tools on the left panel of the screen including the values for categories such as Location, Period, etc.</p>	<p>Searching and UI friendliness functions.</p> <p>Link to: U/Case ID: UC17 U/Story ID: S28.2</p>
4. The need for date field that could be auto-completed for dynamic data being uploaded on the site to show also the frequency of data updates.	<p>Interpretation: Seeking date: created or uploaded and date of last update for datasets</p> <p>Implementation: The implementation which features the display of a metadata completeness table for each dataset including relevant dates and removal of the metadata completeness percentage was carried out after the main usability evaluation workshop which took place August 2016. The feature is shown in the beta version</p> 	<p>Metadata issues link to: U/Case ID: UC2, UC1 U/Story ID: S29.3, S29.1</p>

<p>5. Manner of implementation of the compulsory metadata form should not become a hindrance for data suppliers. Include help prompts for data publisher To avoid incomplete dataset file, when uploading dataset, Data Officer should be warned against missing fields – Error alerts</p>	<p>Interpretation: Data suppliers should not be restricted from uploading dataset due to inability to supply all fields of metadata table. Provide reminders for relevant fields to be supplied Implementation: Offer of suggestion of what field of metadata table has not been supplied instead of restricting data upload.</p> 	<p>Metadata issues link to: U/Case ID: UC2, UC1 U/Story ID: S29.3, S29.1</p>
<p>6. On interface friendliness and learnability, there should be auto-suggestions while searching. Users would have to go through all datasets to be sure datasets displayed are relevant</p>	<p>Interpretation: Include refining tools to improve search result relevance Implementation: Inclusion of general filter options on the left panel of the screen: Location, Period, etc. and specific filtering options: Relevance, Name and Date. Also provided for by including dataset description both in the search result (under each title) and displayed in the description box when a dataset is selected.</p> 	<p>Searching and system usability, learnability UI friendliness functions. Link to: U/Case ID: UC17 U/Story ID: S28.2</p>
<p>7. Need to see the table before charting or as Tabular mode</p>	<p>Interpretation: Maybe charting is not intuitive enough – requires user knowledge of dataset in its tabular form to be able to understand it and then chart it properly. Implementation: Provided the following groups of views as individual box with relevant details about the dataset selected: Immediate view:</p> <ol style="list-style-type: none"> 1. Description – about the dataset 2. Metadata elements and relevant dates 3. Dataset quality description box 4. Table structure 5. Discussions on SPOD <p>Further view: under the 'View the Summary' button</p> <ol style="list-style-type: none"> 1. Box plot – shows the Median, quartiles and whiskers. 2. Correlation matrix – shows relationship between parameters in the dataset 3. Summary of the dataset field 4. Description – Description of dataset in text 5. Details – of the field and counts of parameters in the dataset 	<p>New feature supporting UI friendliness and system learnability linked to Analyse/visualise a dataset: U/Case ID: UC10 U/Story ID: S1.4</p>
<p>8. Data load time is long – slow loading</p>	<p>Interpretation: Poor system performance Implementation: Performance enhancement was carried out for this comment</p>	<p>New feature demanding system performance. Linked to Analyse/visualise a dataset: U/Case ID: UC10 U/Story ID: S1.4</p>
<p>9. Social Media-like link in TET to transfer to SPOD 'My space'. Both [TET and SPOD] platforms</p>	<p>Interpretation: Create a button in TET to switch a user to SPOD Implementation: This was implemented by TET/SPOD integration and then creating a 'Export to SPOD' button in TET - inside the PV Table view – to seamlessly switch a user from TET to SPOD. However, entering SPOD in this manner requires user login. The process login at this point</p>	<p>New feature: UI integration issues for navigability. Link to: System usability for main operations:</p>

should be more integrated.	has not fully implemented but has been marked down for next release of beta version.	Analyse/visualise a dataset: U/Case ID: UC10 U/Story ID: S1.4
10. I want popup tooltips. Tooltips and user demo video are needed to provide hints to users in the interfaces	Interpretation: lack of user help or user guide on platform. Implementation: Tooltips provided for buttons and icons to improve learnability and interface navigability as well as friendliness. YouTube videos have been created and published as well. A link will be provided on the TET interface leading users to the YouTube videos as demo materials on how to use system tools	New feature demanding system usability, learnability & navigability UI friendliness functions. Link to: U/Case ID: UC17 U/Story ID: S28.2
11. When several values are selected in the bar chart, some bars cover others in the bar chart.	Interpretation: This comment is asking for enhanced visualisation of datasets Implementation: This issue arises from the CKAN problem and at the current version has not been updated. It has been marked down for upgrade through visualisation enhancement in next release of beta version.	New feature demanding system performance. linked to Analyse/visualise a dataset: U/Case ID: UC10 U/Story ID: S1.4
I Like the auto generated graphs; allows the user to explore the dataset	Interpretation: Auto-generated graph is important to users Implementation: This feature has been enhanced in the beta version to enable fast display of graph and a clearer version than before – performance enhancement. More needs to be done in next release.	Demanding better system performance. linked to Analyse/visualise a dataset: U/Case ID: UC10 U/Story ID: S1.4
12. I want a process or method used to determine Datasets	Interpretation: This comment was interpreted as requesting for a method to determine or find a relevant dataset for use Implementation: We provided an additional 'SQL Query' button within the dataset view as a table window to enable users query all of the selected fields and parameters within the dataset. 	New feature requesting inside dataset searching function. Link to: U/Case ID: UC17 U/Story ID: S28.2
13. Accessibility issue with transparency in searching function	Interpretation: Making datasets more transparent to searching functionality and to users on platform. This asks for 'Related dataset suggestions' to users so they can see those datasets and view them Implementation: This was catered for the provision and enhancement system tools to provide dataset recommendation of the related datasets to the dataset searched by user. This is seen under the 'Related Datasets' box within the selected dataset view. We provided a slider button to be used in relaxing or reducing the level of relatedness of datasets recommendations 	Requesting assurance of dataset searching function support for data visibility through system recommendations. Link to: U/Case ID: UC17 U/Story ID: S28.2

3.3 SYSTEM REQUIREMENTS

3.3.1 EXPLANATIONS OF SYSTEM REQUIREMENTS

As explained in D4.2, that the functional requirements (use cases) were derived from a consideration of user stories and the breakdown of the specific and unique functionalities and capabilities that the system must offer to meet the needs of users expressed in the user stories that were provided by stakeholders. The earlier stories

have been replaced by usability evaluation processes carried out by potential users and data publishers to see how well the system functionalities so far implemented addressing their desired needs. However, we have to ensure also that the system (non-function) requirements are adjusted to enable improved quality of tools and also to ensure that the system can facilitate open data transparency and social discussion on open data. In this section, we present facts and relationships that relate to system support for the intended aims and objectives of democratic values embedded in the ROUTE-TO-PA concept of transparency enhancement and social platform for open data to the extent of the current release of the beta version.

Accessibility

The beta release ensures users gain access the system tool at all times and login into the system or click on a button to invoke the function. Login is not required for some functions such as searching and viewing a dataset; however, when a user wishes to export a dataset into SPOD for further analysis and visualisation, the user will be required to login. System performance has been enhanced to enable system performance operates as intended and reduce user frustration or delays in displaying outputs. Accessibility is closely related to the understandability and availability qualities in that if the system poses a difficulty in assimilating the instructions on how to use a tool. To reduce this problem, we have included tooltips and demo videos on the platform to serve as user help and guides on how to deploy system functionalities. Although no special tools have been implemented to support adaptability for use by people of special considerations such as people with disabilities, this plan will be actualised in future releases of the platform. For data transparency purpose, the implemented searching and filtering tools should ensure visibility of open datasets on portal and this is supported by the implementation of the dataset recommendation functionality which has been commended by users.

Accuracy

The system should work as expected and produce the result that is expected based on the searching and query mechanisms for output relevance. Thus, a user should find the system in operation according to how it is intended to function and produce expected output. The filtering options both on the right (including **Location**, **Period**, **Themes**, etc.) and sorting options (including **Relevance**, **Name** and **Data**) are tools to improve accuracy of searching results.

Auditability

As explained in D4.2, this quality relates more to the capability of TET tool to permit an efficiency check rather on the data transparency feature it tends to support than the working capability of the functions itself. Its auditability feature will permit a systematic check or assessment of the open data practice or standards it supports to allow a user understand the accuracy, efficiency or effectiveness of the dataset or other open data resource operations it supports. Such understanding may help a user decide on certain issues pertaining to use of a given dataset resource on the platform. To support this quality, we implemented a strong metadata completeness functions that alert data publishers to complete all required field provided in the metadata table. Although a publisher is not stopped from uploading a dataset if all fields are not completed, however, the implementation ensures that suggestions of fields not yet completed are made to the publisher as reminder to do so.

Availability

The beta release strengthens the tools alpha version to ensure better system performances so that as explained under accessibility, availability quality, should uphold the feature or tool or functionality of the system in place for the user to use at all times. The system must not leave the user in any form of frustration while attempting to use the resources due to the absence of a tool or lack of help. Again, we provided for easy learnability of system tools to facilitate usability thereby making the platform available for user's use. The quality of system availability extends to the capability to keep uptime in good measure – that is high – over a given period of time.

Data Privacy and Security

In cases of login with personal details, the systems must adhere to the required standard of maintaining the safety of user personal details and maintain the user account and profile data against unauthorised access. This was not compromised during the implementation of the login use case.

Integrity

The system should maintain a level of standard that can be adhered to as moral principles of completeness and wholeness in output and the quality thereof as stated governing professional or occupational standards in the domain. We implemented functionalities that ensure standard out structures and clarity for better understandability and obey the format of other known systems for compatibility and interoperability purposes.

Interface and User friendliness

We improved interface friendliness especially the landing page from alpha versions through to current beta version. We now have a clutter-free Google style landing page and enable keyword search with recommendation of popular topics. This helps users who are already used to searching topics on Google platform and which they also recommended as usability evaluation feedback. The system interface is the point of interaction between the system and the user or another system. On the human user interface point, the system should present an attractive easy to engage windows containing screen features that are not hostile to the eyes. Generally, the interface should adhere to ergonomic standards and contain readable fonts that the user would always like to return to at any time.

Operability

The buttons and tools implemented must be deployed in the manner intended to allow the user to use them to achieve the desired goals. Thus, there should be no unexpected errors or hitches capable of forestalling the appropriate interaction of the user with the system. While this a continuous improvement consideration, we have ensured enhanced system operations as much as possible through system integration between TET and SPOD interfaces as advancement from alpha to beta releases, simple to use analytic and visualisation tools that are no so technical in nature. We have hidden from view all technical tools by putting them into the CKAN classic view for the expert users. Simple drag and drop functions in the PV table view enable non-expert users to build a visualisation chart from datasets of interest to them.

Performance

Performance quality relates to the capability of the system to execute operations with high efficiency and speed. Alpha version was slow to use as revealed from the pre-usability evaluation testing. However, we continued to improve on the quality through the releases with improved feedback from the main usability evaluation feedback. Now is much better in the current beta version having done a great deal performance enhancement on the system.

Reliability

Reliability quality calls for the capability of the system to maintain the expectations that a user places on the system. Reliability, therefore, relates to the factors like performance, availability or meeting up to the standard of delivering the query results with accuracy. Again the upgrade work on performance support reliability as well as the tools implemented to improve searching accuracy such as result filtering and sorting options. Nonetheless, this quality cannot be fully attained until completion of the project when all features would have been implemented.

Simplicity

Simplicity is the quality that ensures that users are able to learn the application and use it without much stress in understanding how to apply the tools. User friendliness of the interface may be enhanced by the quality of simplicity which avoids unnecessary complexity in the interface or windows of the application. Part of our aim is to produce a system usable by non-computer experts. Therefore we have continued to remove very technical tools from the interface (as in CKAN Classic view) and introduce more user-friendly features as in the introduction

of help facilities such as tooltips and videos for better learnability and simplification of applications. Clutter-free interface supports user-friendliness.

Usability

Usability quality sums up all features into the quality offered a user to use the system for a specific purpose in a given context. We are trying to learn more about the usability of the system by conducting various usability evaluation exercises and collecting evaluators' feedbacks and implementing their complaints and suggestions. All these are geared towards attainment of better usability quality of ROUTE-TO-PA system.

Supportability and Inter-operability

We know that the system will have to operate with other systems, and it is hard to find a standalone system that does not communicate with other systems or platforms. To operate smoothly with other systems, components and interfaces, a system must maintain the capability to utilise standardised file formats, synchronised data components and other required system protocols. In this direction, we have started implementing login authentication from other platforms such as Facebook although this is at very early stage of development. TET and SPOD integration (intra-operability) has been achieved to a great extent, however, future releases will see enhance support for interoperability.

4 TET ARCHITECTURE AND DESIGN

This section documents the TET Architecture, its components and relationships between them in order to achieve the requirements for Transparency-Enhancing Toolset (TET).

4.1 ARCHITECTURE PRINCIPLES

In order to understand the requirements for Transparency-Enhancing Toolset (TET) Architecture we analyzed a selection of existing Open Data Platforms and followed User Stories described in D2.3 (ref. *D2.3 User stories on Open Data and Transparency*). The selection of Open Data Platforms was made by the analysis of the latest publications concerning Linked Open Data and of the usage of the available platforms. Moreover based on publicly accessible documentation (projects websites, documentation, publications, press releases) each platform was analyzed in order to investigate the key components and relations between them. The patterns observed across the reviewed Open Data Platforms, and research results were used to create the TET Architecture.

Architecture is defined as *the fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution* (IEEE-SA Standards Board, 2000). The purpose of architecture is to guide the development and the implementations of the platform. Moreover, it allows understanding of the requirements, the workflow and the challenges faced during the deployment of the solution.

4.2 TET REFERENCE ARCHITECTURE

Generic TET reference architecture is presented in Figure 6 , and it is organised in three Architectural Layers: (i) User Interface, (ii) Services, and (iii) Storage. A detailed description of the layers is described in section 4.3 while the implementation details are explained in section 0.

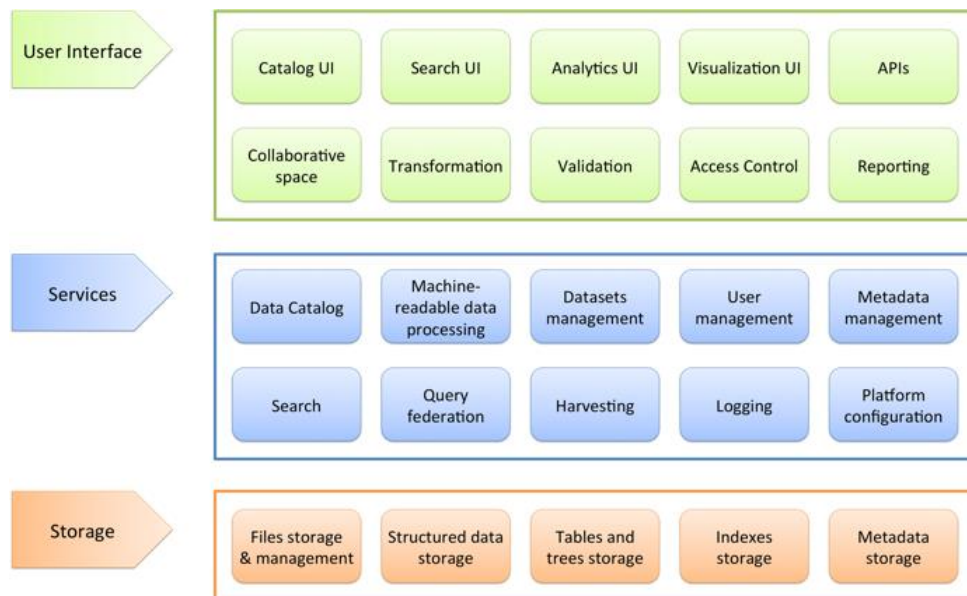


Figure 5 TET Reference Architecture

4.3 ARCHITECTURAL LAYERS

Different layers in the software architecture are allocating the different responsibilities. In our case, the layer at the bottom handles the data. The middle layer can be viewed as a layer of the high level of services provided on top the data while the top layer provides services to users and applications. Details of each layer are as follows:

User Interface layer

The User Interface layer provides basic portal functions such as access to the data, search interface, personalization and customization features, etc. The search feature allows users to find quickly information stored

at the portal, while analysis and visualizations features allow users to explore, analyse and visualize various types of data, such as tabular and geospatial data. Various APIs allows external applications to consume services offered by the platform.

Services layer

The Services layer provides services on top of Storage layer that can be exploited by the User Interface layer. Data Catalog services are used to list the details of datasets and associated metadata stored in the platform. Search service uses the index to search relevant content. Platform extensions services allow external applications to use the platform services. All these services have the corresponding features in the interface layer.

Storage layer

The Storage layer is concerned with persistence of data and information and provides all the tools for data storing and efficient retrieving of Open Data. This layer is responsible for storing the files, structured data, tables and trees as well as the indexes and the metadata. Data can be stored directly in file system storage or in the structured data store. The low level description of the components is available in Deliverable D4.2 Alpha version of TET.

4.4 IMPLEMENTATION ARCHITECTURE

TET implementation architecture comprises of Frontend, Data Platform and TET Analytics. The frontend is used to handle interaction with the user and provides neat content management capabilities. Data platform handles data management related issues while TET Analytics is used to offer advance data analysis over a large volume of data. The implementation architecture has not changed significant since the alpha release, the only noticeable is the change in the frontend which now standalone python web application. The architecture is presented in Figure 6.

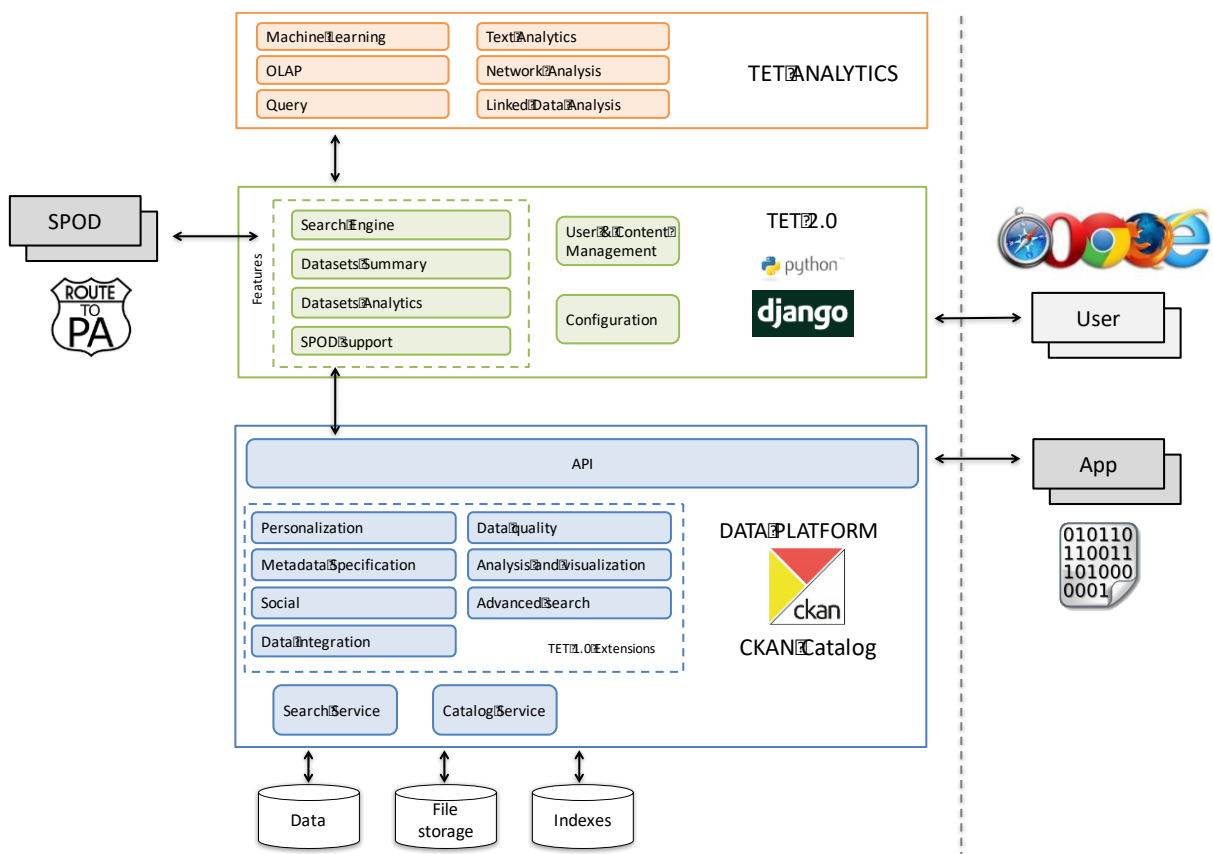


Figure 6 TET implemented architecture

Frontend

Beta release of TET is build as stand alone Django application and acts as a frontend interface for the TET platform. It consumes API from CKAN and TET Anyaltics in order to present views on the datasets for the users. The frontend layer provides a set of powerfull tools for discovering relatated datasets and fecilitates user in exploring and understanding datasets stored on the platform. The layer also handles interaction with SPOD other external tools. The fronted layer is multilingual and supports the following languages: english, italian, dutch and french.

Data Platform

CKAN is used as data management middleware for TET; it provides data catalogue services and supports searching. Additional extensions are added to CKAN to enable support for personalization, rich metadata specification, analysis and visualization, advance search over structured data, data integration, sharing, and assessing data quality. The layer uses a database, file storage and fulltext search index to offer a wide range of services over data to users and to Apps via API. Moreover, its role is to store, manage and serve all kinds of data to the user in a unified view. It covers user Interface, a catalogue of the datasets.

Future development plans includes e.g. support for multiple data point access and enrichment of the results.

TET Analytics

TET analytics platform offers advanced analytics services like Machine Learning, OLAP, Text Analytics, Network Analytics and Linked Data Analysis, etc. to the data platform. Existing analytics tools like R, Python and Spark, etc. will be used to build an intelligent Open Data platform that will enable users to make better use of the Open Data as key instrument for better transparency and better decision-making.

Extensions

Extensions provide a link between the User Interface and the Services. Is exposes all kind of analytics and visualization services available in TET. Moreover, APIs & widgets are providing a bridge between SPOD and TET to keep a unified state of the data between platforms: i.e. user status. Moreover, it enables functionalities of the two platforms in each other in a seamless way (i.e. link to start the discussion).

5 IMPLEMENTATION

This section describes the implementation of TET. It defines the details of the components included in TET, functional and implementation descriptions; the methodology followed as well as the technologies used during the development.

5.1 APPROACH

To effectively manage the software development process and to deliver high-quality software, we used agile development methodology. Agile methodology encourages continuous iteration of development and testing throughout the software development lifecycle [7]. This way the software evolves through communication between self-organizing teams with cross-functional expertise.

Agile methodologies have been proposed by the design science and software engineering communities as a set of flexible techniques that can adapt to the changing of user requirements [1]. The flexibility in software development projects is always important in order to increase the response to requirements changes (e.g., Byrd and Turner 2000, Duncan 1995, Gefen and Keil 1998, Lee and Xia 2005, MacCormack et al. 2001) - the positive relationship between flexible development architecture and software project performance as well as a positive relationship between team flexibility and end-user satisfaction with a system (Lee and Xia (2005). However, flexibility is not without cost and also tightly inimical to the degree of structure embedded in a development process [1]. It has been always the business customer responsibility for the success or failure of the product as a product owner. Hence, some agile technologies including SCRUM define this role by key functions, which include maintaining and prioritizing the product backlog, specifying and gathering individual user stories and their acceptance criteria [2].

Traditional SCRUM centres the product owner's interactions on two main artefacts which are a project plan (which relates the owner to the stakeholders) and a backlog (which relates the owner to the development team), where the product owner plays two roles in the development team but also deal with logistical challenges like customer bug reports for example [2].

SCRUM is defined as a framework within which people can address complex adaptive problems while productively and creatively delivering products of the highest possible value. Where, main features of any SCRUM is lightweight, simple to understand and difficult to master [3]. SCRUM started since the early 1990s to manage complex product developments. It deals not only with the process but also the framework where you can employ many techniques and processes. It binds events, roles and artefacts as well as managing the relationships and interaction between them [3]. SCRUM team contains the product owner, development team and SCRUM master. However Scrum events include the sprint, sprint planning, daily scrum, sprint review and finally sprint retrospective [3].

Requirements for TET were collected by analysing the project deliverables: D2.1, D2.3, D2.4, and D4.1 as well as the Description of Work (DoW), see Figure 7 below. The aforementioned deliverables are:

- D2.1: State-of-the-art Report and Evaluation of Existing Open Data Platforms
- D2.3: User stories on Open Data and Transparency
- D2.4: Requirement Specification and Use Case Models for TET and SPOD Subsystems
- D4.1: Alpha version of SPOD and D4.2: Alpha version of TET

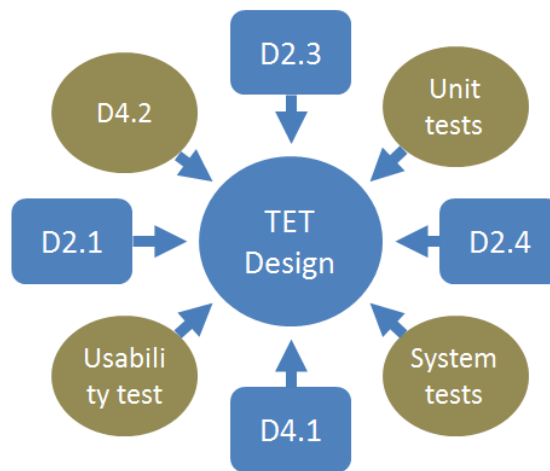


Figure 7: Requirements gathering

As feedback come from various system testing exercises (including unit tests – by TET team, Galway) conducted in various locations by various groups of stakeholders including the major usability workshop held in Galway, we incorporate them as updates to the initial functional and system quality requirements. Result details of these usability and technology quality evaluations are provided in section 7 while evidence from field work are exhibited in appendices Appendix 1A: Pre-usability evaluation testing – 2 Samples of manual with user comments, through to Appendix 4: Github Issues List.

We adopted SCRUM to manage product development. Tasks are divided into sprints and useable software with incremental changes is delivered at the end each sprint. Planning meetings at the start of each sprint are organized to define clearly the objectives of each sprint and to plan development tasks. To track the project progress; daily short stand-up meeting are held. A retrospective session is conducted at the end of each sprint to learn from the sprint experience and use it for improving future work. Initially, we started with long sprints but gradually will use short sprints. The software is extensively tested before each release. Feedback from the pilots is used to prioritize development tasks, bug fixes and enhance implemented features.

5.2 IMPLEMENTATION TECHNOLOGY

TET is set of tools to build on top popular open data platform CKAN. The core advantage of TET beta is the new user centric user interface for CKAN which is implemented using Django Framework². Django is free open source widely used python based development framework. Additional scientific and visualizations libraries such pandas³ and matplotlib⁴ etc are used for analysis and visualizations. CKAN is a de facto standard for open data portals; it provides comprehensive data management services. Comprehensive Knowledge Archive Network (CKAN) is the free open source web-based data management solution. It's written in Python and uses Pylons framework. It uses Postgres as database and apache solar for the index to fulltext search. CKAN facilitates easy discovery of data and has a broad user base. Governments around the world use it for serving open data, which includes US, UK, Ireland and Australia. CKAN provides provision for theming and extensions. CKAN provides many essential features required to publish, share and visualize datasets and offers powerful cataloguing, searching and storing capabilities [5]. CKAN software development began in March 2006 and first release in July 2007. CKAN project is implemented and managed by Open Knowledge Foundation (OKF) [6]. The main objectives of CKAN can be summarized by:

- 1) Freedom of access, creation and dissemination of knowledge

² <https://www.djangoproject.com/>

³ <http://pandas.pydata.org/>

⁴ <http://matplotlib.org/>

- 2) Develop, support and promote projects, communities and tools that foster and facilitate the creation, access to and dissemination of knowledge.
- 3) To campaign against restrictions both legal and non-legal on the creation, access to and dissemination of knowledge.
- 4) Act as an intermediary between funding institutions, the creation and diffusion of knowledge projects [6]

TET exploits API and extension capabilities offered by CKAN to deliver rich user experience, and powerful data discovery and analysis features to help users in leveraging the real potential of open data. TET reuses existing free opensource tools and libraries wherever possible. The majority of TET code is written in Python⁵ and JavaScript⁶ programming languages. Most of TET functionality relay on API provided by CKAN, however, external libraries and API are used where there is no native support available in CKAN. In some cases where there is an existing solution already available, we have tried to reuse it.

5.3 IMPLEMENTATION DETAILS

The core implementation work is done using the Python programming language. Python is high level programming language that offer syntax that is ideal for data processing, therefore it was an ideal choice for the our development, in additional it offers several rich set of data analysis, machine learning and visualization libraries that are helpful in development analytics and visualization features for TET. Django is used as core development framework; it is powerful MVC framework with all essential features for developing robust web application. It uses Jinja⁷ as template engine. JavaScript and associated tools are widely used to handle client side interactions and visualization. Additional JavaScript libraries for visualization and handling AJAX are used to build client end of application. Bootstrap⁸ is used for organizing the user interface elements. Themes are designed to match common look and feel with SPOD. Additional plugin developed previously are improved wherever required, all the plugin were reused in beta release TET.

5.4 IMPLEMENTED FEATURES

TET beta offers significant improvement from previous Alpha release. The user experience has redesign and considerably improved, however users are provided option to switch back to classic experience introduced in alpha release. Search and exploration is simplified and enhanced. Easy to use analytics features are added to guides users in understanding the data associated with datasets without having deep technical knowhow of data analysis.

5.4.1 USER INTERFACE

The user interface is redesign to meet the concerns of ordinary citizen, inspiration were taken from popular search portals to come up with a new user centric design. Interaction between TET and SPOD is made seamless as possible. All the previous implemented features are preserved and are incorporated in the new user interface and users are provided option to switch between new and old user interface. All unnecessary user interface element are removed which were causing distractions for the users. The new user interface enables users to focus on core purpose of open data platform of search and discovery. Tooltips and other clues are provided to users to assist them exploring TET.

⁵ <https://www.python.org/>

⁶ <https://developer.mozilla.org/en-US/docs/Web/JavaScript>

⁷ <http://jinja.pocoo.org/>

⁸ <http://getbootstrap.com/>

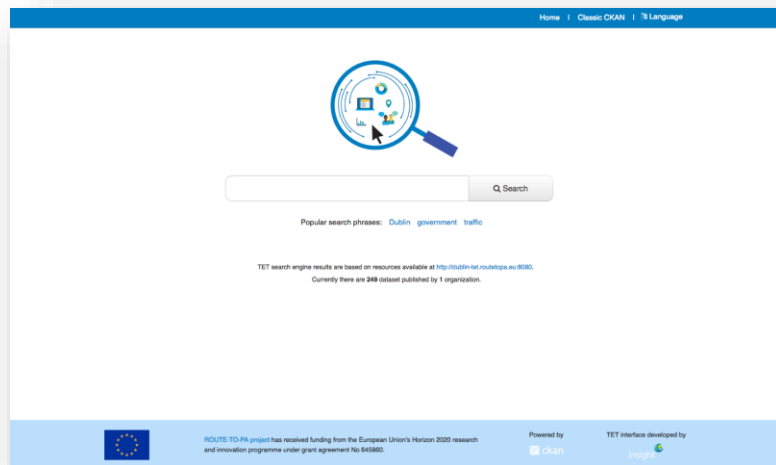


Figure 8: User interface - Homepage of TET

5.4.2 IMPROVED SEARCH EXPERIENCE

Search experience is enhanced to make the process of discovering relevant dataset as easy as possible. Autosuggestion feature is added to guide user to make correct and relevant queries. Users are automatically suggested once they enter a term related to category or role. Search feature is present throughout the interface, to enable users to flexibly explore datasets related to their query. The landing page only highlights the search feature with any distraction.

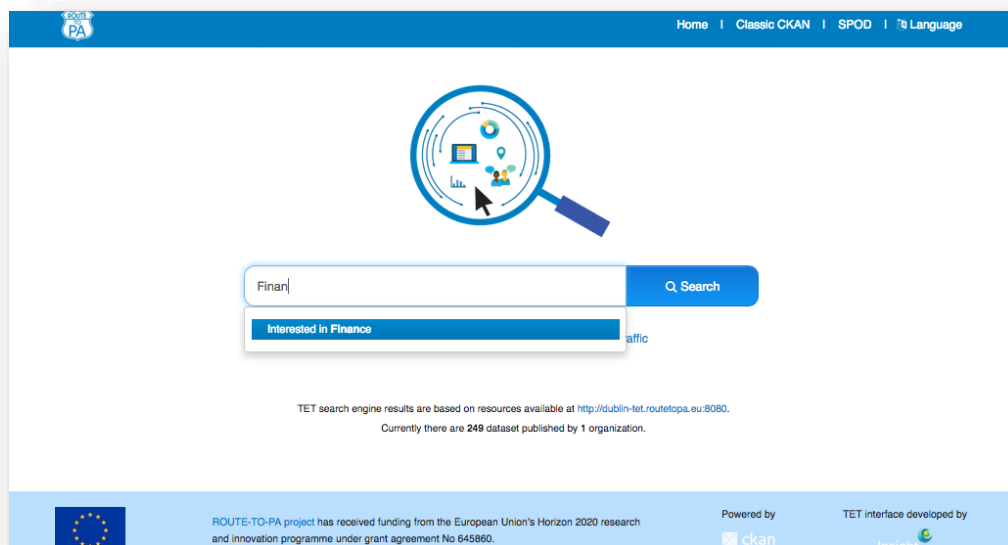


Figure 9: Search experience improved with automatic keyword suggestion

5.4.3 SEARCH RESULTS FILTERING

More option for filtering and ordering search results are added to let users to get more precise results. User can filter results by date, location, theme, file formats etc. and sort using different options. After search all the facts are updated with counts. Users are provided with number of results and results can be sorted by relevance, name and date. Result page provide name, description, and file formats available for each dataset. Datasets can be viewed, analysed and visualized with single click from the search results page.

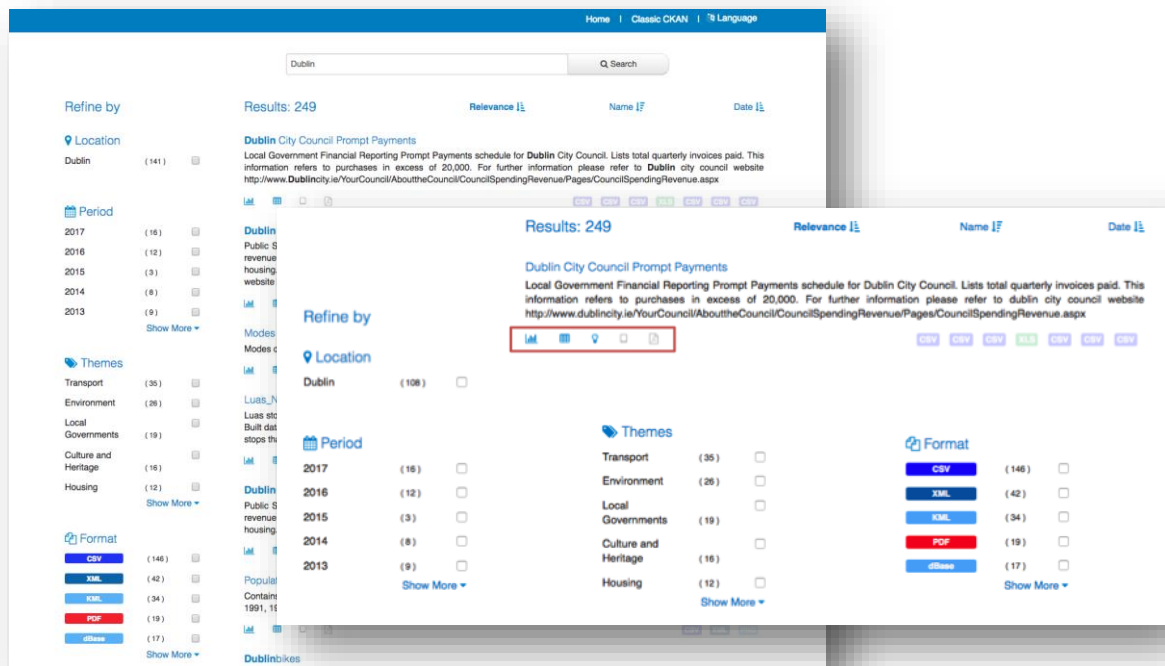






Figure 10: Screenshots showing search result filtering tools

5.4.4 DATASET VIEW

The dataset preview is enriched with new option that helps users in better understanding the dataset and related files in single glance. Charts are automatically generated from associated data and displayed to the user on the dataset page. Metadata is provided in easy to understand text rather than table. A dataset can be shared directly from the dataset page on popular social media outlets with a single click. Discussion from SPOD can be viewed directly from the dataset view. Quality indicators, tabular structure, dataset recommendations, tags and files associated with datasets can be viewed on the dataset page in different panels. Data summary and tabular views can be invoked with a single click from the dataset view.





View as:    

Description

Local Government Financial Reporting: Prompt Payments schedule for Dublin City Council. Lists total quarterly invoices paid. This information refers to purchases in excess of 20,000. For further information please refer to dublin city council website <http://www.dublincity.ie/YourCouncil/AbouttheCouncil/CouncilSpendingRevenue/Pages/CouncilSpendingRevenue.aspx>

Share this page

Use the icons below in order to share this page in your favourite Social Media.

Metadata

This Dataset was created at **14 June% 2016, 16:38** and last modified at **21 September% 2016, 12:34**.

The data was published by **DublinCity**.

If you need more details, maintainer can be contacted at: **Not supplied**.



The data is available in 7 formats: **CSV CSV CSV XLS CSV CSV CSV**

Discussions at SPOD

This dataset has not been discussed in SPOD, be the first to discuss the dataset!

[Start the discussion](#)

Dataset Quality

Indicator	Value	
Metadata	Fair	<div><div></div></div>
Content	Very good	<div><div></div></div>
License	 Not provided	
Version	 1.0	
Last updated	138 days ago	

Tags

[Government and Participation](#) [OpenData](#) [Local Government](#) [Citizen Participation](#)

[Transparency](#) [Local government spending](#) [public finance](#)


Table Structure


Field name	Type
<code>id</code>	Integer numbers, e.g. 42, 7
<code>Vendor Name</code>	Arbitrary text data, e.g. Some text
<code>Item Description</code>	Arbitrary text data, e.g. Some text
<code>Price Paid</code>	Numbers, e.g. 1, 2.4, 4.7


Related Datasets


- Litter Fines 2011
- Litter Warden Inspections
- Log of Litter fines issued 2003-2012
- Property Price Register
- Annual Budget 2013
- Councillors
- Councillors Allowances and Expenses
- Councillors Conferences
- Councillors Representative Payments
- Dublin City Council Spending & Revenue Budget for 2014 (CSV)
- DLR Councillor Contact Details
- Dublin City Council Annual Financial Statements
- Dublin City Council EU IMF Financial Reporting
- Dublin City Council Prompt Payments
- Dublin City Council Spending and Revenue budget
- Fingert County Council Budget 2016
- Freedom of Information Requests
- Fire Brigade and Ambulance Call Outs


Files


CSV  [DOC_PromptPayments2013Q1R2.csv](#)


CSV  [DOC_PromptPayments2013Q1R3.csv](#)

CSV  [DOC_PromptPayments2014Q1R1.csv](#)

XLS  [DOC_PromptPayments2013Q1R4.xls](#)

CSV  [DOC_PromptPayments2013Q1R1.csv](#)

CSV  [DOC_PromptPayments2013Q1R2.csv](#)

CSV  [DOC_PromptPayments2013Q1R3.csv](#)

3

Figure 11: Screenshot of a selected dataset showing various boxes of information on the dataset

5.4.5 TABULAR VIEW

The tabular preview present similar options as in the previous version, feature such as data explore and PivotTable are reused from previous implementation. New SQL query console is added for more technical users. Description, Metadata and SPOD discussion panel are added to list details related to the dataset. Tabular view is an effective way to view, analyse and visualize actual data in detail.

Dublin City Council Prompt Payments Tabular View

Data Explorer
PivotTable
SQL Query

Grid
Graph
Map
338 records
1
100

Search data
Go
Filters

ID	Vendor	Item Co.	Paid A...
1	HANLE	CSV S E	11077.0
2	BKMF CO	CCN207	480.000
3	DUEGA	CCN207	411280
4	BKMF CO	CCN207	388.000
5	GAASCO	CCN207	216.687
6	GAASCO	CCN207	212.885
7	DUEGA	CCN207	291729
8	GAASCO	CCN207	202709
9	TOWNL	CCN207	79486.3
10	KELLY	CCN207	81880
11	ISSB NE	ISSB SU	28283.2

API Link
http://dublin-bf.routetopa.eu:8080/api/action/datasetone_search?resource_id=bf101a2e-f101-603e-bd

View as:
Bar chart
Lightbulb
Table
Document

Description

Local Government Financial Reporting Prompt Payments schedule for Dublin City Council. Lists total quarterly invoices paid. This information refers to purchase s in excess of 20,000. For further information please refer to dublin city council website <http://www.dublincity.ie/YourCouncil/RouteCouncil/CouncilSpendingRevenuePages/CouncilSpendingRevenue.aspx>.

Discussions at SPOD

This dataset has not been discussed in SPOD, be the first to discuss the dataset!

[Start the discussion](#)

Metadata

This Dataset was created at 16 June% 2016, 14:38 and last modified at 01 September% 2016, 12:34.

The data was published by **Dublinlink**.

If you need more details, maintainer can be contacted at: **Not supplied**.

The data is available in 7 formats: **CSV CSV CSV XLS CSV CSV CSV**

Figure 12: Tabular view of a selected dataset

5.4.6 DATASET SUMMARY

Dataset summary feature provides users descriptive statistics related to the data associated with dataset. Summary panel provides maximum, minimum and average for each field in the dataset as textual description. The details panel lists mean, median, mode and quartiles for each variable in the datasets; the box plot provides the same information in visual way. Correlation between fields is represented as correlation matrix plot. Together all panels on the dataset summary page help users to understand the data in details and to make useful inferences.

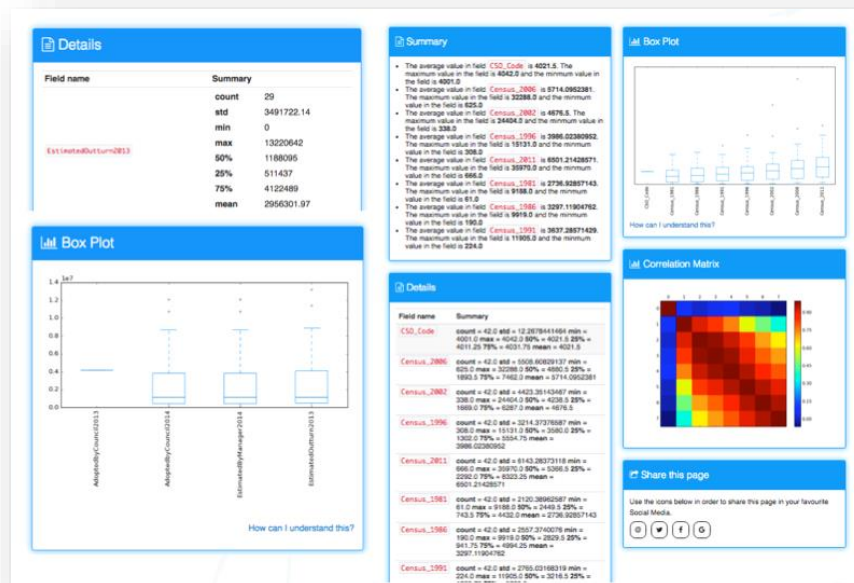


Figure 13: Summary View screen offers further information on a dataset

5.4.7 AUTOGENERATED CHARTS AND ANALYTICS

Datasets are analysed and useful visualizations from datasets are generated automatically, which are helpful, in understanding and interpreting datasets quickly. Generated visualisation scrolls on the screen highlighting different characteristics of the dataset. This feature makes use of the SPOD charting feature to render the charts. All the charts are labelled appropriately to communicate the intent of each chart to the user.

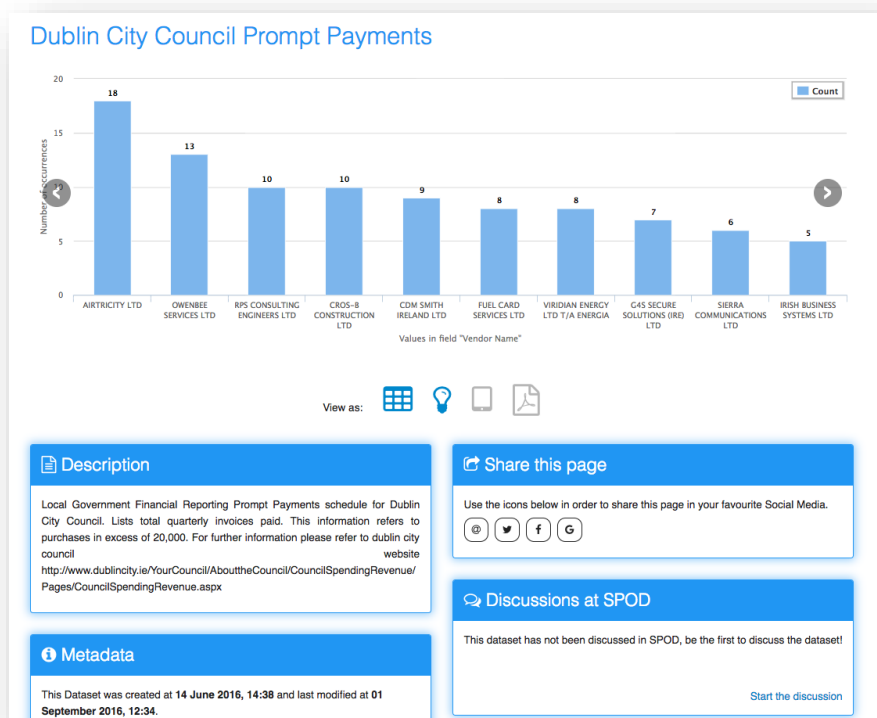


Figure 14: Autogenerated chart on a dataset

5.4.8 DATASET QUALITY METRICS

The dataset quality panel displays set of measurements that help users in gauging quality of data associated with the dataset. Measurements are displayed both as text and as visual indicators. These measurements provide details related to actual resources associated with datasets as well as the metadata. Quality indicators are key to ascertain the quality dataset before utilizing it for analysis and visualisation purposes.

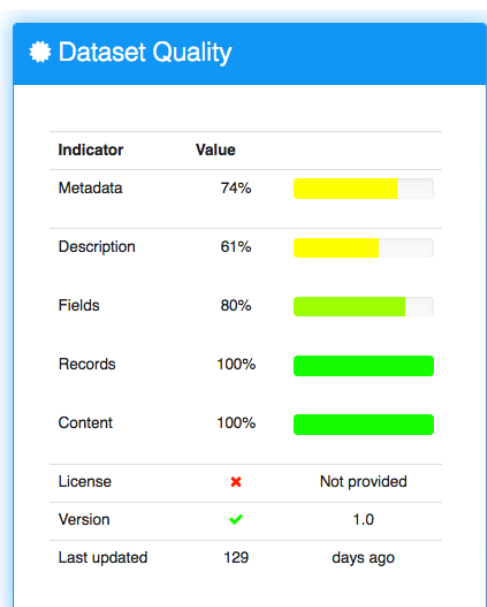


Figure 15: Dataset quality metrics from the metadata

5.4.9 USER READABLE METADATA

Metadata is converted into readable text to allow ordinary citizens to understand metadata in natural text. In the previous version metadata information was presented in tabular format and the user had difficulty in understanding the terminology used. To fix this issue, the values are converted into a readable textual description, which is easy to understand.

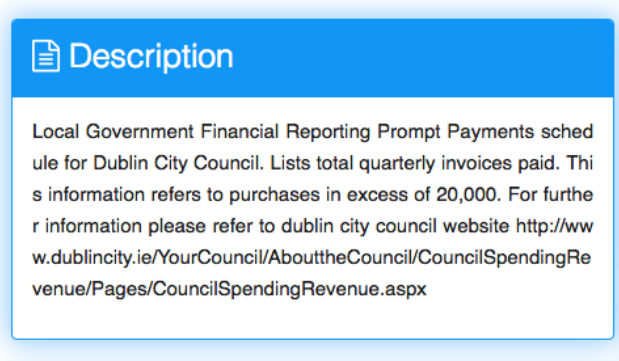


Figure 16: Metadata information in natural language

5.4.10 SQL QUERING

An SQL query console is added to allow power users to query datasets for advance data analysis tasks. The console provides colour-coded editor for entering SQL queries and highlights any errors. Queries can be executed against multiple datasets resources. Query can contain many useful functions provided by the SQL language. SQL console can be used to combine, filter, aggregate and group data using standard SQL syntax.

Dublin City Council Prompt Payments Tabular View

Data Explorer PivotTable SQL Query

1 SELECT * From "8e761a2a-f16f-4c9e-8d4e-2ebc42f56b4c" LIMIT 10

Run

Vendor Name	Item Description	Paid Amt
HANLEY PEPPER LTD	CIVIL ENGINEER PROFESSIONAL SERVICES	17077.090000000000014551915228366851806640625
BAM CIVIL LTD.	CONSTRUCTION OF HOUSING	485000
DUGGAN BROS LTD	CONSTRUCTION OF HOUSING	417280.41999999999837018549442291259765625
BAM CIVIL LTD.	CONSTRUCTION OF HOUSING	388000
GANSON BUILDING & CIVIL ENG CONTRACTORS LTD	CONSTRUCTION OF HOUSING	276487.04999999999883584678173065185546875
GANSON BUILDING & CIVIL ENG CONTRACTORS LTD	CONSTRUCTION OF HOUSING	272865.219999999997206032276153564453125
DUGGAN BROS LTD	CONSTRUCTION OF HOUSING	257726.92999999999930150806903839111328125
GANSON BUILDING & CIVIL ENG CONTRACTORS LTD	CONSTRUCTION OF HOUSING	202139.89999999999417923390865325927734375
TOWNLINK CONSTRUCTION LTD	CONSTRUCTION OF HOUSING	78466.3800000000004656612873077392578125
KELLY & COGAN ARCHITECTS CONSUL	CONSULTANT PROPERTY	81885

Figure 17: A screenshot from SQL query on a dataset

5.4.11 RELATED DATASETS RECOMMENADTION

There are two versions of dataset recommendation. The first version is imported from the previous version uses tags similarity to recommend new datasets. The new smart algorithm is based on sophisticated Self Organizing Maps (SOM) that intelligently recommend related datasets to the users by identifying unique relationships in the data. TET can be configured to use any of these two versions, in both implementations related datasets are automatically recommended to the users when the user navigates to the dataset page. Users can adjust relatedness by adjusting the scrollbar provided in the case of smart recommendation. The feature can be used to discover related datasets. The feature will be enhanced to let users to combine recommended datasets together for analysis and visualization.

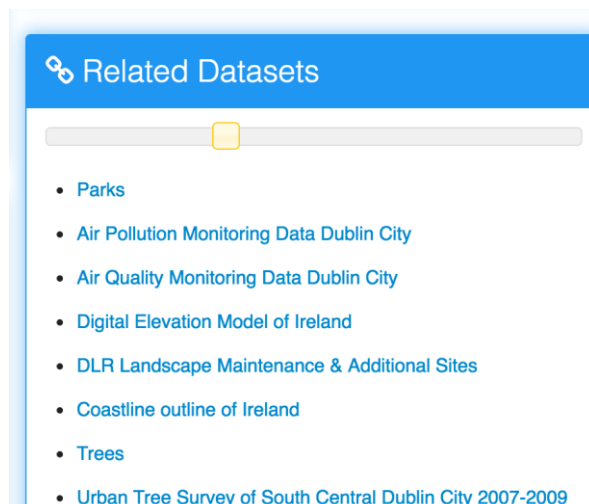


Figure 18: System recommendation of related dataset with a slider to adjust relatedness of items on list

5.4.12 ANALYSIS AND INTEGRATION OF RELATED DATASETS

Compatible resources both within and outside the datasets can be automatically combined, viewed, analysed and visualized together. Users can select datasets from the results page to identify related datasets. The merge dataset feature recommended files that can combine from the selected dataset, and indicate to the user where

it isn't possible. Users can select combination of datasets for one of the compatible dataset clusters and can view, chart, analyse or download the dataset to explore it using a third-party tool.

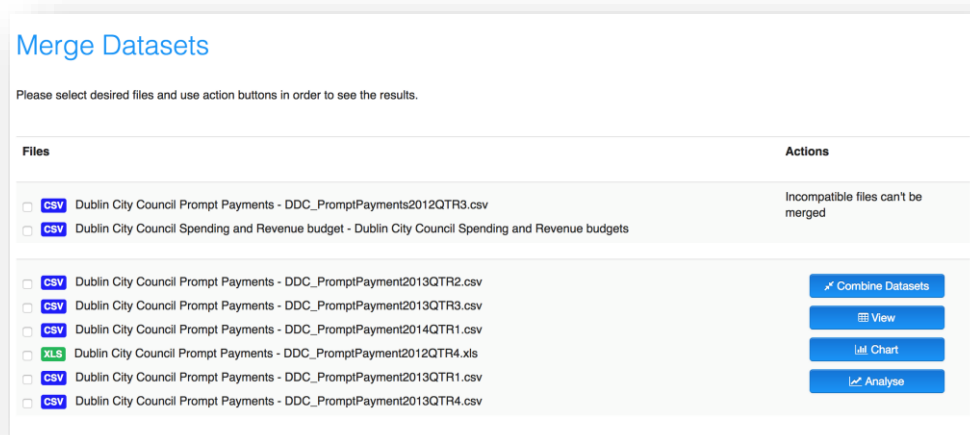


Figure 19: Screenshot showing compatible datasets that can be merged and analysed together & incompatible dataset that cannot be merged

5.4.13 LANGUAGE SUPPORT

Language support in TET alpha has been improved with build-in features provided by Django Framework. TET supports English, Italian, Dutch and French languages and new languages can be easily incorporated.

5.5 INTEGRATION BETWEEN TET AND SPOD

Integration between TET and SPOD has significantly improved in beta release of both platforms. Interaction between TET and SPOD has been made as seamless as possible. In order to enable seamless user experience across TET and SPOD, a unified user interface was designed and implemented. Both TET and SPOD have been specifically designed within the ROUTE-TO-PA project, and they are based and compliant with the material design⁹. Both platforms share common theme / styling and shared authentication mechanism. TET uses SPOD charting tools for displaying auto-generated charts. SPOD discussion can be started and viewed on the datasets. PivotTable results can be exported to SPOD for visualization and discussion. An API link can be easily copied from TET to SPOD or any other application with a single click.

Roadmap for future development includes series of features that will enable close integration between TET and SPOD.

⁹ <https://design.google.com/resources/>



Figure 20: Screenshot showing user discussions on a dataset imported from SPOD

6 DEPLOYMENT

This section defines the system requirements for TET, its deployment process and discusses the issues that may be encountered during the deployment process.

6.1 DEPLOYMENT REQUIREMENTS

TET implementation is platform independent and can be deployed in a variety of environments. The preferred environment for TET is Ubuntu Linux 12.04 64bit which has been used for development, deployment and testing of the TET alpha version. TET user interface for CKAN requires CKAN instance to be installed on the same or on a remote machine. CKAN requires a machine with python, java virtual machine, apache solr and instance of Postgres or MySQL. Details related CKAN installation can be found on CKAN website¹⁰. Apart from standard CKAN installation TET requires DataStore and some TET extension plugin to be installed on CKAN details related to this can be found in Table 9.

Table 9: Requirements for CKAN with TET extensions

Software Requirements	More information
<ul style="list-style-type: none">• Python 2.6.x/2.7x• CKAN• Apache Solr• Postgres 9.3 or• MySQL Server• DataStore + DataPusher• Extensions<ul style="list-style-type: none">○ RTPA extension○ PivotTable○ OpenID○ RTPA theme○ DataStore	https://www.python.org/doc/ http://ckan.org/ http://lucene.apache.org/solr/ http://www.postgresql.org/docs/ , http://dev.mysql.com/doc/ http://docs.ckan.org/en/latest/maintaining/datastore.html https://github.com/routetopa/ckanext-routetopa https://github.com/routetopa/ckanext-pivottable https://github.com/routetopa/ckanext-openid https://github.com/routetopa/ckanext-ckan-theme http://docs.ckan.org/en/latest/maintaining/datastore.html

TET and CKAN can be installed on a single machine or TET can be configured to use an existing CKAN instance. Additional extension and DataStore need to be enabled in order to use the TET interface. The required capacity of the system varies and depends on the expected workload that is required to be handled by the server; at a minimum, a server should have 2 CPU cores, 4GB RAM and 60 GB of disk space. The medium size workload should be sufficient for the city level portal while a large instance should be sufficient for a national portal. Table 10 provides recommendations for hardware requirements.

Table 10: Hardware Requirements

Workload	Recommended hardware
Small	2 CPU cores, 4GB RAM and 80 GB of disk space
Medium	2 Servers with 4GB RAM and 80 GB of disk space
Large	2 Servers with 8GB RAM and 160 GB of disk space

¹⁰ <http://docs.ckan.org/en/latest/maintaining/installing/install-from-source.html>

For better performance and security, the database and solr for CKAN installation should be deployed on a separate server.

6.2 DEPLOYMENT PROCESS

The deployment process involves setting up the server, installing the required software and relevant TET components and configuring the instance. TET can be either used as a fully integrated open data solution, or some components can be adopted to enhance the capability of the existing portal. Detailed documentation for installation is provided with each release and can be also found in the code repository. Each repository contains a detailed readme file, which explains in detail the installation and configuration options available. Table 11 lists and corresponding links to the code repository of the projects that are produced as part of the TET alpha release.

Table 11: Links to the repositories

Project Name	Project Type	Description	Link to repository
TET CKAN Extension	CKAN Extension	Extends CKAN capabilities with rich metadata specification, personalized search and recommendations.	https://github.com/routetopa/ckanext-routetopa
TET	Standalone Application	TET User Interface for CKAN	https://github.com/routetopa/tet
RTPA CKAN Theme	CKAN Theme	RTPA theme for CKAN	https://gitlab.insight-centre.org/egov/ckanext-rtpa_theme
PivotTable Extension	CKAN extension	Provide easy analysis and visualization over tabular data using PivotTable	https://github.com/routetopa/ckanext-pivottable
CKAN OpenID Extension	CKAN Extension	Provides authentication with OpenID	https://github.com/routetopa/ckanext-openid
RTPA-devenv	Vagrant Project	Setups the development environment for TET	https://github.com/routetopa/tet-devenv

For full installation, there should be an existing CKAN installation with DataStore enabled. All the TET plugins should be installed on a CKAN instance, most TET extensions can be installed as any other standard CKAN extension, detailed instruction for each plugin are provided in a respective repository. TET front application is Django based application and preferably should be installed in a virtual environment. The instructions for installation are provided in the respective repository, the dependencies can be installed using the requirements file provided.

TET development environment is managed with Vagrant¹¹. RTPA-devenv contains the Vagrant configuration and associated data file, which are required to configure the development environment. Once the development is finished, the application will move to a staging area where new features are tested before moving on to the production environment as illustrated in Figure 21: TET deployment process.

¹¹ <https://www.vagrantup.com/>

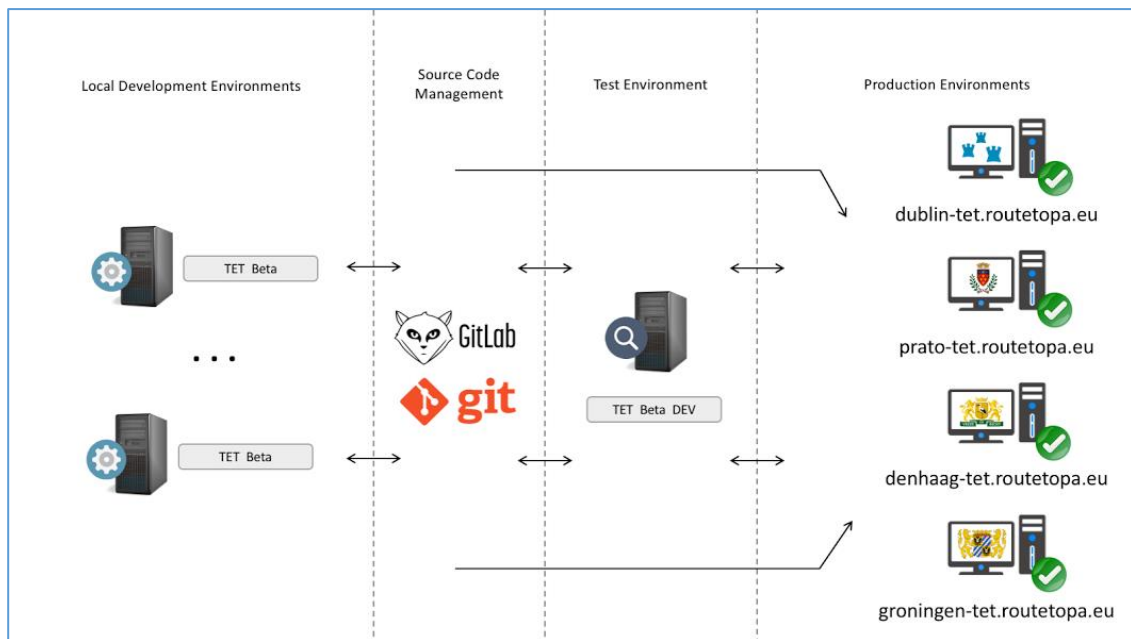


Figure 21: TET deployment process

6.3 DEPLOYMENT ISSUES

Migration, software versions and metadata schema mismatch are common issues that can be encountered during the deployment process. Migration data could be tricky and depends on the source platform. Most of the open data platforms have implemented the harvesting feature that enables transfer of the data from one platform to another. In some cases, custom scripts development may be required in order to transfer the data.

TET has been built on CKAN in version 2.6, and some features may not be compatible with its previous releases. Therefore, in some cases, a CKAN version needs to be upgraded in order to maintain compatibility with developed extensions.

Each organization has their own way of describing metadata associated with the data. TET defines a metadata schema that is compliant with the W3C recommendations for data on the web. Changes may be required in order to adopt the TET schema ideally. However, it is also possible to configure the TET schema in such a way, that it could be easily adopted with existing metadata schema.

Categories and roles are essential metadata fields for enabling personalized search and recommendations. These fields could be missing, or list of categories and roles may be different than provided. While the categories issue can be easily fixed by changing the settings in the configuration file, the roles issue would either require manual data entry or the development of a custom script.

7 TESTING AND EVALUATION

7.1 VERIFICATION AND VALIDATION CONCEPTS

In deliverable D4.2, this section is entitled 'Verification and Validation' and explains that software product development generally combines two groups of tasks – the **Verification** processes and the **Validation** processes. While the verification processes describe the activities of consultation with users, gathering requirements and analysing them to ascertain alliance of a system with users' expectation, usability and performance; validation processes involve checking and testing the functionalities to ascertain correctness of the system with its specifications derived during verification phase (Tutorial points, online). Using *Figure 22* which shows two-legged transition processes involved in converting user needs into the requirement specifications and validation of the requirement as well as the system capability to address the needs, we place our system development processes as follows Table 12:

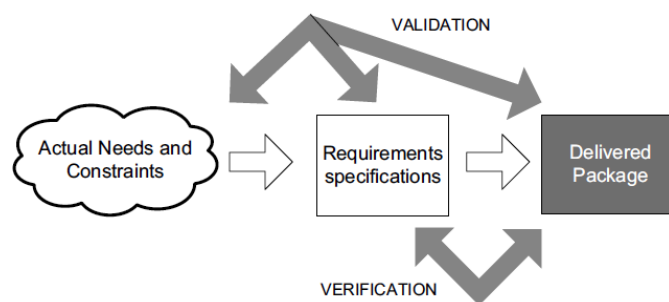


Figure 22: Validation and Verification Processes in Software Development (Baresi, L & Pezze, M. 2006)

Table 12: Development Processes in Categories

Processes	Verification or Validation processes
Requirement elicitation, analysis and confirmation with stakeholders to ensure uniform understanding by all parties	Verification
Unit testing tasks done within our TET team at Insight (NUIG) along with implementation of individual components, providing results on Github as testing took place to serve as feedback for developers to upgrade system. (Unit testing)	Verification
<div> <div> RTPA: Some Issues from Usability Evaluation #20 opened 3 hours ago by eosagie </div> <div> TET Beta Issues #19 opened 4 hours ago by eosagie </div> <div> TET on Internet Explorer #4 opened on Nov 3, 2016 by eosagie </div> <div> TET Review on IE environment #3 opened on Nov 3, 2016 by eosagie </div> </div>	
Pre-usability evaluation testing exercises performed by recruited testers – some technically advanced and others not. (System testing)	Verification / Validation
Usability Evaluation performed by recruited evaluators – including ordinary citizens, Public administrators, researchers, students, etc. (Usability testing)	Validation

The processes of verification and validation took place simultaneously and it applied the principles of the V-Model which we adopted for this project as explained in deliverable D4.2. The iterative processes of build, test and then implementation of test findings as system upgrade also as explained in details in deliverable D4.2 are put into practice as in the manner of: build and first test by Insight (NUIG) TET team – adjust system with feedback – 2nd time retest with recruited team – adjust system with feedback – 3rd time retest system for usability by external evaluators – adjust system for optimal performance. For details of the explanation of the verification and validation, the V-model and related iterative processes, please refer to deliverable D4.2. Figure 23 shows that

verification is not directly concerned with how the solution addresses the client's problems because it ensures only the correctness or alliance of built components with the collected and confirmed requirements. On the other hand, validation tests the solution to see how best it addresses the initial problems of the users or clients (A). The interconnections between verification and validation activities are shown in (B) side of the figure.

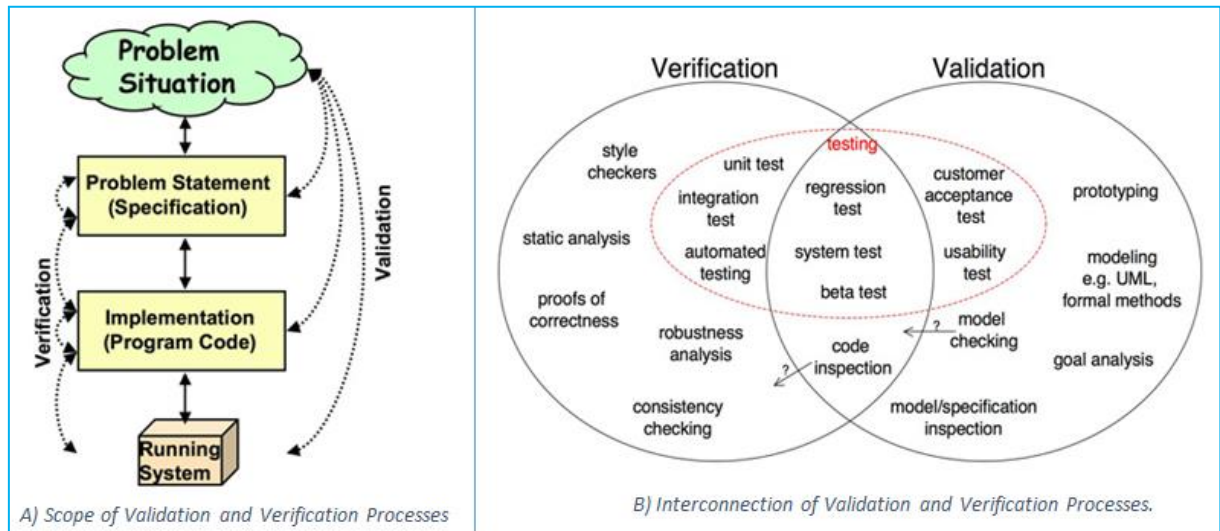


Figure 23: A) V&V Scope, B) V&V Interconnections. (Easterbrook, S. (2010))

7.2 FEEDBACK FROM TESTING AND EVALUATION PROCESSES

7.2.1 SUMMARY OF PRE-USABILITY EVALUATION TESTING

As stated in subsection 3.2.1, there were a total of 11 testers of males and females with some playing the roles of an open data publishers while others played the roles of open data platform user. Due to the division of roles, we compiled two instruction manuals for use by these testers; however, the introduction sections of the manual are the same and thus have no need to repeat them in the appendix that exhibits samples (see Appendix 1A: Pre-usability evaluation testing – 2 Samples of manual)

Check Metadata Quality and Metadata Completeness Rating

On the tools for checking the quality of metadata and metadata completeness rating, 50% of evaluators believe that the feature is good enough and would improve the accessibility and usability of datasets. However, one evaluator reported on the need to be consistent in use of terminology with existing platforms such as DublinCore and should be aligned with it. Help prompts such as dialogue boxes, improve naming of file and text to texts that describes buttons and terms are needed on the platform. If metadata provided on the platform correctly describes available datasets, evaluators (63%) believe that the tools will improve searching relevant datasets on the portal. Nevertheless, that would be based on the search word used. If you search with category search word, you may end up with a broad result, some of which might not be very relevant to your need. By compelling data suppliers to complete a metadata form for all datasets they supply, 75% of evaluators agree that open data may become more available and accessible to potential consumers. Evaluators suggested the need for a date field that could be auto-completed for dynamic data being uploaded on the site to show also the frequency of data update. Manner of implementation of the compulsory metadata form should not become a hindrance for data suppliers;

the functionality does add some important dimension such as more complete metadata and quick decision on relevance of datasets and help for searching the existing platform.

Personalized Search (User with Admin Login) and Personalized Search (User is logged in)

2/8 of evaluators believed that Personalized Search (User with Admin Login) is user-friendly while 3/8 believes it only reasonably user-friendly. Suggestions include more roles which should be listed alphabetically and navigation should be improved. For the Profile-based Personalization (user, not logged in), 1/3 of evaluators believe the implementation is OK whereas others commented that there are too many role options to search through. For the personalised search – user logged in, all evaluators believe the function worked well for them and they got accurate result from their searches.

Add Dataset (Admin login) – Adding dataset to the portal is rate easy (2/8), difficult (2/8) and not so difficult chosen by another 2/8 of evaluators. This means evaluators (33%) are a little tilted towards the easy side of the task of adding a dataset. Other comments include: terminology should be made clearer, internal server error, a few bugs to fix and improvement to make it more intuitive. For some evaluators, adding a dataset wasn't a difficult task, but there is a need for more explanation as in tooltips to help users, improve intuitiveness, add more information on the markdown formatting on the spatial notes as well as refine the spatial coverage

Analyse and visualise a dataset – to decide its quality and meaning – one of the evaluators was unable to say if visualising a dataset could reveal its quality. However, the functionality of the tools to visualise the dataset is reported to be good enough in visualising the dataset and making decision on the type of graphical representation is easy. Learnability, performance and satisfaction is rated on the positive for this functionality. However, not many evaluators had the opportunity to evaluate this tool in the workshop.

Linking/Unlinking a dataset – Evaluators (data publishers) believe this is a relevant to dataset publicity functionality and useful and can be used to target other datasets and can be used to enhance the recommendation dataset system. However, a user may need to have knowledge of the data structures to know where to go find the linked data types and “what happens if a user keys in a non-existent dataset name?” queried one of the evaluators. In general, satisfaction on the manner of implementation is rated above average by evaluators. On the second task of unlinking a dataset, evaluators believe this undo aspect will enable users update dataset linking. However, there is need to keep the operation simple. Evaluators suggested that users should be able to receive a notification indicating a previously linked dataset is no longer existing and needs to be unlinked. Lastly, one evaluator in a different environment was unable to find the dataset linking “Manage button”.

7.2.2 USABILITY EVALUATION RESULT

The usability exercise was conducted in the form of a workshop in which we had a cross-section of stakeholders representing different sectors of interest in ROUTE-TO-PA system. A breakdown of the stakeholder composition is provided in subsection 3.2.3. During the workshop, users were shown a demonstration of the web-based ROUTE-TO-PA system to get familiarised with the system after which they are handed over the user manual to be used in executing the Dublin Traffic problem scenario described in usability manual (see

Appendix 2: Usability Evaluation – Instruction Manual). Working on the system, users were guided through the processes of problem identification, discussions on issues arising or matters of interest such as the suggested moped bikes and finally attempting the co-creation of a solution to the identified Dublin City traffic congestion problem. As they worked with ROUTE-TO-PA tools, users were asked to note down their experiences with the system functionalities and the perceptions they felt. At the end of the workshop, users were asked to complete the Google survey containing the usability statements or questionnaires (Table 13). This gave them the opportunity to rate their levels of satisfaction on the various aspects of the system they used.

Table 13: ROUTE-TO-PA Tech Evaluation Survey – System Features and Usability statements

Functions	Features	Usability Statements (used for survey)
Search for datasets	1) Relevance (Keyword search)	The result list of datasets returned is relevant to the search keywords entered
	2) Accuracy (search result refinement)	The filters options enhance the data search accuracy
	3) Clarity	The presentation of search results is clear and easy to read
Data resource views	4) Simplicity (Data charts and tables)	The data tables and charts are simple and easy to read
	5) Understandability (charts and tables)	The dataset presented as data tables are understandable
	6) Structure of pdf output (charts/tables)	The datasets presented as charts are understandable.
Data resource descript.	8) Understandability	I understand the description of the dataset.
	9) Informativeness	Data description provides sufficient information about the content and meaning of the dataset I viewed.
Social discussions on data resource	10) Relevance	I find the discussions from SPOD on my dataset relevant to my need about the dataset
	11) Informativeness	The discussions provide some sense about the content of the dataset
	12) Understandability	The discussions about data on SPOD are clear and understandable to me
Private space	13) Simplicity	The tools to create and store data artifacts in 'My space' are simple
	14) Familiarity	I am familiar with most of the buttons in 'My space'
	15) Self-descriptiveness	Tools, buttons and functions in 'My space' are self-descriptive
	16) Minimal action	I think the number of steps to create a chart while in 'My space' are few enough
Public space	17) Simplicity	I think using the tools to achieve discussions in 'Agora' is simple
	18) Familiarity	I am familiar with the majority of the buttons in 'Agora' and their uses too
	19) Self-descriptiveness	Tools, buttons and functions in 'Agora' are self-descriptive
	20) Minimal action	Creating a chart while in 'Agora' takes few steps
Collaboration space	21) Simplicity	Using the tools to achieve the goals in 'Cocreation room' is simple
	22) Familiarity	I am familiar with the majority of the buttons in 'Cocreation' room and their uses
	23) Self-descriptiveness	Tools, buttons and functions in 'Cocreation' room are self-descriptive
	24) Minimal action	Creation of a chart while in 'Cocreation' room takes few steps

Analysis of Results

Table 15 assembles the results of the usability survey conducted after the workshop showing the various scores users gave to the criteria that were evaluated in the scenario-based exercise. In the table, we associated the equivalent QUIM criteria with the ROUTE-TO-PA criteria and showed the usability factors that each line of criteria measured as prescribed by QUIM model (Table 14).

Table 14: QUIM usability factors and criteria (excerpt)

Usability Criteria	Usability Factors			
	Effectiveness	Satisfaction	Learnability	Accessibility
Attractiveness		●		
Minimal action		●	●	●
Consistency	●		●	●
Self-descriptiveness			●	●
Accuracy	●			
Readability				●
Simplicity			●	●
Familiarity			●	

Users rated their levels of perceptions for each usability criterion by selecting a corresponding figure from 1 to 5 in response to the usability statement for that criterion (Table 15). Statistically, we believe that the average score, of all the scores for a particular feature, provides a better value to judge the usability of that features. On the bad side, a mean value of 1.0 indicates that the feature has been rated as having a very bad technology usability rating by many users which translates into the technology design is very bad and unable to support a good usability. On the good side, a mean value of 5.0 indicates that many evaluators believe the usability of the system feature is very good and that means it is supported by a very good technology design.

Table 15: Computation of average score from usability rating

Features Tested	Criteria Tested (ROUTE-TO-PA)	Equivalent QUIM model Criteria	1.) Very Bad	2.) Bad	3.) Can't Say	4.) Good	5.) Very Good	Av Score	QUIM usability Factors
Search dataset	Relevance	Accuracy	3	2	2	4	8	3.6	Effectiveness
	Accuracy	Accuracy	2	3	3	8	3	3.4	Effectiveness
	Clarity	Readability	2	3	4	6	4	3.4	Accessibility
Data resource views	Simplicity	Simplicity	2	2	5	6	4	3.4	Learnability, Accessibility
	Understandability	Simplicity	2	1	4	10	2	3.5	Learnability, Accessibility
	Structure	Attractiveness, Readability	2	2	0	6	7	3.8	Satisfaction, Accessibility
Description of data resource	Understandability	Simplicity	4	2	1	9	3	3.3	Learnability, Accessibility
	Informativeness	?	4	1	4	5	5	3.3	?
Social discussions on data resource	Relevance	Accuracy	3	3	2	10	1	3.2	Effectiveness
	Informativeness	?	3	3	4	7	2	3.1	?
	Understandability	Simplicity	3	1	3	8	4	3.5	Learnability, Accessibility
	Familiarity	Consistency	3	2	7	6	1	3.0	Effectiveness, Learnability, accessibility
	Self-descriptiveness	Self-descriptiveness	3	3	6	7	0	2.9	Learnability, Accessibility

	Speed, number of clicks	Minimal action	3	2	6	7	1	3.1	Satisfaction, Learnability, Accessibility
--	-------------------------	----------------	---	---	---	---	---	-----	---

Logically, a mean value of say 3.0 which is above the median value (2.5) is considered above average quotient and is accepted as a reasonably good usability score for any feature evaluated by the users. An overview of the average scores on each of the evaluated features in each functional area of ROUTE-TO-PA system is provided by Figure 24.

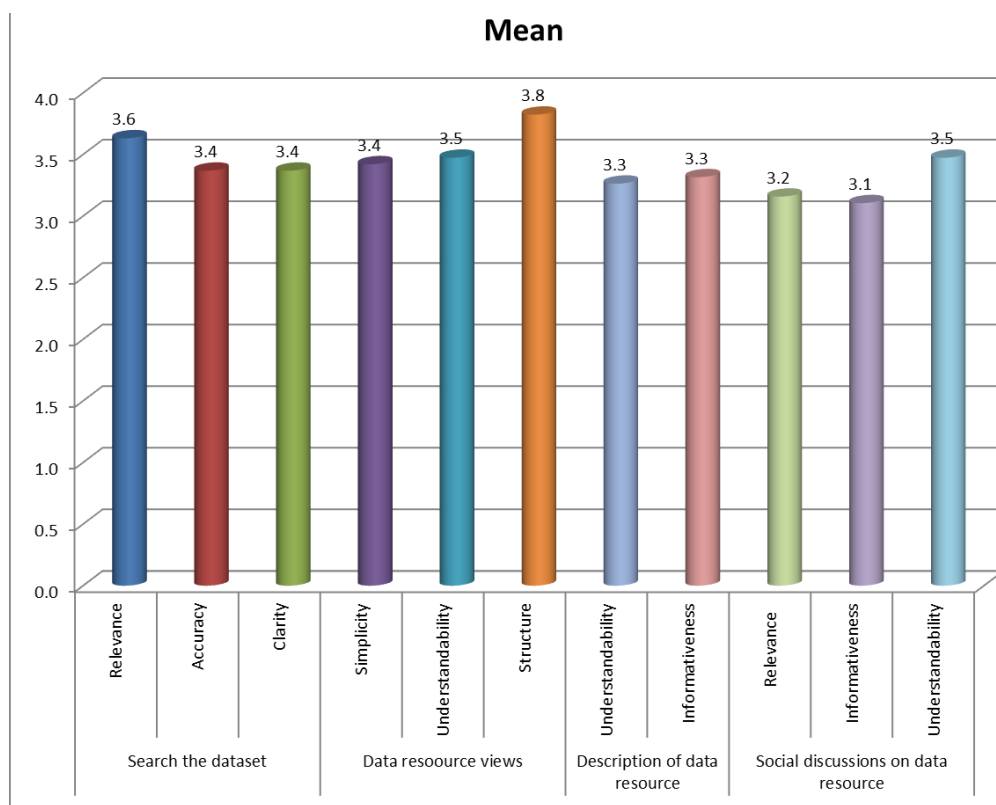


Figure 24: Evaluators' rating across functional areas of ROUTE-TO-PA system

We observed that the majority of the features have mean values above the 3.0 mark, that is, they are considered to be fairly usable in the context of this case study. This reveals that the structure of the data resource views – output presentation and layout on screen has the best rating (3.8) while informativeness features under ‘social discussion on data resource’ as it appears on the TET interface for users to learn more about the dataset before deciding to select for use received the poorest ratings (3.1). The poor scores [informativeness: 3.1, Relevance: 3.2, (both under social discussions on data resource) and Understandability and informativeness (both 3.3 and under description of data resource) were justified by other comments such as the lack of tooltips and on-screen help information for users which make learnability and issue also. These issues have now been resolved in the beta version of TET.

The major problem with the first generation open data platforms such as CKAN is that the system is complicated and difficult for non-tech savvy citizen users to use in consuming the data resources available on the data portal. ROUTE-TO-PA's major aim is to provide tools for ordinary citizens with little computing skills to consume open data resources with relative ease. In the usability evaluation workshop, evaluators provided both positive and negative comments regarding the usability of the technology of ROUTE-TO-PA and some of the comments show that the system is addressing the main objectives of the project.

The Positive Comments

Improved User Interface: Overall, users believed that the TET subsystem now, has simple and clear user interfaces especially the dataset search and selection window. Other comments add that there has been a big improvement on the user-friendliness of the data analysis tools from the initial alpha release which was a mere extension of the CKAN platform, to the second version (beta) that was a total rethink of the technology design. In the current version, we adopted a “non-cluttered Google-style” landing page where users are given a familiar-looking search bar to type in their query to search a topic of interest to them. The simplicity which also comes with the opportunity for users to enhance their search result using filtering tools made a lot of difference in usability improvement. The on-screen presentation structure which affects satisfaction has also improved tremendously along with clarity as well as familiarity because the new TET is more consistent with other platforms which users are already used to.

Improved User Experience: The improved user experience perceived by evaluators comes as a combination of the improved data processing and presentation tools in data search and analysis. Compared with standard CKAN and other open data portals, users have the advantage of being able to see a list of datasets in the topic of relevance to them at the second click and the availability of many helpful filtering options and data description with the metadata rating enabled users to understand more about the dataset they were viewing and that helped them to decide whether such datasets were of good quality or not and whether it was relevant to them. First generation open data platforms are more difficult to use because certain level of computing skills is required to make sense of datasets available on a portal. So, unlike earlier data platforms, in ROUTE-TO-PA, users are able to view simple data tables and to create visualisations of the datasets on TET by following the instruction manual. User comments support the fact that usability criteria such as understandability, relevance of search results, informativeness of textual materials on platform, minimal action, simplicity, accuracy, clarity, structure (attractiveness) fall above the mean threshold of 3.0 (Figure 24). The availability of tools on ROUTE-TO-PA made it possible for better and easier consumption of datasets because they enabled users to make better sense of the datasets of interest through simple table presentation, geo maps and other visualisation charts. These possibilities informed user comments that the system is more data-usage centric than, not just the first version of ROUTE-TO-PA, but perhaps than other traditional open data portals.

The Shortcomings

Interface integration problem: On the negative side, users perceived a seamless transition problem between the TET and SPOD interfaces because in the version that was evaluated, these two interfaces had not been integrated completely. During evaluation process, users were made to copy dataset API links from TET to SPOD interface in order to study the datasets information further through the use of the advanced visualisation tools and also in order to discuss matters arising on the visualised datasets with other participants on platform. Furthermore, many users raised the issue of informativeness. Users complained about not having enough user help documentation such tooltips and pop-up notes on platform to explain unfamiliar terms and guide users on what do at certain points or with certain buttons.

The usability manual containing the Dublin traffic problem scenario and details of the Google survey are provided as appendices:

Appendix 2: Usability Evaluation – Instruction Manual and

Appendix 3: Usability Evaluation – Google Survey Report respectively. Based on the iterative processes of component development followed by testing and usability evaluation and by system upgrade with evaluation findings, we hope to carry out further testing and evaluation processes on the just release beta version later in February, 2017. The result of that evaluation will be used to improve TET usability as necessary.

8 CROSS-CUTTING DESIGN CONCERNS

This section describes mainly how Privacy concerns were addressed during TET development.

8.1 PRIVACY

During the development of TET 2.0 platform the main concern was related to Data Privacy and Security. One of the system requirements was to *ensure privacy & security of the user data it stores with user permission to enable the user to understand what type of data is stored and how it is used.*

At this stage the platform provides all the functionality to a non-logged in user and no data related to the user is stored on the system.

Non-publicly released features (under development and candidates to release in the next version of TET) will provide a login form which will allow visitors to create and to store a user account which will enable the user to experience a better interaction with the platform. This will be realised by the personalised experience.

In cases of logging in with personal details, the system will adhere to the required standard of maintaining the safety of user personal details and maintain the user account and profile data against unauthorised access.

Moreover, the user will be informed what data is stored and a dedicated page with the TET data policy will be available.

8.2 RESPONSIBLE RESEARCH AND INNOVATION COMPLIANCE

Responsible Research and Innovation is a cross-cutting issue of the entire EU framework programme for Research and Innovation. In 2012, the European Commission defined RRI as follows: Responsible Research and Innovation means that societal actors work together during the whole research and innovation process in order to better align both the process and its outcomes, with the values, needs and expectations of European society. RRI is an ambitious challenge for the creation of a Research and Innovation policy driven by the needs of society and engaging all societal actors via inclusive participatory approaches.

In ROUTE-TO-PA, deliverables we provide an explicit indication of how the RRI criteria are addressed. Therefore following the RRI recommended structure now we elaborate how we addressed the key issues:

Public engagement

In order to perform the requirement elicitation activity, we involved public administrators, active citizens, researchers and data journalists, social organisations and entrepreneurs through the organisation of formal workshops. Workshops have been organised in each of the five pilots.

Gender equality

Through workshops for requirements elicitation participants roughly half of them were female.

Science education

N.A.

Open access

All the scenarios and user stories which are referred by this deliverable, and the deliverable itself, are offered openly through the project website, at www.routetopa.eu/public-deliverables/. This deliverable will be also available in its final format.

Ethics

According to D1.5 Ethical Policy, in order to conduct the user workshops, each research partner submitted a proposal for ethical clearance to the relevant authorities in their context (University, Research Center/Company). The main goal of this ethical clearance is to guarantee the right, safety and risks-free for participants in the scientific research activities. Each partner got the ethical approval all of them attached to the Deliverable D1.5 Ethical Policy.

According to the process submitted to ethical clearance, workshops to gather requirements have been organised giving material (consent forms, questionnaires etc.) during and right after the workshops. So, workshop organisers obtained informed consent to participants.

In addition, ethical issues have been considered within the collected requirements. In particular, the UC23 Requirement states that Platform ensures privacy & security of user data it stores with user permission to enable user understand what type of his data is stored and how it is used.

Governance

N.A.

Sustainability

N.A.

Social Justice / Inclusion

N.A.

9 FUTURE DEVELOPMENT PLANS

For the final release of the TET platform, a thorough roadmap for development activities will be created together with pilots and based on the feedback gathered from other stockholders. The creation of a roadmap will be a collaborative process and will be tightly aligned with goals and objectives of the RTPA project. Apart from planned features new experimental features will be developed to find creative ways to address the challenges related to open data adoption and exploitation, especially we would like to exploit state of the art developments in Artificial Intelligence and Machine Learning to add smart features to open data platforms. Table 16 lists the summary of candidate features for the final release.

Table 16 Candidate Feature for TET final release

Category of Features	Details
ETL Aspects	<ul style="list-style-type: none"> Workflow system: an easy to use workflow system that will allow data publishers to extract data from existing repositories, clean it and transform it to appropriate structure that is publishable on open data portals. Interface to tools for cleaning and transforming data e.g. Open Refine to enable a data publisher to apply more complex ETL operation on their data.
Analysis & Analytics Workbench	<ul style="list-style-type: none"> Analysis and Integration of time series data: The feature will help in making sense of time series data scattered around the open data portal and will help in exploring new facts and answering complex questions. Export to existing data analysis tools e.g. Excel, Kibi, Google Sheets etc. Support for mature and familiar data analysis tools can significantly enhance the chances of open data exploitation by a wide range of users. Text analytics: Text is one of the commonly found formats on open data portals, extracting meaning from a long text document is very useful for citizens looking for quick answers to simple queries. Tools for handling Linked Data: use of open standards such as linked data standards can add value to open data, however existing open data platforms have limited support for linked data. We plan to use linked data not just as a format for publishing but as a means, to enhance querying and search features offered by the platform. Advance data analytics workflows. Rich visualizations.
Dataset Recommendation	<ul style="list-style-type: none"> This feature will be improved by creating a model with features extracted from actual content rather than just metadata. Explanation for links between the datasets discovered by recommender
Visualization	<ul style="list-style-type: none"> Dashboard for visualizing key indicators: There are many key indicators related to different sectors which are relevant to a wide range of audience, however these indicators are scatter around different files. This feature will let data publishers easily extract these indicators and display them using a rich dashboard. Specialized visualization for specific datasets with pre-defined templates. Chart recommender: automatically will recommend an appropriate chart for the datasets. Export to existing visualizations tools.
Advanced Searching & Filtering	<ul style="list-style-type: none"> Fine-grained searching: search within the content of datasets. Q&A interface: Helps user to get instant answers to common asked questions. More options filtering
Interaction between TET and SPOD	<ul style="list-style-type: none"> Improved interaction between TET and SPOD with enhanced API for data exchange. More SPOD widgets on TET user interface 'Share on SPOD' button Integrated authentication mechanism
Help System	<ul style="list-style-type: none"> Help system and more video for end user training. Guided tour of TET
User Interface	<ul style="list-style-type: none"> Improve cross-browser support of TET

	<ul style="list-style-type: none"> • Improve mobile version of TET
--	---

All the new features will go through the same rigorous testing and evaluation as previous releases. Agile development will help in delivering key features on time; dropping any failing feature and to have low risk all the time.

Features such as dataset recommendation and data integration will be enhanced with new training features extracted from the content of the datasets; current implementation uses training features mostly extracted from metadata to discover new relationships between datasets. The feature will be enhanced to explain user links discovered between the datasets. Dataset recommendation and integration features will be combined together to allow easy integration of datasets. Support for time series data integration and analysis will be added to address various use cases related to that. New features such as configurable dashboards for key indicators and an analytics workbench will provide advance analytics capabilities to users. Integration between TET and SPOD will improve with an API for sharing data between the platforms. The same principles of design, implementation and evaluation discussed earlier will be employed to refine the implementation in an iterative way towards the final release of TET.

10 CONCLUSION

The deliverable 4.5 beta version of TET presents the details of requirements, design, implementation and deployment activities performed towards the development of the beta version of TET. The TET platform has gone through the process of significant enhancement since the alpha release. The user experience has been significantly improved to meet the requirements of ordinary users-citizens, at the same time, the original tech-user-oriented interface has been preserved as an option for end users.

The focus of the beta release has been mostly the usability, human-data interaction and system stability which were the areas highlighted in requirements elicited from user-feedback collected to guide the design and implementation for TET beta. Updates to the initial system requirements through updates to functional requirements were made possible via the comments and suggestions from a wide range of user evaluators. These evaluators (testers) played various roles – citizen open data platform users, researchers, developers, public administrators and students across genders and a significant spread of age brackets. The resulting implementation evaluation indicates a significant improvement in usability, cutting across UI friendliness, uncluttered “Google-style” home page, support for keyword search, various views on dataset information understandability, sharing features and help tips, including demo videos. The TET-SPOD integration has been improved by importing and including the discussions around datasets on SPOD platform. Moreover, the visual part has been improved by introducing clear styling and simpler structure of the output on screen.

Agile methodology has been used as the primary software development approach for TET development. This enabled the pilots to get fully-working solutions at the end of each cycle. The method, giving more ownership to project stakeholders, also made the adjustments better aligned to user-needs. A simple and easy to use interface is the hallmark of the TET beta release. The design and the implementation were inspired by popular search engines and ecommerce sites which resulted with a truly user-centric interface for open data portals. The new user interface guides users in searching and discovering relevant datasets and exploring resources associated with datasets in a very easy way. TET provides users with powerful yet easy to use data analysis features, which allows non-technical users to make sense of available data in intuitive way. More advanced users can use TET’s query console, PivotTable and Data Integration features to perform advanced analysis of the data. The interaction between TET and SPOD is improved with inclusion of SPOD charts, the export to SPOD implementation and the SPOD discussion widget on the TET dataset page. Advanced machine learning algorithms are used to intelligently match and recommend related datasets, which is a big leap towards the development of an innovative, AI-driven open data platform.

During the TET evaluation process many areas for improvement were identified and will be improved in the upcoming releases of TET. A community will be built around the TET tools and additional documentation will be added to support the developer community. The final release of the platform will focus on data analytics, better data sharing with SPOD and integration of related datasets. The work will be performed with the feedback provided by the pilots and gathered from evaluation workshops aligned to project goals. All the bugs and open issues will be resolved and the platform will be stabilized to ensure that the final product is delivered free of major defect and up highest standards.

11 APPENDICES

APPENDIX 1A: PRE-USABILITY EVALUATION TESTING – 2 SAMPLES OF MANUAL WITH USER COMMENTS

Note: This appendix features two manuals one for testers playing the roles of open data publishers and the other for testers playing the roles of open data platform users. Both manuals have common sections – the introductory section and other relevant sections as information necessary for tester to understand their roles. These sections are shown in this appendix only for the manual for testers playing the roles of publishers – just below this note. Similarly, the conclusion and references section of the manuals are same and hence not repeated in this appendix for reason already mentioned.

The main instruction section begin from the heading '**Solution Evaluation Instructions for TET Components (Alpha Release)**'. Appendix 1B features details of the comments from Pre-Usability Evaluation Tests and analysis of the results.

Testing Instructions for Tester as an Open Data Publisher

Introduction

This document serves as a user guide for the purpose of evaluating the components (or functions) of the TET subsystem so far developed by the Development Team (Insight Centre, NUIG). Insight Centre is a consortium partner of the Raising Open and User-friendly Transparency-Enabling Technologies for Public Administration (ROUTE-TO-PA) project. ROUTE-TO-PA is an innovative project focused on prototyping and piloting the integration of platforms for Open Data and a social network engine. The system is enhanced with tools to facilitate Open Data understanding, metadata linking, and personalization of data usage. Two of the primary ROUTE-TO-PA project objectives are:

1. To facilitate the transition into the next generation of Open Data portals by creating tools that will enable citizens to engage themselves socially over Open Data resources. This is termed *Social Platform for Open Data (SPOD)*.
2. To provide tools that could be integrated into existing Open Data platform to deliver greater data transparency, better quality and improved understandability. This component is termed, the *Transparency Enhancing Toolset (TET)* and its aim is to provide basic analysis tools for reducing datasets into more understandable forms including visual form for users.

Dataset Handling Activities: TET Alpha release supports three sets of user activities namely –

- Data publishing activities
- Data search and discovery activities, and
- Analyse and visualise dataset activities

The TET Solution Evaluation Guide, produced by Insight Centre (NUI Galway), is split into two documents: The Data Publisher's Guide (this document) provides Data Publishers (acting as dataset uploaders) with a set of instructions to test and evaluate the functionalities supporting **dataset publishing activities**. The second document which is the End User's Guide, provides the data consumers with a set of instructions to test and evaluate the functionalities supporting dataset search and discovery activities as well as data analysis and visualisation activities. Similar to TET Solution Evaluation Guide, the development of SPOD Solution Evaluation Guide is being handled by another partner of the ROUTE-TO-PA consortium and is concerned with the evaluation of SPOD functionalities as regards the goal of enabling the interactions between Open Data users over the integrated SPOD and TET system.

Solution Evaluation and Recording Observations

Please read each set of scenario, test data and instructions given under each component to be tested and evaluated by you before starting to execute the test instructions. This is to enable you understand the context of the user roles and objectives, the TET solution and the direction of the evaluation instructions before you begin to click around the windows. The scenarios mimic the goals of the users while on the platform and TET functions are the capabilities that ROUTE-TO-PA TET subsystem provides the user to achieve these goals. Follow the instructions as given below each solution evaluation heading (representing a TET component) and record your actual observations (actual results) against the expected results in case of deviations using the **Comments** section of the survey tables provided. Please note that your comments should relate only to the context of the component you evaluate and not anything else within the whole evaluation exercise. This means that you should not add comments that do not relate to that particular component you have tested so as to avoid confusion with other components' test or evaluation results. If you find no deviations from the expected results documented in the guide, please provide simple and concise responses to the survey questions. The first part of the table contains a few more '**Direct**' questions to seek your opinion regarding your satisfaction with the functionalities of TET that you have tested one after the other. The questions in this section will attempt to relate TET functionality to the objectives of ROUTE-TO-PA system components as solutions to the needs of the users in accordance with the requirements that were gathered from stakeholders earlier in the project lifecycle. The ***Solution Evaluation and Validation*** process will use the questions embedded with ***solution performance measurement metrics*** to seek and measure your opinions regarding the capability of each solution component to meet one or more user needs. These needs are expressed in the accompanying scenarios in the forms that:

- support **Data Transparency** for users
- facilitate **Data Accessibility** to users
- enable users to (pre-) **Check the Metadata** elements of datasets to decide on suitability for their use
- enable users view the **Measure and Visualisation of the Completeness of the metadata** provided for each dataset on the portal by the suppliers
- enable users to **Personalise Search criteria for Datasets** (while logged in and while not logged into the system)
- enable users to **Request Recommendation for Datasets** from the system search functionality
- enable users to **Link Datasets** together so that the datasets becomes more visible or accessible to users
- enable users **Analyse and Visualise Datasets** in order to assess the quality of the dataset and/or gain better understanding of the meaning contained in the datasets, etc.

Definitions of Qualities Supporting Data Transparency

The definition of terms below is offered in the context of data Transparency objectives set out by ROUTE-TO-PA project. In this context, data transparency to the citizens under given political governance refers to how well the open government data meets the qualities of **Accessibility, Usability, Auditability, Informativeness and Understandability**. Each of these qualities are further reduced to elements that most closely describe the qualities of an open dataset in the context of public administrators' initiatives (and others alike) to proactively provide government data, non-proprietary and non-private datasets for citizens consumption and allowing the citizens free access to and usage of these datasets.

Table 1: Transparency support qualities (Cappelli, 2009)

Usability	Accessibility	Auditability	Informativeness	Understandability
<ul style="list-style-type: none"> • Uniformity • Simplicity • Operability • Performability • Adaptability • User-friendliness • Intuitiveness 	<ul style="list-style-type: none"> • Portability • Availability • Publicity 	<ul style="list-style-type: none"> • Validity • Controllability • Verifiability • Traceability • Accountability 	<ul style="list-style-type: none"> • Consistency • Integrity • Accuracy • Completeness • Clarity • Comparability • Correctness • Currency 	<ul style="list-style-type: none"> • Conciseness • Composability • Decomposability • Extensibility • Dependability

Usability:

Usability quality refers to the capabilities provided by ROUTE-TO-PA solutions to enable users utilise open data resources by improving the **simplicity** of the tools to reach the datasets and use it in the interest of the user. It improves the **uniformity** in the supply standards and presentation manners of the datasets by creating a metadata form for completion at time of dataset release by data suppliers. ROUTE-TO-PA solutions seek to enhance system **performance** in the delivery of open data services and to enrich the technical systems into **intuitive user-friendly** portal **adaptable** to the needs of the non-tech savvy ordinary citizens. The user-friendliness quality of the system also improves the obligation of data suppliers to upload datasets much more easily in a standardised format as facilitated with the use of the metadata form.

Accessibility:

Accessibility quality defines the ease of reaching open datasets by making the datasets more **available** and **portable** by delivering them to consumers through multiple devices such as desktops, laptops, tablets and smartphones. In this sense, ROUTE-TO-PA solutions enable improved accessibility of open datasets through multiple portals and hence **availability** to users over free and open source portals. In the long run, improved accessibility via many types of platform, including on-the-go platforms, increases the **publicity** of open data availability, uses and awareness about the concept in general to the public.

Auditability:

This quality enables open data consumers to carry out a quick check on the metadata and other extra information provided by the supplier on the dataset. Extra information could be the description of the dataset, frequency of upload/update, name of maintainer and the contact information. Thus auditability quality supports user ability to **verify** the **validity** of the dataset and understand more about **traceability** factors affecting the dataset. This quality therefore supports **accountability** and **responsibility** duties which the data supplier owes to the consumer community because knowing about the above criteria of auditability, data suppliers are influenced to cooperate. Auditability factor provides criteria that enable users to **control** (influence) the cooperation of the data suppliers by holding the supplier responsible for inadequate or poor quality data supplies through quality rating and remarks. In general, the elements of auditability helps not only to improve the quality of the datasets being published, they also have the capability to improve the integrity of the sources of the datasets so long as the suppliers cooperate with the demands.

Informativeness:

Informativeness quality directly impacts on the quality of data supply by causing suppliers to conform to certain level of **consistency** in dataset upload duties. For example, ROUTE-TO-PA solutions have the functionality to compel data supplier to complete all relevant metadata fields before being able to upload the dataset onto the portal. Simply put, the functionality compels data suppliers to conform to the metadata **completeness** quality; and together with the consistency quality, the dataset dealing on same topic can be **compared** more easily thereby providing better information on the **correctness** of one dataset against another. The **currency** quality refers to the **timing of validity** of the dataset – the **freshness** or **newness** of which may be predicted by such metadata items like date of upload and frequency of update combined. By meeting the consistency, completeness and accuracy qualities, data users can predict the level of **integrity** of the supplier; and if all factors are met, then the informativeness quality of the dataset is significantly enhanced.

Understandability:

ROUTE-TO-PA solutions provide tools that make open data more understandable to users. These tools enable data users to **de-compose** and **re-compose** datasets into other forms that are more easily understandable or **comprehensible**. For example, tools that can be used to create visual graphics on datasets, and the “drag and drop” functionalities which creates dashboard effects that explains the meaning of a dataset quickly. For instance, TET tools that enable components of a tabular dataset to be de-composed to reveal missing values such as blank cells of a columnar data array. TET analytic capabilities can **de-compose** a dataset to reveal dataset qualities in a **concise** manner which can help users decide on the suitability of that dataset for further uses. TET basic analytics and visualisation tools can **compose** or **re-compose** the components of a dataset to reveal dashboard values and graphics such as bar charts, pie charts, stacked bar charts, heat maps etc., not originally supplied with the dataset and which make the meaning of the dataset more comprehensible to users. To support these functionalities, the tools of ROUTE-TO-PA are **extensible** and always **dependable** for them to be able to serve the needs of all data consumer community. TET functional requirements (Alpha version), the details of the TET objectives as well as the system qualities are clearly laid out. The table further presents arguments that justify the supports that the system qualities offer TET objectives.

Table 2: System (Solution) Requirements - TET (implemented components only)

Functional Requirements	TET Objectives	System Qualities Supporting TET Objectives	System Quality Justifications as Support for TET Objectives
User Login	System provides login form or data fields for user to complete and save	Accessibility, Usability, Data Security and	In so far as this function has no direct link to the TET objective, Privacy & Data Security, and simplicity of use are however, relevant to users while using TET tools. In terms of systems performance quality, ROUTE-TO-PA solutions objectives demand that the login functionality must be Usable

	to account so that user can enjoy a better experience of using the platform.	Privacy, Simplicity	– meaning that it must be operable, simple, adaptable to multiple platforms and other social medial login systems, and be user-friendly as well. Furthermore, the login must be Accessible (portable or available) via multi-platforms to enable users to access their account and consume open data through many types of computing devices including smartphones
Check Metadata Quality and Completeness	Visualise metadata completeness quality for a dataset by displaying the measurement bar, show percentage rating for the completed metadata fields for that dataset to enable users decide on the suitability for their intended purpose.	Usability, Accessibility, Auditability and Understandability	Metadata completeness rating supports usability of dataset by enabling users to predict how appropriate and current a dataset is; know the dataset source and the integrity of the supplier. This tool compels data suppliers to provide more metadata on the dataset they supply. Availability of metadata on a dataset enables users evaluate the quality of the dataset in a way as well as carry out audit of the source, integrity of the supplier and how current the dataset is. All these are quality metrics that inform users about how suitable (usable or fit-for-purpose) the dataset is. The accuracy of the dataset metadata with regards to how well they describe the dataset helps in searching it out thereby improving its accessibility and relevance to user when correctly retrieved by user query parameters. In terms of the system quality, metadata completeness tool must meet the demand for simplicity, operability and user-friendliness and in addition, data required in the metadata form must be concise and dependable to promote usability . Besides, the tools must be portable and available via multi-platform capable of enabling users to truly verify and validate as well as trace more information about the dataset to the source/provider who should be accountable for the datasets they provide. These quality components promote dataset auditability .
Add / Update / View Provenance and metadata related to a Dataset	Support for Provenance records: these are the values entered for the fields: Source, Versions (Current & Previous) and Frequency of publication fields in the metadata form screen of the dataset.	Usability, Auditability, and Informativeness	The possibility to view the provenance records (i.e. data source, ownership & history – upload/update dates) are additions to the metadata quality metrics e.g. auditability (traceability) and which make users more confidence of the appropriateness and integrity of the dataset selection. The usability quality demands that TET functionality is operable, simple and easy to learn by users so they can understand the other quality features of the system and apply them correctly (system usability). More on this is presented in the case of “ <i>Check Metadata Completeness</i> ” above.
Analyse a Dataset	System provides simple analysis, analytics and visualisation tools to enable users (including non-tech savvy) to analyse datasets & visualise results in common graphs to enhance comprehension of data meaning	Usability, informativeness, Correctness, Intuitiveness, Performance, Accuracy	System capability to enhance data informativeness through further analyses of a dataset to improve the clarity of data meaning, enables comparability between dataset components or between different datasets; enables users judge the accuracy of datasets, and all ensures that users can comprehend the meaning easily and facilitate user confidence. The results of analyses should facilitate users’ decisions on suitability (usability, correctness and accuracy) for their purposes. However, system analytic tools must be supported by route-to-pa’s state of technology with simplicity, intuitiveness and reasonable performance efficiency while producing correct/accurate and reliable results. Finally, the tools should provide understandability capabilities for users to be able to de-compose, re-compose dataset components so as to reduce the dataset into concise portions.
Link a Dataset	The part of this objective so far implemented is the part that enables dataset linking with other dataset(s) for searchability enhancement and recommendation purposes; as support for dataset accessibility. The aspect that enables dataset integration is not yet implemented.	Usability (Supportability, Operability, Extensibility, Performance) and Accessibility (Availability and Publicity)	System capability that links datasets together helps in realising the recommendation and searchability goals in that if any of the linked datasets are searched out by a user, the system then auto-recommends the other(s) to the user as related datasets. In this way, datasets linking improves user awareness of and accessibility to the recommended dataset through exposure to searching and recommendation functionalities. Linking datasets together publicises datasets better and makes them visible to users and traceable to their repositories and retrievable by querying criteria (making the dataset more available to users). To achieve these aims, TET data linking tool must possess the quality to support several types of standardised file formats. It must meet usability qualities (simplicity, operability, user-friendliness with excellent level of performance) by data/platform administrators and data uploaders/suppliers. Moreover, the functionality entails system extensibility quality that the tool should possess especially in sub-functions such as data integration, merging and synchronisation (not yet implemented).
Personalized Search (User, not logged in)	System provide options e.g. role, location, dataset	Usability, Informativeness (Accuracy,	In personalised search, while user is not logged in, the system uses predefined data categories such as <i>Unemployed Persons, Climate, Education, Finance, Economy, Business, Person with disabilities, etc.,</i> to

	category, etc., for user to personalise search as support for data accessibility and improved search experience	currency, correctness, consistency), Reliability, dependability,	direct users to datasets that could be of interest to them. These dataset categories are iconised on the interface to improve usability (performance, operability and interface-friendliness) thereby improving searching experience for those not technically savvy. The system has to be accurate and consistent in fetching the datasets in the category with further improvement through a range of filtering tags when already inside dataset category to enhance the reliability and dependability (informativeness) on the search results. By applying all these tools put together, users can enjoy a level of performance efficiency and operability of searching experience with reliable and informative search results.
Personalized Search (User logged in)	System provide option e.g. interest, role, location, dataset category, etc., for user to personalise search as support for dataset accessibility and improved search experience	Usability, Reliability, Accuracy, Data Security & Privacy, accessibility, availability	Instead of using the iconised data categories, users can actually prearrange their user profiles to enable the system apply their specific profile data in searching repositories for datasets that best match their individual interests and profile details. In this situation, system must maintain accuracy and reliability in using the details of each user to retrieve relevant datasets to the user while also strictly adhering to the data privacy and security requirements. In pre-defining their profile data plus personal interests e.g. occupation, stated interests, hobbies, age, location, etc., the system should possess usability (operability and performance) qualities while searching and displaying results of searches to users. The result of this functionality supports data accessibility and availability to data consumers and such search results should be reliable .
Recommendation for Datasets (User logged in & Not logged in)	TET system provides user with dataset recommendation in Dataset categories based on user predefined account profile options (location, role, age, email, sex, marital status, disability, interest, etc.) when user is logged in) or based on popular data searches made by user community (when user is not logged in).	Usability, Personalisation features, Simplicity, Data Security, Interface friendliness and Performance	Data recommendation is based on user profile information on user account while user is logged in. The system uses the user data to personalise dataset that are recommended by default to the user. To enable users set their profile correctly so as to enhance recommendation of relevant resources, the system interface must be user-friendly, simple, operable and easy to learn (usability). Because it concerns using personal data, data privacy and security is important quality the tool must possess. As in the case of the Personalised search, the system utilises the personalisation features of the user account which must be provided by ROUTE-TO-PA system. Features include – personal details plus interest, occupation, education, social and health status. While the user is not logged in, the system recommendation of datasets is based on the popular data searches made by user community.
Request Recommendation for Datasets	Provide user with options (location, role, age, email, sex, marital status, disability, interest, etc.) to enable user personalise account and receive content suggestions based on account data. (Option not yet implemented: User alert via email)	Usability, Personalisation, Simplicity, Interface friendliness, Privacy & Security, Integrity and Performance	Request for Data Recommendation functionality goes a little further than the system recommendation by default using the user profile details. In the case of request, the system provides extra features to enable user proactively seek data supply notification . So the system offers advanced profile fields for user to enter extra user details and email , and enrol for data upload and update notification via emails or instant messaging system such as Facebook. If the user opts for social media notification, user must integrate his/her social media account details with the ROUTE-TO-PA platform. All other system qualities applicable to Data Recommendation are also applicable to Request for Data Recommendation . Please note that the request for dataset recommendation function is not yet implemented in the TET Alpha release. It will be released in the subsequent version to be determined.
Enrich Profile	System provides more dataset accessibility options to enable users add more data to profile so that user can search & receive more relevant resources based on his/her enriched profile details	Operability, Accuracy, Accessibility, Dependability, Data Privacy & Security	For a better feel and to enjoy more functionalities of the ROUTE-TO-PA system, the user will be given the option to enter more advanced details into their profile accounts on platform. The fields for the details are same as those that would enable user to explore the advantages for personalised search, recommendation from more relevant resources, location-and social status–related recommendations, enablement of data supply alert etc.

Solution Evaluation Instructions for TET Components (Alpha Release)

Evaluation 1: Check Metadata Quality and Metadata Completeness Rating

Assumed Scenario:

You are a user in need of a specific dataset for planning purposes. You have no time for repeating the job of planning and cannot afford to have a data-driven plan that is not reliable or does not meet the needs of the users of the plan. You want to check on the metadata quality of some datasets of interest to you (one by one) to see if and all the datasets of interest meet the metadata quality specification for the use into which you intend to put them. For example, knowing the source and update frequency can raise or reduce reliability of the dataset, the metadata completeness rating can inform you of the use into which the dataset is capable of being put. For example, linking or integrating a couple of datasets together before analysing them to gain better insight necessary to support planning purposes.

Test Data:

- URL: <http://vmdatagov01.deri.ie:8080/dataset>
- Dataset name: **Journey Times 11122015 13.01**

Actions and Expectations/Results/Screen displays:

Type the URL on your browser bar and press enter to open the dataset listing page of TET/SPOD enhanced CKAN Platform.

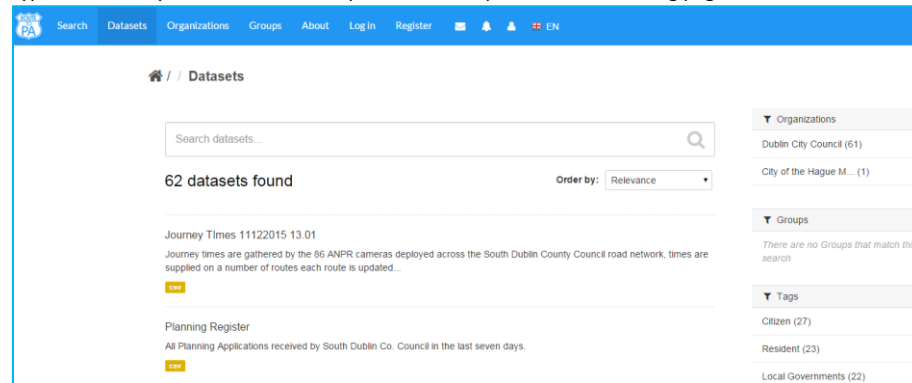


Figure 1: Dataset listing screen of TET/SPOD enhance CKAN Platform

You should see on your screen, a number of available datasets (**xx datasets found**) and, **Order by:** (drop-down menu) with sort tags: **Relevance, Name Ascending, Name Descending and Last modified**. Click on the dataset name: **"Journey Times 11122015 13.01"** (or any other dataset name) on the page. The system should display the metadata screen similar to figure below.

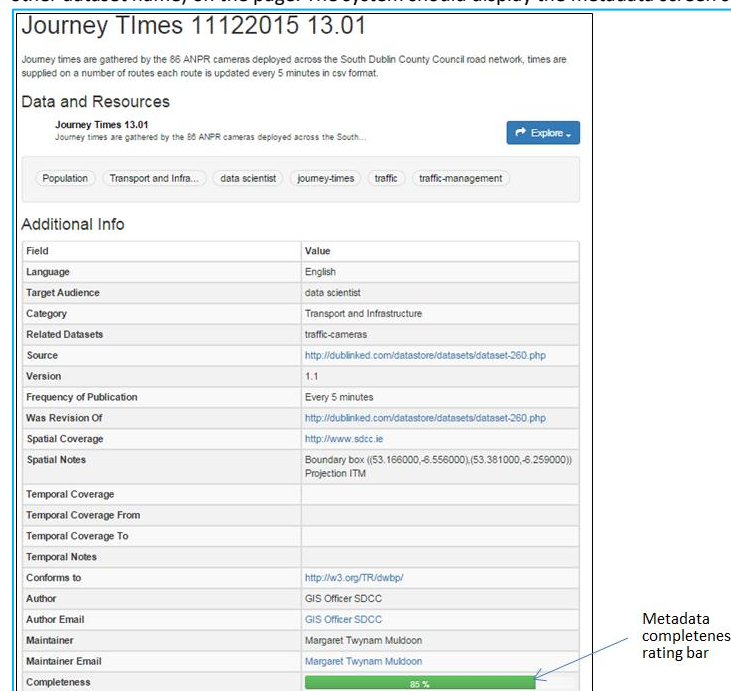


Figure 2: Metadata screen for a dataset

Observe the metadata fields on the left of the table and their corresponding values supplied by the data provider on the right. The metadata completeness rating is displayed at the bottom of the table as a horizontal bar and it measures the completeness (in percentage) of the metadata supplied. The more values of metadata supplied for the fields, the higher the quality of the metadata, the longer is the bar and the higher is the percentage rating displayed.

Table 3: TET User Survey

Questions	Responses
1. Metadata helps in searching out the dataset and in ensuring users of the appropriateness of the dataset for their uses. Does this functionality seem good enough to improve <i>accessibility</i> and <i>usability</i> of datasets?	Some of the metadata may require explanation. What is 'temporal coverage' and how is it different from 'Temporal Coverage From' etc. Perhaps a hover-over quick help function would improve functionality.
2. Assume that metadata provided for a dataset correctly describes the nature of the dataset. In your opinion, will this situation improve retrieval of relevant datasets to match user searches (improve data <i>searchability</i>)?	It depends if you can search based on every meta-data heading. For example 'Spatial Coverage'

	would bring back many datasets in an area that may not be at all related or relevant.
3. By compelling data suppliers to complete metadata form for all datasets they supply, open data may become more <i>available</i> and <i>accessible</i> to potential consumers. Do you agree or disagree and why?	Agree unless too much metadata is overwhelming to the consumer and not considered relevant to them.
4. Will the possibility for you to see the metadata of the dataset beforehand affect your decision to use or not to use the dataset and also save you time for searching for the relevant datasets (improved platform <i>experience & usability</i>)?	If the metadata made it clear this was the dataset I needed then it would affect my decision but I think I would preview the dataset to be sure.
5. What value does this system capability add to the existing open data platforms?	Too much metadata may put off users.
6. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	5
7. Additional Comments	

Evaluation 2: Personalized Search (User with Admin Login)

Assumed Scenario:

As a regular open data provider or consumer, you are perhaps, public sector personnel, a professional Data Journalist, a Data Scientist or a Community Leader/Activist who is always interested in government open data. You could also be an Open Data System Developer or any other data consumer. You have registered your profile and data interest on the platform so that dataset retrieval for your frequent uses is done seamlessly and more accurately.

Test Data:

- **URL:** <http://vmdatagov01.deri.ie:8080/user/login>
- **Dataset:** User details with an administrative capacity – Use the following:
 - User name: **test1**
 - Password: **test123** (After login, complete your account preferences by clicking on **Settings** button on main menu. Select your '**Role**' and '**Category of Interest**' fields)

Actions and Expectations/Results/Screen displays:

Go to the URL above and enter admin login details provided on the user account form that opens, then click on **Login** button. Note, you may tick on the *Remember me* box for auto login next time you visit the systems.

Figure 3: User Login (Admin Responsibility)

Once the **Login** button is clicked, the Dashboard screen opens showing: **News feed** (opened by default), **My Datasets**, **My Organisation**, and **My Group**. The **Edit Settings** button is shown on the right side of the screen.

Figure 4: Dashboard - News feed screen

Click on "**My Datasets**" button which is on the left side of top menu next to the **News feed** button to display the available datasets under the current admin login.

Figure 5: My Datasets screen

There is an “Add Dataset” button on top of the list of datasets displayed on the screen which the user could use to add more datasets to the system. The user can click on any dataset filename to view the metadata and the completeness bar and rating of the metadata of the dataset.

Table 4: TET User Survey

Questions	Responses
1. Do you consider this functionality to be well <i>personalised</i> and <i>user-friendly</i> ?	I think My Dataset should be the default tab rather than Newsfeed.
2. How <i>accurate</i> do you think the result of the personalised search is?	I didn't upload the datasets so I can only assume they are displaying the correct ones.
3. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	4-5
4. Additional Comments	

Evaluation 3: Add Dataset (Admin login)

Assumed Scenario:

As a regular open data provider, part of your responsibilities is to upload datasets onto open data portal. This time you want to upload a dataset to the TET/SPOD-enabled CKAN platform so that users (including non-tech-savvy citizens) can make the best of the dataset. For example, do basic analysis with the dataset using the TET tools and/or share and discuss with network members using the dataset or a processed component of it. You have created an admin account and registered your profile on the platform.

Test Data:

- URL: <http://srvgal100.derl.ie:8081/>
- Dataset:
 - User name: **test1**
 - Password: **test123** (use your set of username & password if you were assigned another)
 - Link to sample dataset to be added/uploaded unto the portal: http://roscocommon.roscommon.roscommon.opendata.arcgis.com/datasets/fbae9c86e6e1441bbea951893744f3bc_0.csv

Actions and Expectations/Results/Screen displays:


Prepare sample dataset for your uploading:

1. Download the sample csv dataset unto your local drive using the link provided above
2. Save the file but **rename** it by giving it the filename **roscocommon-roscommon-opendata-xxxxx** but in your case, replace 'xxxxx' with a 5 digit alpha-numeric code selected by you. Example of filename should resemble: **roscocommon-roscommon-opendata-ak140** or **roscocommon-roscommon-opendata-c220d** **Do not change the file extension and avoid spaces and cap letters.**
3. File naming methodology in 2 should enable filenames created by various evaluators to be unique.
4. Upload the renamed dataset from your local drive unto the platform following the instructions below
5. Provide the name of the file you uploaded in the corresponding feedback survey table 5 for that evaluation.

Further Instructions to upload dataset

Go to the URL provided (<http://srvgal100.derl.ie:8081/>) and click on the login icon located at the right corner of the top menu. Then complete the login form using the details above and click on the **Login** button. On the screen that opens, locate the **Datasets** button on the left corner of the top menu and click on it. The system should display a screen that contains an “Add Dataset” button and datasets found with the dataset filenames. Click on the **Add Dataset** button to reveal the following blank metadata screen:

Top of screen



Bottom of screen

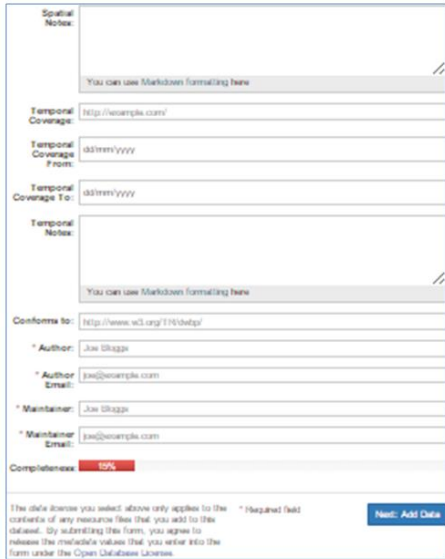


Figure 6: Metadata form screen

Complete the metadata form for at least the mandatory fields:

- **Title:** Give any title of your choice e.g. Roscommon dataset
- **URL:** Use the filename you created when you saved the csv file
- **Description:** Describe the file in your own words. One sentence is OK. See example on the field
- **Target audience:** See examples provided, you can use any word also

- **Category:** You can use – Tourism, Culture, Heritage, History, etc
- **License:** Select from menu
- **Organisation:** Select from menu
- **Source:** Use the Link to the Dataset you downloaded (i.e. the sample dataset link)
- **Version:** Use any e.g. 1.0
- **Author, Maintainer, Email:** make up your values and reuse – see example on the fields

When done, click on the **Next: Add Data** button at the bottom of the screen. If there is any error, rectify it and repeat click on the **Next: Add Data**. The system should move to the next screen to complete the dataset upload:

Figure 7: Data upload screen

Click on the **Upload** button, and then select the file that you downloaded and renamed as per the instructions above. Complete the remaining fields:

Name: provide any name for the dataset – could be same as you used in the previous screen

Description: Re-use the description provided before or similar one

Format: Use the format of the file which, in this case, is **csv**

When complete, click on the **Finish** button. The system should save entries and return to the metadata screen of the just uploaded dataset.

Table 5: TET User Survey

Questions	Responses
1. In your opinion, the process of adding a dataset is: difficult, not so difficult; easy, very easy or can't properly rate the process (<i>learnability/simplicity</i>). Please select applicable option.	Easy
2. Is the interface <i>user-friendly</i> , easy to learn with enough help documentation?	On the metadata screen the URL wouldn't accept '.csv', gave an error about ascii code. But URL on the Upload screen does allow '.csv' in URL. The Name and Description field in upload should be auto populated from the Metadata screen but editable.
3. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	3-4. Had to deal with ASCII error on URL, had to remove '.csv' to solve it. The validation rule may need to be changed.
4. Additional Comments	roscommon-rosco-opendata-sh123.csv

Evaluation 4: Dataset Linking

Assumed Scenario:

Note that on this version; only portal administrators or data uploaders can use this functionality because it is not yet available for ordinary data consumers. Therefore backend administrator's login detail is required. Subsequent version will be incorporated with front-end version of the tools for linking dataset and made available to data consumers. As a data provider, assume that you have the need to inform or enable data users that one or more datasets are related to one and another by whatever nature. Perhaps the latter dataset is an update to the former and would make a better meaning by using both together; or similarly, by understanding something about the latter version would improve the use into which the former or both datasets could be put. Linking datasets enables the system to recommend the linked dataset(s) as related datasets when one of them is searched out by a user because linking functionality treats datasets as related. Secondly, linking dataset enables both datasets to be integrated into one dataset (*future version of TET*)

Test Data:

- **Login**
 - User name: **test1**
 - Password: **test123** (*use your set of username & password if you were assigned another*)
- **URL:** <http://srvgal100.deri.ie:8081/>
- **Dataset:**
 - **Dublin City Libraries Accessibility Audit**
 - Dataset to be linked to Dublin City Libraries Accessibility: **Planning-register**
 - You may use any set of datasets name you find on your screen.

Actions and Expectations/Results/Screen displays:

Open the given URL and arrive at the TET home page. Then click on the login icon on the top right corner of the page. System should display the login form. Complete the login field with your admin data and you may tick-off the box for **Remember me** option before clicking on the **login** button. This operation should bring you to the following screen:

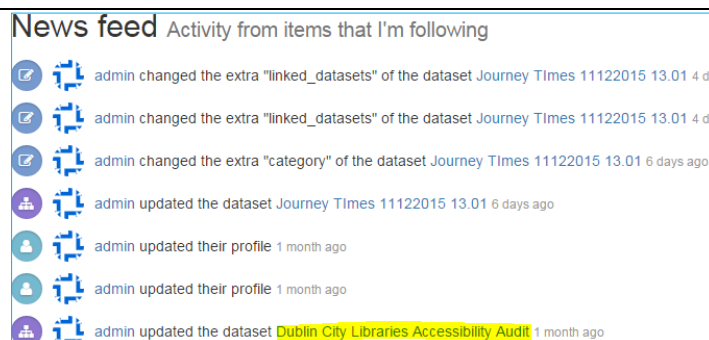


Figure 8: Dataset listing on the platform

On the top main menu, locate the “**Datasets**” button and click on it to bring you to the dataset listing (catalog) screen. On this window, select the dataset named above (i.e. **Dublin City Libraries Accessibility Audit**) to open it. (Note: you can select any dataset of your choice). On the screen that opens, locate the **Manage** dataset button; then click on it and the system should open the **Edit Metadata** table. Note that this figure which shows the field for **Related Datasets** is only a portion of the Edit Metadata table.

Figure 9: A portion of the “Edit metadata screen”

In the table, locate the **Related Datasets** field towards the bottom of the table, and as you begin to type in the name of the dataset (e.g. *Planning-register*) to be linked with the first dataset (e.g. *Dublin City Libraries Accessibility Audit*), you should get name(s) suggestion(s). Click on the relevant name to complete action. When done, click on the **Update Dataset** button at the bottom right side of the page. The system should save your work and return to the metadata table of the first dataset. On this screen observe that the value shown against the **Related Datasets** field is the filename of dataset you have just linked.

Category	Finance
Related Datasets	planning-register

Figure 10: A portion of metadata table showing the filename (of a given dataset) that has been linked

Table 6: TET User Survey

Questions	Short Responses
1. Linking datasets aims to make dataset more visible to users by exposing them. It can help integration datasets for better quality. Do you believe this functionality will enhance datasets <i>publicity & accessibility</i> and also improve <i>understandability, usability</i> and collectively improve <i>transparency</i> ?	It will advertise datasets to the end user.
2 How relevant is this feature considering the already available recommendation of dataset and the personalised search functions?	It is a good feature to have.
3. Comment on the <i>learnability/simplicity</i> of linking datasets?	Very easy.
4. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	5-6
5. Additional Comments	Can't link to related datasets uploaded by a different organisation. If I was an end user looking for a particular dataset type if would be useful for the DLR, SDCC, Fingal etc datasets could be a related dataset of the DCC dataset.

Evaluation 5: Unlink Previously Linked Datasets

Assumed Scenario:

It may become necessary to unlink previously linked datasets for various reasons which could be:

- The datasets are not quite related and they were earlier linked in error
- One or more of previously linked datasets have become obsolete and that using that version instead of a newer version would be misleading.
- One or more of the datasets need to be removed from public consumption for whatever reasons: e.g. for national security, data protection or that the dataset might have been inappropriately classified as open data earlier on.

Note that this operation can only be performed by a system administrator.

Test Data:

- URL: <http://vmdatagov01.deri.ie:8080/dataset>
- Dataset: Planning Register, Dublin City Libraries Accessibility Audit or Sculpture in Dublin City Council Parks

Actions and Expectations/Results/Screen displays:

Log in as an admin user as in the case of solution evaluation 4 above. From the metadata screen of a linked dataset, click on the **Manage** button and arrive at the **Edit Metadata** table. Go to the **Related Datasets** field and just click off the **x** button on the linked dataset to remove it. Click on the **Update Dataset** button to save changes.

* Category:

✕ Finance

Related Datasets:

✕ planning-register

* License:

Other (Open)

Figure 11: Linked dataset to be remove is highlighted

Table 7: TET User Survey

Questions	Responses
1. What importance does this functionality serve to uphold dataset qualities: source <i>integrity</i> , data <i>accuracy</i> and <i>relevance</i> ?	Can undo if the wrong dataset is selected.
2. Do you find the operation, <i>navigability</i> of the function easy and the <i>performance</i> smart or fast?	Easy to do.
3 Do you think this functionality will remove the frustration that users could encounter if datasets were removed from the portal without unlinking them from datasets they were previously linked to?	I don't know if a dataset is automatically unlinked if it is removed.
4. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	5
5. Additional Comments	

Conclusion

This Solution Evaluation User Guide is a quick guide that covers the instructions to test and evaluate a few of the TET functionalities so far developed and released as TET Alpha version of the ROUTE-TO-PA project. This guide does not cover solution components of SPOD; however, SPOD objectives are currently being developed simultaneously by another ROUTE-TO-PA consortium partner. The TET and SPOD components will be integrated in the TET/SPOD-enabled CKAN platform in due course according to the project schedule. Your cooperation is highly solicited regarding the completion of the questions in short survey tables attached to each evaluation exercise. These questions are intended to supply the metrics that measures the extent to which TET functions meet user needs in terms of qualities of the criteria that support data transparency based on transparency construct framework offered by Cappelli (2009). The survey responses are also intended to measure the levels of satisfaction users will derive from TET tools when the product goes live.

References

Cappelli, C. (2009). An approach for Business Processes Transparency Using Aspects, Doctoral Thesis, Departamento de Informática, PUC-Rio, Ago. 2009 (in Portuguese).

Testing Instructions for Test as an Open Data Platform User

Solution Evaluation Instructions for TET Components (Alpha Release)

Evaluation 1: Check Metadata Quality and Metadata Completeness Rating

Assumed Scenario:

You are a user in need of a specific dataset for planning purposes. You have no time for repeating the job of planning and cannot afford to have a data-driven plan that is not reliable or does not meet the needs of the users of the plan. You want to check on the metadata quality of some datasets of interest to you (one by one) to see if and all the datasets of interest meet the metadata quality specification for the use into which you intend to put them. For example, knowing the source and update frequency can raise or reduce reliability of the dataset, the metadata completeness rating can inform you of the use into which the dataset is capable of being put e.g. linking or integrating a couple of datasets together before analysing them to gain better insight necessary to support planning purposes.

Test Data:

- URL: <http://vmdatagov01.deri.ie:8080/dataset>
- Dataset name: **Journey Times 11122015 13.01**

Actions and Expectations/Results/Screen displays:

Type the URL on your browser bar and press enter to open the dataset listing page of TET/SPOD enhanced CKAN Platform

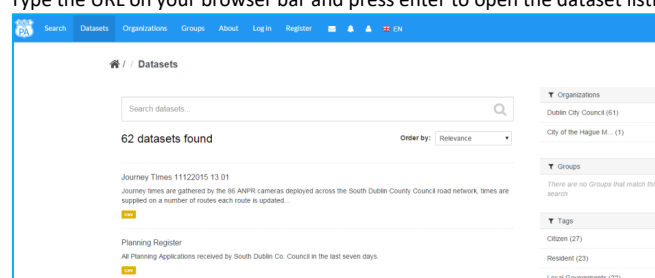


Figure 1: Dataset listing screen of TET/SPOD enhance CKAN Platform

You should see on your screen: a number of available datasets (**xx datasets found**) and, **Order by:** (drop-down menu) with sort tags: **Relevance, Name Ascending, Name Descending and Last modified**. Click on the dataset name, **“Journey Times 11122015 13.01”** (or any other dataset name) on the page. The system should display the metadata screen similar to figure below.

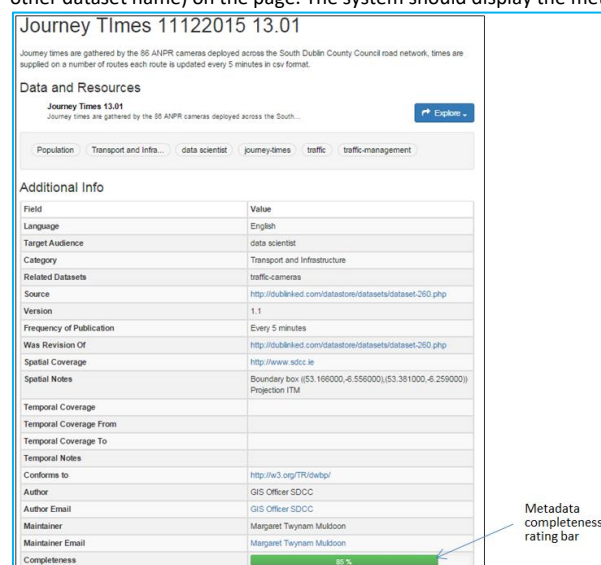


Figure 2: Metadata screen for a dataset

Observe the metadata fields on the left of the table and their corresponding values supplied by the data provider on the right. The metadata completeness rating is displayed at the bottom of the table as a horizontal bar and it measures the completeness (in percentage) of the metadata supplied. The more values of metadata supplied for the fields, the higher the quality of the metadata, the longer is the bar and the higher is the percentage rating displayed.

Table 3: TET User Survey

Questions	Responses
1. Metadata helps in searching out the dataset and in ensuring users of the appropriateness of the dataset for their uses. Does this functionality seem good enough to improve <i>accessibility</i> and <i>usability</i> of datasets?	Yes
2. Assume that metadata provided for a dataset correctly describes the nature of the dataset. In your opinion, will this situation improve retrieval of relevant datasets to match user searches (improve data <i>searchability</i>)?	Yes
3. By compelling data suppliers to complete metadata form for all datasets they supply, open data may become more <i>available</i> and <i>accessible</i> to potential consumers. Do you agree or disagree and why?	I agree. Because of public accessibility.

4. Will the possibility for you to see the metadata of the dataset beforehand affect your decision to use or not to use the dataset and also save you time for searching for the relevant datasets (improved platform <i>experience & usability</i>)?	Yes
5. What value does this system capability add to the existing open data platforms?	Meta data quality and completeness
6. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	10
7. Additional Comments	Overall idea design good. Need some issue fixing. Such as – after clicking the links provided in metadata fields always goes in different links. (Spatial Coverage) Time (5.06pm to 5.49pm)

Evaluation 2: Profile-based Personalization (user, not logged in)

Assumed Scenario:

As a user, you are not interested in creating a profile on the platform or you do not want to log in at this moment. However, you are interested in searching for datasets that meet a given set of criteria; perhaps, different from what were earlier saved on your profile account. You provide the system with a defined set of criteria this time by completing a search query form and clicking on the search button afterwards. In this scenario, the search criteria are not saved on the system and can only be repeated through the same process as before. On clicking the **search** button, you should retrieve from the system the available datasets that match the criteria you provided.

Test Data:

- URL: <http://vmdatagov01.deri.ie:8080/>
- Dataset: Data Scientist or any other datasets in the *Role* category, or in any other category on the screen.

Actions and Expectations/Results/Screen displays:

Type the given URL on your browser; the system should open the page below and by default should display datasets under **Role** (*I am...*). Other available data categories include **Datasets by Category** and **Datasets Recommended for me**

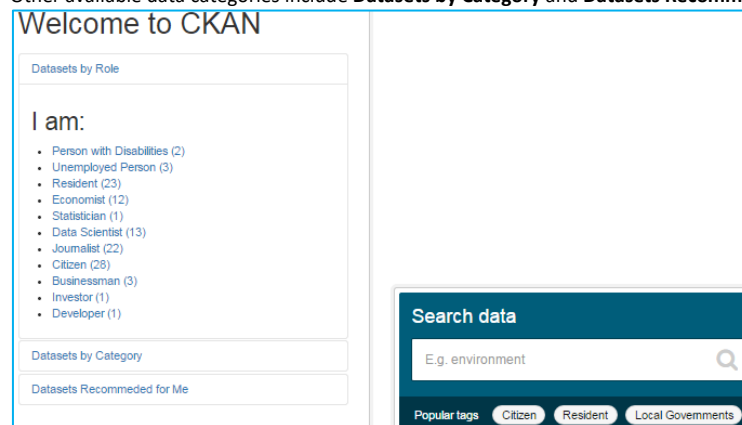


Figure 3: Dataset display by categories to improve user personalised searches

Click on the dataset name specified above (or on another name of your choice from the list) and system should display the screen below or something similar: You can select any dataset of interest to you on this screen by clicking on the filename.

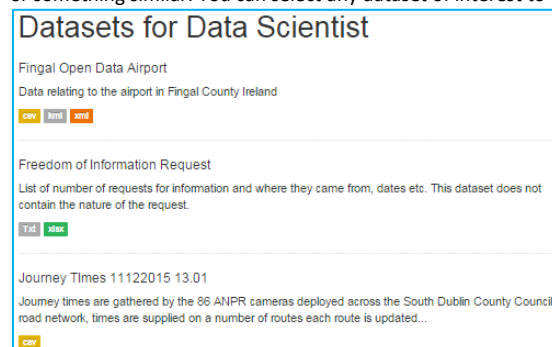


Figure 4: Dataset screen of "Data Scientist" category

Alternatively, you may personalise your search by using the search box provided on the first screen after the URL has been entered on the browser. First, click on the **back** button on the left end of the browser bar. Type in any search word/tag e.g. "**Economist**" and click on the search icon (**Q shape**) or press **enter** key. The system should return the number of datasets matching the search tag you entered and should display the filenames. If the files found are more than can be displayed on one screen, the system will display them in multiple screens ('pages'). Scroll down the screen to view the multiple screens numbers by selecting the numbers one by one to view the datasets filenames.

Table 4: TET User Survey

Questions	Responses
1. Does this functionality appear good enough to improve data <i>accessibility</i> to the users and more importantly searching out <i>relevant</i> datasets through responsive searching based on pre-defined search criteria enhanced with filtering tags?	yes
2. How <i>relevant</i> and <i>reliable</i> are the outcomes produced by using this feature?	Good, but numbers of dataset are less. So need more dataset to evaluate this.

3. Comment on the <i>interface friendliness</i> and <i>learnability</i> qualities of this functionality	There should include the autosuggestions while searching.
4. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	5
5. Additional Comments	The UI has changed. It was not expected according the description. But search is working fine. Time: 12.11 pm to 12.27pm

Evaluation 3: Profile-based Personalization (User not logged in) - Alternative

Assumed Scenario:

This is an alternative way to achieve **Personalized Search** without logging into the TET subsystem. The scenario is the same as in solution evaluation 2, but this time the user needs to use the iconised quick-links to get to the categories of datasets stored under various names in the datasets catalogue.

Test Data:

- URL: <http://srvgal100.deri.ie:8081/>
- Dataset: Person with Disability

Actions and Expectations/Results/Screen displays:

Enter the URL on your browser and press **Enter**. The TET home page opens showing the Welcome Message. Scroll to the bottom of the screen and observe various icons under the heading “**What are you looking for?**” Each icon represents a category of datasets and showing the number of datasets currently stored under the category.

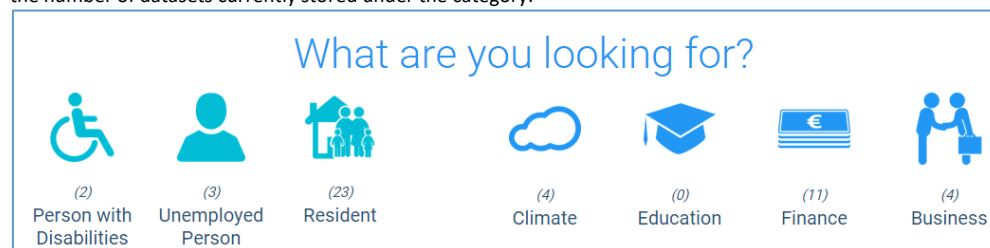


Figure 5: Iconised dataset categories

Click on the dataset icon of any dataset category above (e.g. iconised by the wheelchair) or select any one of them which is of interest to you. The system should display the number of datasets found in that category you selected and the datasets filenames.

Table 5: TET User Survey

Questions	Responses
1. Does this functionality appear good enough and does it improve <i>performance</i> (speed, efficiency) and <i>simplicity</i> of searching for <i>relevant</i> datasets?	Friendly and simple UI and Performance good
2. How about the <i>accuracy</i> of fetching the <i>relevant</i> datasets on the some of the categories tested by you. (<i>Note: relevance of searches depends on the accuracy of the metadata and tags used by the supplier in describing the dataset</i>)	Found relevant dataset
3. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 10 = excellent satisfaction)	10
4. Additional Comments	Time: 12.27pm to 12.33

Evaluation 4: Personalized Search (User is logged in)

Assumed Scenario:

Assume again in this scenario, you're a regular open data consumer (but not a provider), and you're perhaps a Data Journalist or a Data Scientist, Community Leader or Activist who is always interested in government open data. Alternatively, assume you are an Open Data System Developer or any other data consumer in your community. You have registered your profile and data interest on the ROUTE-TO-PA platform so that datasets retrieval for your frequent uses is done seamlessly and more accurately.

Test Data:

- URL: <http://vmdatagov01.deri.ie:8080/user/register>
- Dataset: For User details as a consumer of datasets (a common portal user), use the following:
 - User name: **tetuser**
 - Password: ********* (*Select your password*)

Actions and Expectations/Results/Screen displays:

Go to the URL above and click on **Register** button. The system should display the form below:

Register for an Account

Username:

Full Name:

Role:

Category of Interest:

Twitter ID:

Email:

Password:

Confirm:

[Create Account](#)

Figure 6: User account registration form

In the form, type in the required details above and complete the rest of the fields as you wish. Note that compulsory (required) field must be completed else the system will not save your details. When done, click on **Create Account** button.

In the screen that displays, click on **Datasets** button on the top menu.

The system should display a list of datasets available on the platform repository and **Personalised Search** button for the user to filter the result based on his/her user profile data.

Search datasets...

Personalized Results

Order by: Relevance

62 datasets found

Journey Times 11122015 13.01
Journey times are gathered by the 86 ANPR cameras deployed across the South Dublin County Council road network, times are supplied on a number of routes each route is updated...

Planning Register
All Planning Applications received by South Dublin Co. Council in the last seven days.

Organizations
Dublin City Council (61)
City of the Hague M... (1)

Groups
There are no Groups that match your search

Tags
Citizen (27)
Resident (23)
Local Governments (22)

Figure 7: Dataset listing screen

Click on the **Personalized Search** button to display the result of datasets matching the user's profile and interests.

Search datasets...

All Results

Order by: Relevance

17 datasets found

Freedom of Information Request
List of number of requests for information and where they came from, dates etc. This dataset does not contain the nature of the request.

Dublin Economic Monitor
Quarterly Economic Indicators compiled by the four Dublin Local Authorities to monitor economic development and enterprise support

Organizations
Dublin City Council (17)

Groups
There are no Groups that match your search

Tags
Data Scientist (13)
Citizens (9)
Residents (8)
Local Government (8)
Local Governments (7)

Figure 8: Personalised search result screen

Observe that the **Personalized Search** button has changed to **All Results** button and the number of datasets displayed is reduced according to the matches found. By clicking on the **All Results** button, you can toggle between the total datasets displayed and the number of datasets matching the user profile. On the right side of the screen are tags and headings to filter the search results much further to narrow into more relevant search result.

Table 6: TET User Survey

Questions		Responses
1. Is the searching based on personalised profile account fast, seamless and user-friendly?		yes
2. How accurate are the results of the searches based on your profile details and interests which the system uses to carry out data searches?		All are related to my interest
3. How relevant are the results of the personalised searches to you?		Related
4. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)		5
5. Additional Comments	I can't select more than one interest or category. I think system should give users to select multiple are of interest while registration or after login. Time: 2.35 to 2.39pm	

Evaluation 5: Personalized Recommendations
Assumed Scenario:

This scenario is similar to the one in which the user needs to use personalised search functionality (Solution Evaluation 6) so that the system can generate the matching datasets based on the user preferred profile data. Additionally in the case of **Personalised Data Recommendations**, user enters more details such as *Datasets category of interest, email address, the Role of the user* e.g. Professional/occupational status (e.g. Data Scientist, Economist, Businessman, etc.); Employment status, Disability status, etc. In the following versions, post TET Alpha release, there will be possibility to add a choice of email alert for users to receive update to datasets or a fresh upload in their areas of interests.

Test Data:

- URL: <http://vmdatagov01.der.i.e:8080/>
- Dataset: Planning Register.

Actions and Expectations/Results/Screen displays:

Go to the URL specified above; ensure that you are not logged into the system yet. Then locate the button “**Datasets Recommended for me**” at the bottom of the box that is displayed.

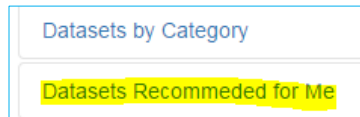


Figure 9: Display of Datasets recommended for user (User, not logged in)

This recommendations are system-generated based on popular searches by system users as default criteria.

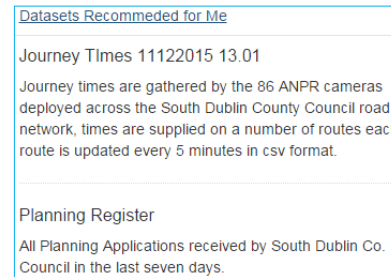


Figure 10: Display of filenames recommended for user (user not logged in)

Click on the button and observe the datasets displayed by the system. You may take a screen shot of the screen to compare with the next result. The datasets displayed in this occasion (not logged in) are based on the popular datasets searches by users visiting the platform. Now, log into the system by clicking on the **Login** button on the top menu and entering the required user (your) details. Once logged in, click or the **Search** button on left side of the top menu. That should bring you back to the screen on Now, click on the “**Datasets Recommended for Me**”, again (this time, while you are logged into TET systems), you should see the screen :

Display of “Datasets by Category” & “Datasets Recommended for me” is same for both user login and non logged instances

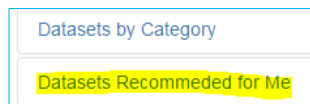


Figure 11: Display of Datasets recommended for user (user logged in)

This recommendations are system-generated based on User profile data and interests pre-registered on user account.

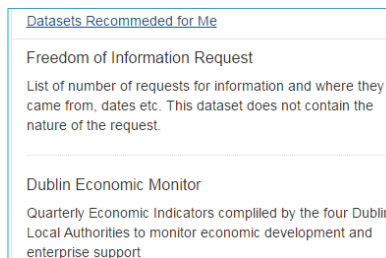


Figure 12: Display of filenames recommended for user (user, logged in)

Observe that the datasets displayed in the case while you are logged into the system are slightly different from those displayed while you were not logged into the system. The datasets recommended while logged in are personalised based on your profile data and the interests specified by you whereas, those recommended while not logged in are not based on your profile details instead on the popular searches made by the public.

Note: In the future versions of TET, the system will be configured so that users who want dataset recommendation would be able to choose if they would want data update or upload alert sent to them via specified emails or social media platform so that they receive quick notifications of the availability of a new datasets or updated versions for their immediate consumption.

Table 7: TET User Survey

Questions	Responses
1. Do you think this functionality is a relevant improvement over existing traditional open data platform?	
2. Do you think this feature will encourage open data users, facilitate open data use as well as promote <i>platform adoption</i> through seamless presentation of <i>relevant</i> datasets to users?	
3. Would you use this tool to request data recommendations?	
4. In your opinion, do you think personalised data recommendation functionality has the capability to improve <i>transparency</i> through data <i>accessibility</i> and <i>availability</i>	
5. Based on your profile details and interest that you registered on your account, was the recommended result <i>relevant</i> to your interest? In other words, how <i>accurate</i> were your recommended results in terms of <i>relevance</i> to your interest?	

6. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	
7. Additional Comments	Can not reproduced the scenario. There is no Datasets Recommended for Me button Time: 5.03 pm to 5.07

Evaluation 6: Recommend Related Datasets

Assumed Scenario:

This functionality follows on from the other ones already treated above. It uses the dataset types already matching your criteria at any occasion (logged in or not logged in) to recommend further datasets that would possibly be of interest to you. In the case of logged in users, the related datasets would be those datasets that are similar in category or nature to the ones much more closely matching your profile data. On the other hand, when not logged in, the related datasets would be those datasets coming next (by relevance) after the most popular datasets being consumed in general by the user community has been listed on the top level of the recommendation list.

Test Data:

- **URL:** <http://vmdatagov01.deri.ie:8080/dataset>
- **Dataset:** Planning Register, Dublin City Libraries Accessibility Audit or Sculpture in Dublin City Council Parks

Actions and Expectations/Results/Screen displays:

Open the URL above; and you should find the screen below or a similar one displayed by the system:

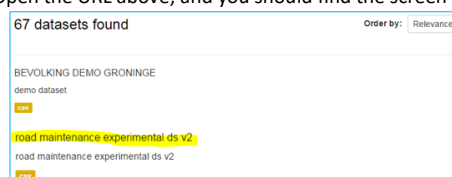


Figure 13: Datasets listing - top of screen

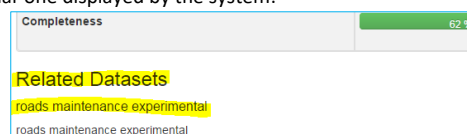


Figure 14: Related Datasets recommended view

Then click on the named dataset **“road maintenance experimental ds v2”** or any other dataset you can find on the page to view the metadata of the dataset. Scroll to the bottom of the metadata table, then locate the dataset(s) displayed under the heading **“Related Datasets”**. Click on the dataset name to view the metadata of the related dataset(s) displayed by the system. You can explore the dataset by using the drop down button **“Explore”** to give you access to actions – *preview* or *download* of dataset. Other search tags are available to trace to related datasets e.g. Public, Land-use, citizen, transport, etc. as they apply to the specific dataset you have chosen.

Note: There are some datasets that do not have related datasets to display when you click on the filename in this alpha version. Example: Customer service request (fixyourstreet.ie) has no related dataset yet on this version of TET portal.

Table 8: TET User Survey

Questions	Responses
1. Do you think automatic display of related datasets to the ones you have searched for will improve dataset visibility and accessibility?	
2 How relevant is this feature considering the already available recommendation of dataset and the personalised search functions?	
3. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	
4. Additional Comments	Can't reproduce according to the description Time: 5.07 to 5.14

Evaluation 7: Analyse and Visualise a Dataset to Determine its Quality

Assumed Scenario:

Majority of Open Data users are people with no technical background. One of the objectives of ROUTE-TO-PA is to make the understanding of open data easy for ordinary citizens without in-depth computer skills to navigate datasets analysis. Assume that you are a non-tech savvy individual or somebody who has not enough time to digest the data arrays presented on a tabular dataset. Your interest is to apply a tool that reproduces the datasets in graphical, colourful figures and in a manner easy and quick to understand so as to facilitate your ability to form opinion on the dataset meaning or use the idea gained from the dataset in further discussions. Using the tabular dataset, TET **Analysis and Visualisation** tools provide to analyse and visualise the dataset so that the user may comprehend its meaning. Example of the graphics include:

- A number of interactive graphics e.g. bar charts, pie charts, heat maps etc.;
- Easy drag-and-drop dataset components such as table column contents e.g. road classification, estimated area of works, final cost, etc., into a pivot table

Test Data:

- **URL:** <http://vmdatagov01.deri.ie:8080/dataset/litter-warden-inspections>
- **Dataset:** Litter Warden Inspections

Actions and Expectations/Results/Screen displays:

Click on the URL provided above to arrive at the dataset name **“Litter Warden Inspections”**. From this point, users can perform the following operations:

Preview the dataset in a table array or any other form:

From the point where you have the dataset name, navigate the platform using the buttons indicated: *Explore* -> *Preview* to arrive at the **Data Explorer** tab where you can view the dataset on a table and the names of the columns. Then click on the tab **Pivot table** default arrangements: **Table** as the default option for the types of visualisation graphics and **count** option for the types of representation of values in each column. You can also view the various names of the columns of the table on the left box of the pivot table. The middle box will display the values you drag in here on the Y- (or vertical) axis while the right box will display the values you drag into on the X- (or horizontal) axis.

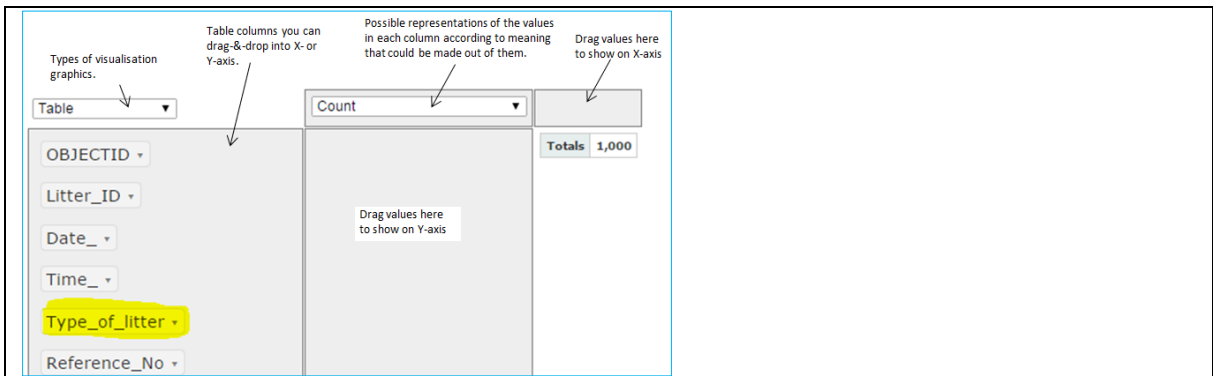


Figure 15: Pivot Table view

Determine the Quality of a dataset

Using the same details as in the table columns, click and drag **"Type_of_litter"** (highlighted) into the Y-axis. View the displayed numerical figures; then drag back **"Type_of_litter"** and drag in **"Type_of_litter 2"**. Also view the displayed numerical values as seen on the figure below (or something similar). Observe that there are some values of litters not named as **"Litter"**, **"Graffiti"** or **"Sign"** and their corresponding cells are blank (or missing values) in the column for all three types of litters.

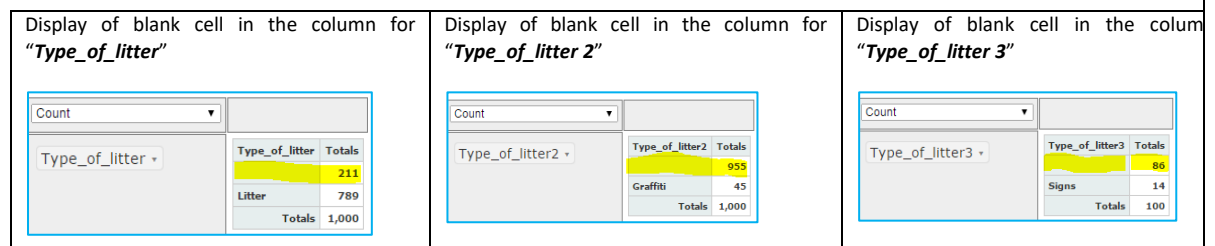


Figure 16: Tabular data quality check screenshots - 3 columns checks

There are three instances of column checks. First, the **"Type_of_litter"** column contains 211 cells without a description of the type of litter while 789 cells contain litter type marked **"litter"**. Second, the column **"Type_of_litter 2"** contains 955 cells with no description of litter type while 45 cells have litter type named **"Graffiti"**; and so on. Note that the blank cells reduce the quality of the dataset and larger the number of blank cells in a table array (as with no text or numerical value), the poorer the quality of the dataset because these blank cells represent missing values that reduces the information presented by the dataset. This functionality enables users to determine the quality of the dataset intended for use. You may drag as many columns of similar data types into the Y-axis box to see their qualities all at once :

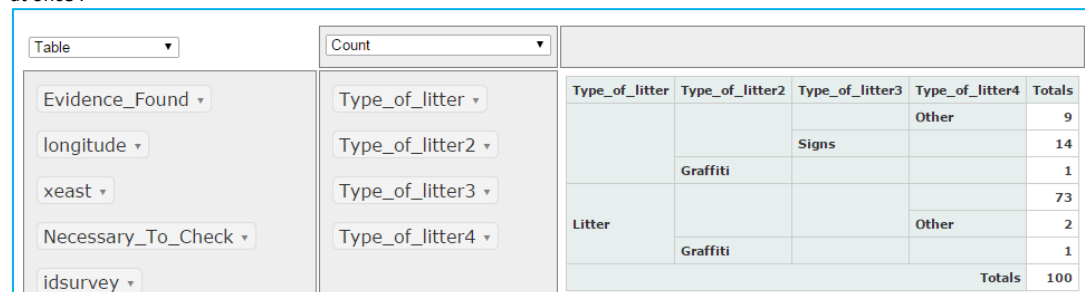


Figure 17: Four columns of similar data types dragged into the Y-axis to view their cell completeness

Furthermore, the above table can even be visualised using the columnar bar chart representation of their non-blank cells by selecting bar chart instead of table on the type of visualisation field. Please note that not all types of visual representations will be suitable for this purpose. Therefore, in each occasion of visualisation, a user will have to try out a couple of chart types to see which type best displays required visualisation.

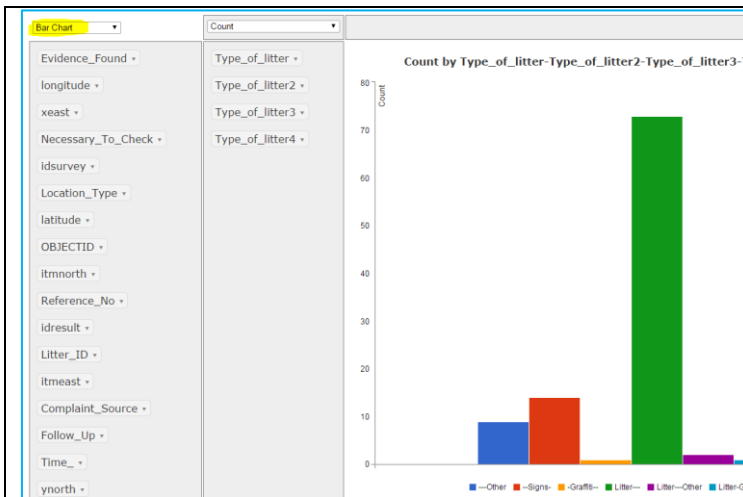


Figure 18: Dataset quality visualisation

Furthermore, Labnol.org (2014)¹² provides a help in deciding the right chart type to visualise dataset based on the need of the user as shown in the figure below.

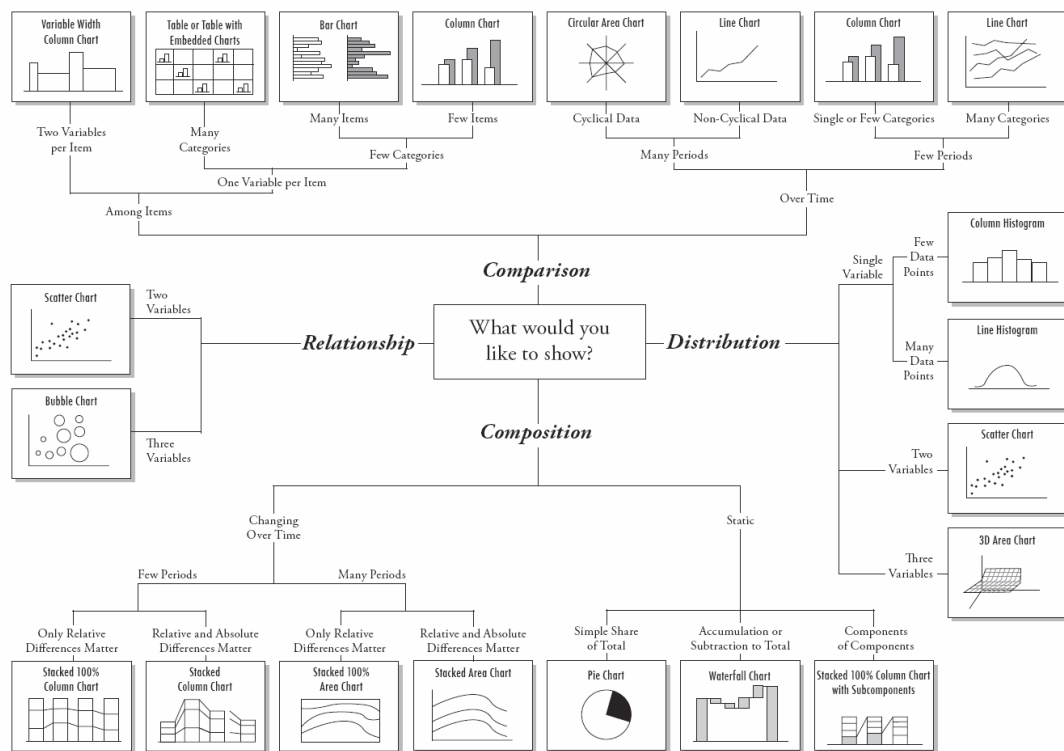


Figure 19: Help for deciding which chart type to use in data visualisation (Source: Labnol, 2014)

In the figures below, the types of litter is further broken down into location where they occur by adding the locations column to the Y- (or vertical) axis box.

¹² Labnol.org (2014) Choose the Right Chart Type for your Data. Source: <http://www.labnol.org/software/find-right-chart-type-for-your-data/6523/> Viewed: 19 Feb., 2016.

Table

Evidence_Found

longitude

Necessary_To_Check

idsurvey

latitude

xeast

OBJECTID

itmnorth

Reference_No

idresult

Litter_ID

itmeast

Complaint_Source

Count

Location_Type

Type_of_litter

Type_of_litter2

Type_of_litter3

Type_of_litter4

Location_Type	Type_of_litter	Type_of_litter2	Type_of_litter3	Type_of_litter4	Totals
Bring Bank	Litter	Graffiti	Signs	Other	3
					3
					24
Commercial	Litter	Graffiti	Signs	Other	1
					4
					9
Footpath	Litter	Graffiti	Signs	Other	24
					1
					1
Industrial	Litter	Graffiti	Signs	Other	1
					1
					13
Lay-By	Litter			Other	1
Open Space	Litter				5
Public Park	Litter				2
Residential	Litter	Graffiti	Signs	Other	3
					1
					1
Totals					100

Figure 20: Type of litter columns broken down into their locations to reveal blank and non-blank cells
The visualisation below reveals the count total of the non-blank cells per location of the litter types.

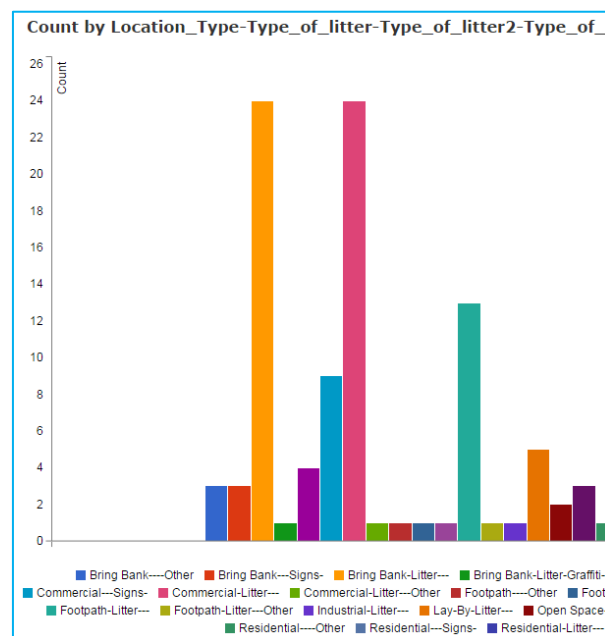


Figure 21: Visualisation of litters broken down into locations to reveal non-blanks cell per location

If you return your visualisation type from **Bar chart** to **Table** and drag the **location_type** over to the X- (horizontal) axis, you get another tabular representation as in figure below. Scroll left and right to view the table in full.

Count

Location_Type

Type_of_litter

Type_of_litter2

Type_of_litter3

Type_of_litter4

Type_of_litter	Type_of_litter2	Type_of_litter3	Type_of_litter4	Location_Type	Bring Bank	Commercial	Footpath	Industrial	Lay-By	Open Space	Public Park
Litter	Graffiti		Other		3	4	1				
			Signs		3	9	1				
							1				
			Other		24	24	13	1	5	2	3
						1	1	1			
	Graffiti				1						
Totals					31	38	17	1	5	2	

Figure 22: Another distribution of values for the Type_of_litter vs Location_type

Visualise this table again by changing **Table** to **Bar chart** in order to see a different display of litter types per location type.

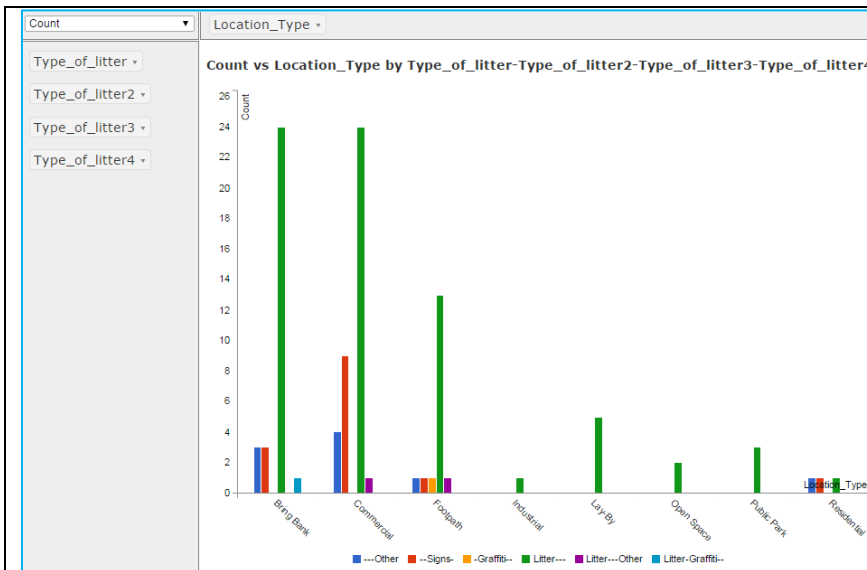


Figure 23: Count of Litter types per location type

This type of quick quality check can reveal the presence of missing values in the dataset as the empty cells are supposed to contain values; therefore, the idea can help you make the fit-for-purpose decision.

Table 9: TET User Survey - Evaluation 7: Determine the quality of a tabular dataset

Questions	Responses
1. Dataset quality is a major concern for all data consumers. In some data portals, it is not possible to verify the quality of a dataset before using it. Would you agree this TET solution presents beyond the state of the art technology with this TET functionality?	yes
2. Will this TET functionality enable you enjoy using open data better by being able to check on the quality of datasets before putting them into actual use?	yes
3. In your opinion, does this tool enhance data <i>informativeness</i> , <i>clarity</i> of meaning and even judge the <i>accuracy</i> and quality of the dataset?	yes
4. Will the result of simple analysis of the dataset components with this tool improve your decisions on suitability (usability , correctness and accuracy) you're your purposes?	yes
5. Comment on how this functionality will support you data usability?	I can decide the quality of the dataset by visualising it easily
6. Comment on the <i>intuitiveness</i> , <i>simplicity</i> , <i>learnability</i> and <i>smartness</i> of this functionality.	good
7. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	5
8. Additional Comments	Time: 5.20pm 5.30pm

Evaluation 8: Analyse and Visualise a Dataset to Understand its Meaning

Assumed Scenario:

The concept of Open Data and open data uses are both new to potential users in the society and majority of the would-be users cannot comprehend the dataset meaning very easily. To facilitate easy understanding of some of the datasets, TET presents tools to enable users easily and quickly reduce the datasets to more understandable figures without possessing much of technical skills. Assume that you are a resident of a community and your local government administrator has published open data on road works in the county. You are interested in this kind of dataset to understand how much or how well the government has been doing in the budget year in terms of road development in accordance to the needs of the people. So you want to view the dataset but also want to analyse and visualise the tabular dataset in a graphical manner to enable you see clearly the jobs so far done by your county administrators based on the dataset published by them.

Test Data:

- URL: <http://vmdatagov01.deri.ie:8080/dataset/roads-maintenance-annual-works-programme>
- Dataset: Roads Maintenance Annual Works Programme

Actions and Expectations/Results/Screen displays:

Click on the URL provided above to arrive at the dataset "Roads Maintenance Annual Works Programme".

Instance 1: Analyse and Visualise Dataset Components to Better Understand their Meaning:

On opening the dataset, follow the instructions below: Explore -> Preview -> Pivot table. Drag **Estimated Area of Work** column to Y-axis and **Final Cost** to X-axis. On the representation field, change **count** to **Sum** and on the field that appears, select **Estimated Area of Works**. Also change type of visualisation from **table** to **bar chart**.

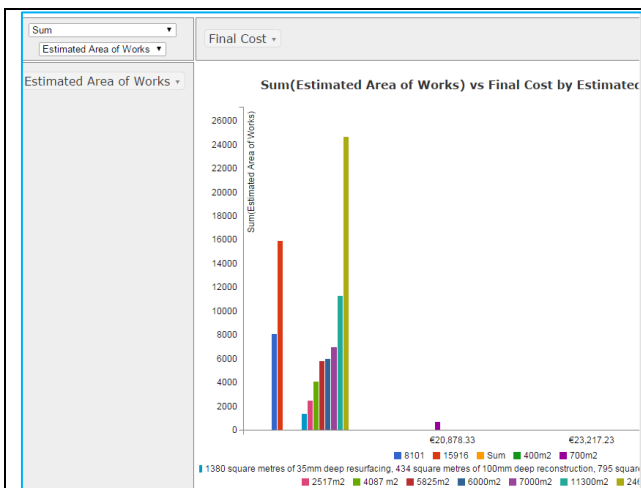


Figure 24: Graphical display of Estimated Area of Works vs the Final Cost Estimated

Note: It is possible to turn off some values of the columns in both the Y- and X-axes if they are meaningless or would distort the meaning of the visualised display. Click on the small arrow next to the **Estimated Area of Works** and click off the value that is undesirable. Figure below displays a stacked bar chart visualisation, but before you do that you can click on the small drop down arrow on the **Road Classification** and click off the first and last values (non-traffic) leaving the Traffic impact values on just to get a better representation.

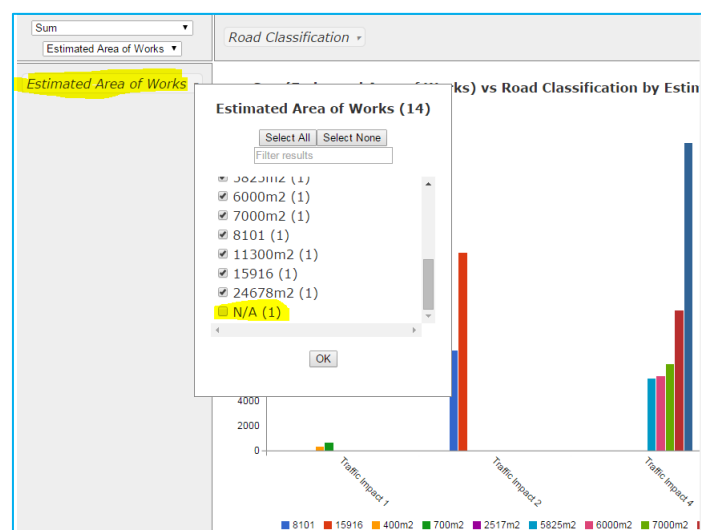


Figure 25: Select values to be represented on the graphic

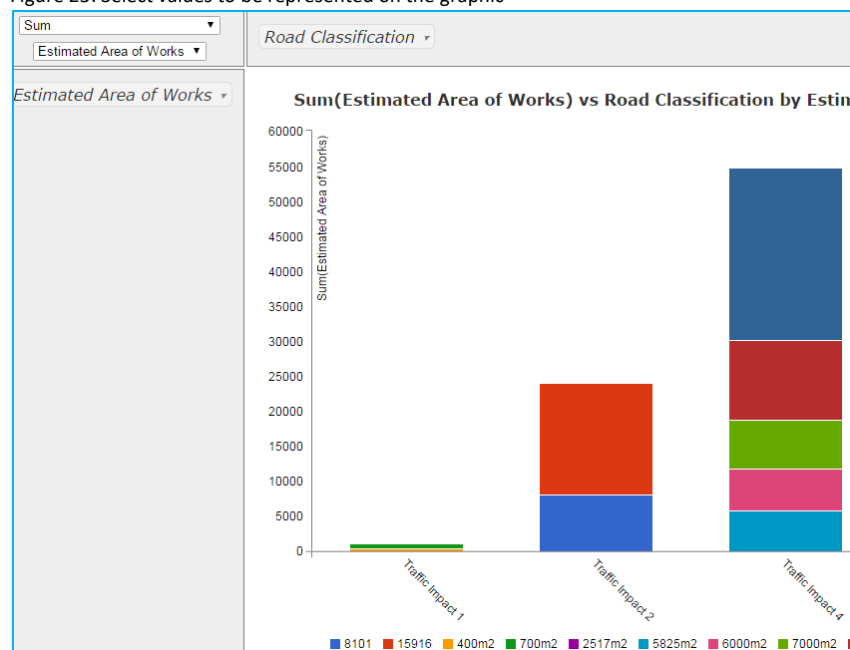


Figure 26: Stacked Bar Chart visualisation

Select various graphics to visualise a given component of a dataset: You can select any type of graphics to visualise the chosen values in your datasets. Example, visualise the **Electoral Area** of the dataset **Roads Maintenance Annual Works Programme** as seen in the figure below. Note that not all types of graphic visualisation would be suitable for the selected data type in the column:

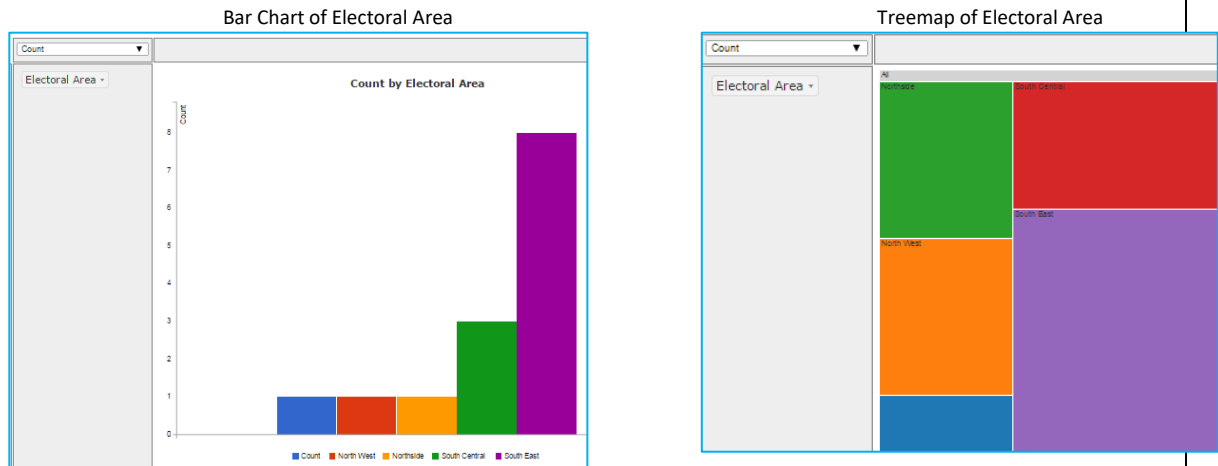


Figure 27: Various visualisation types for Electoral Area column of the dataset

In the Treemap visualisation of count of **Status of Works**, note that the coloured regions are labelled “complete” and “Planned” whereas the unlabelled region represents unclassified data entry in the **Electoral Area** column of the dataset while the labelled regions represent the count of the “complete”, “Planned”, etc.,.

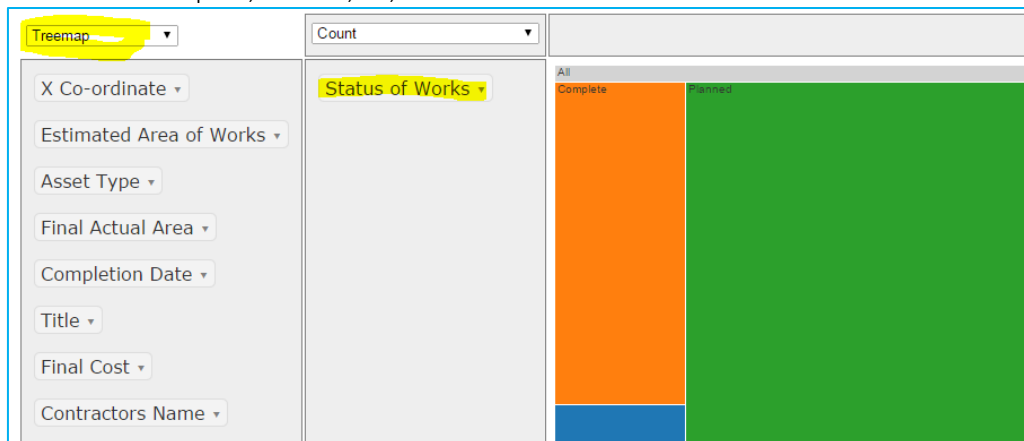


Figure 28: Treemap visualisation of Status of Works values [in a given dataset]

Table 9: TET User Survey - Evaluation 8a: Visualise a dataset to better understand its meaning

Questions	Responses
1. The visualisation of dataset quality attempts to further simplify the understanding of the quality of a given dataset. How easy is it to employ this tool?	Good enough
2. Comment on the learnability or operability of this tool	Overall good
3. Comment on the performance in terms of speed of this tool	Better
4. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	10
5. Additional Comments	Time : 5.30 to 5.36

Instance 2: Analyse a Dataset Components to Better Understand THEIR Meaning:

Assumed Scenario:

Based on instance 1 of the on-going functional evaluation section, you are interested in analysing the dataset to show all works and their status associated with the various work areas contained in the dataset “**Roads Maintenance Annual Works Programme**”.

Test Data:

- URL: <http://vmdatagov01.der.i.e:8080/dataset/roads-maintenance-annual-works-programme>
- Dataset: Roads Maintenance Annual Works Programme

Actions and Expectations/Results/Screen displays:

Open to the dataset “**Roads Maintenance Annual Works Programme**”, then follow the instruction: Explore -> Preview -> Pivot table. Select **Heatmap** in place of **Table**, drag **Electoral Area** and **Status of Work** from the columns into the Y-axis box. Then observe the display of count totals which represent the **Planned** and **Completed** work statuses for the various electoral areas . In the result, note that there is one (1) Electoral Area with **Planned Status of Works** but which has not been assigned to any **Electoral Area**. Similarly, observe that the **South East** Electoral Area, there are 7 areas with status of works already **planned** whereas one of the area in the **South East** has not been

assigned any status of work. The data de-composition results presented by this analysis table can help users understand more about the information contained in the dataset and the quality thereof and thus user can predict the value it will offer for their intended uses.

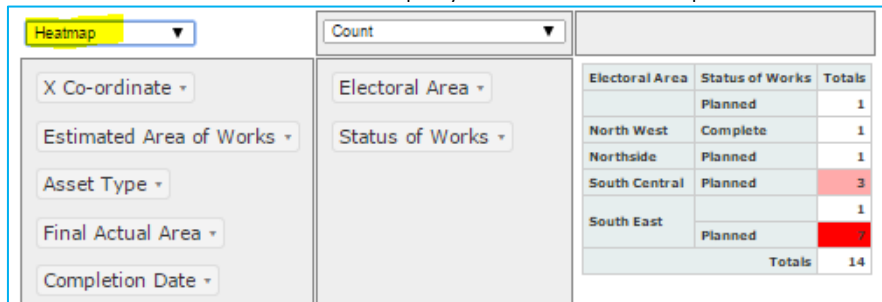


Figure 29: Status of Work Analysed by Electoral Areas

Instance 3: Analyse Dataset Components to Better Understand their meaning:

Assumed Scenario:

Using the same dataset as in the second instance above, however, assume in this case you are interested in analysing the dataset to show:

- The total invoice value for all vendors over a given period and
- The total invoice values by vendor names for first quarter of 2015.

Test Data

- URL: <http://vmdatagov01.derri.ie:8080/dataset/purchase-order-over-20000-2012-2015/resource/14f28fcc-14d5-4178-b36e-dcfc513865a4>
- Dataset name: Purchase order over 20000 (2012-2015)

Actions and Expectations/Results/Screen displays:

(i) Click on the provided URL to arrive at the dataset name given above and click on the **Pivot Table** button. Then change “Count” to “Sum”, and on the dropdown menu that appear, select “Invoice Total” to reveal the total value of the invoices from vendors in the county for the period 2012 to 2015.

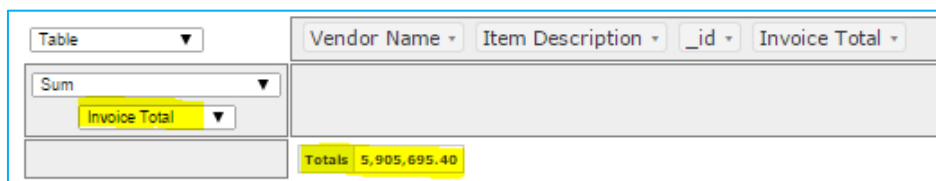


Figure 30: Analysis to reveal invoice total of purchase invoices for the period 2012 – 2015

(ii) To view the invoice totals by vendor names, you can proceed from the last point attained in (i). From the column names displayed in the figure below drag and drop **Vendor Name** into the Y-axis (vertical) box. Observe the displayed invoice total values by their vendor names in the table. To visualise this data array, just change the **Table** on the visualisation type dropdown menu to a suitable chart type as seen in Bar chart) Tree map:

(a) Bar Chart:

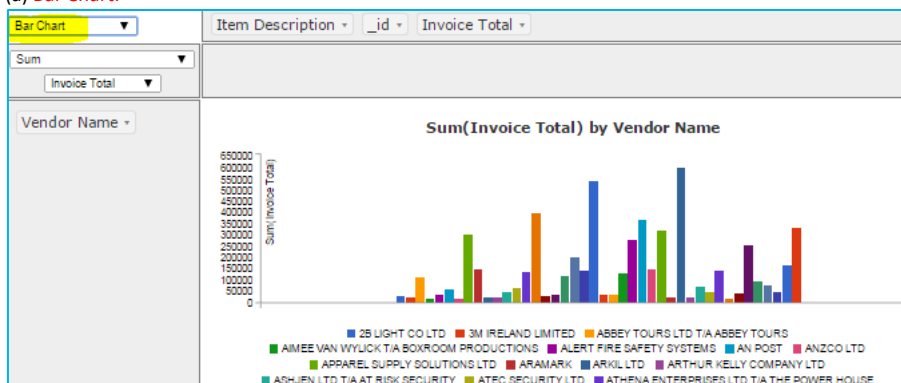


Figure 31: Bar chart view of the invoices totals by vendors

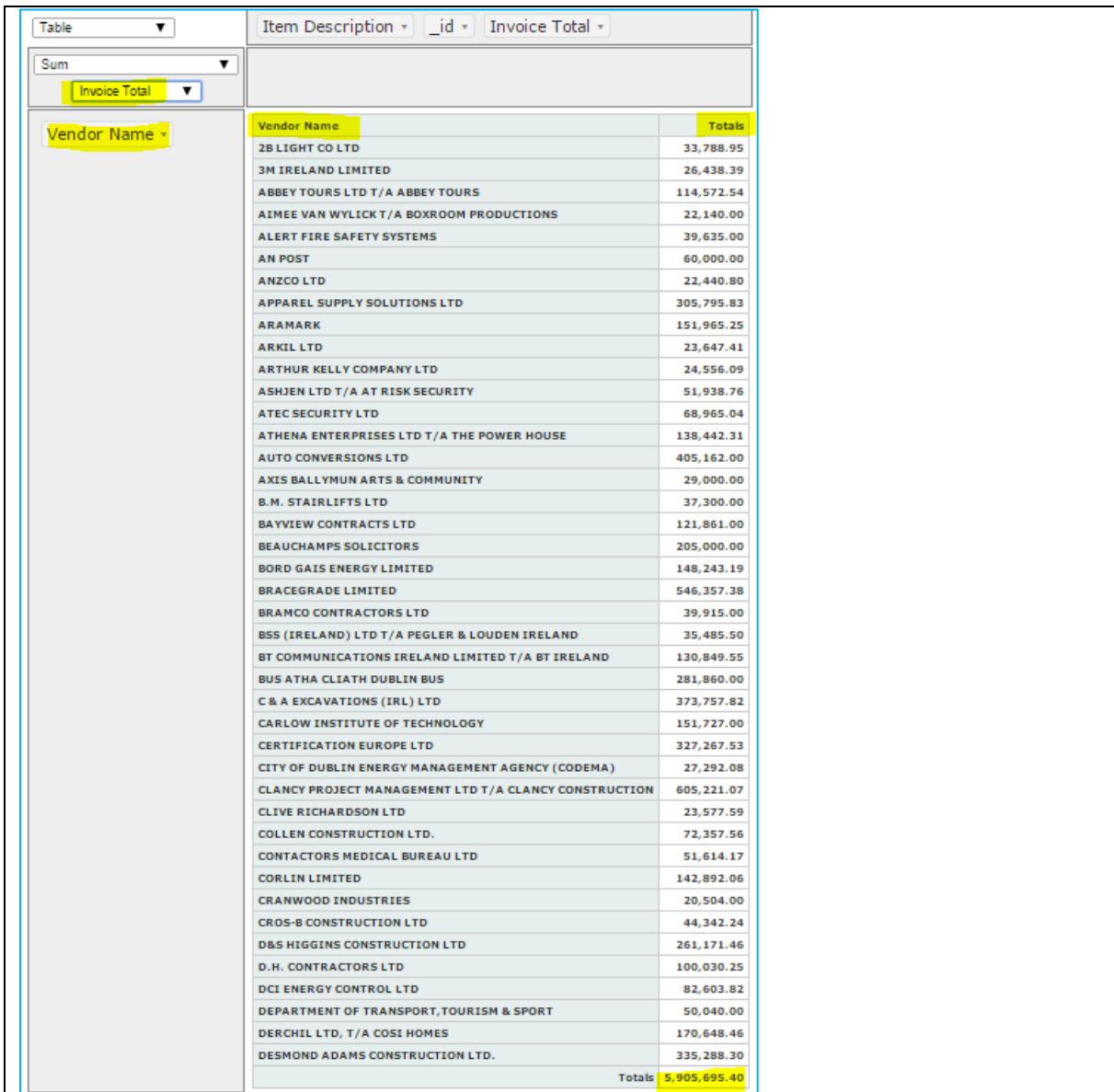


Figure 32: Invoice totals by vendors.

(b) Tree map:

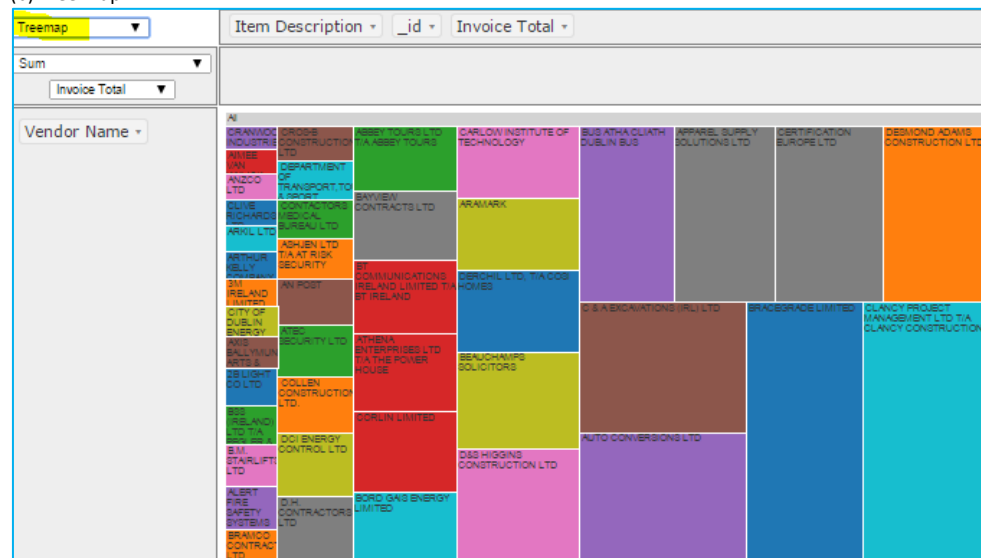


Figure 33: Treemap view of the invoices totals by vendors

Instance 4: Analyse Dataset Components to Better Understand its Meaning:

Assumed Scenario:

Similar to the instances (1 – 3) above, users can analyse datasets of interest to them in order to view the line graph and bar chart (or any other graphical representation) in order to improve dataset meaning to them. In this particular case, change your dataset to another filename.

Test Data:

- URL: <http://vmdata.gov01.deri.ie:8080/dataset/local-election-count-results-2014-balbriggan>
- Dataset name: Local Election Count Results 2014 – Balbriggan

Actions and Expectations/Results/Screen displays:

Click on the URL provided or copy and paste on your internet address browser to open up the dataset name given above; then follow the instruction: Explore -> Preview -> Pivot table. Drag **Party Name** into the vertical box as shown in the presented table. Note that there are 16 cells in the Party_Name column that

Table

Candidate_First_Pref_Votes

Status

Local_Electoral_Area_Number

Surname

Firstname

Count

Party_Name

Party_Name	Totals
	16
Fianna Fail	9
Fine Gael	18
Green Party - Comhaontas Glas	4
Non-Party	35
Sinn Fein	4
The Labour Party	14
Totals	100

Figure 34: Count of Party Name occurrences in the column cells

Change **Table** to **Bar chart** and observe the graphical display of the count of candidate per party name.

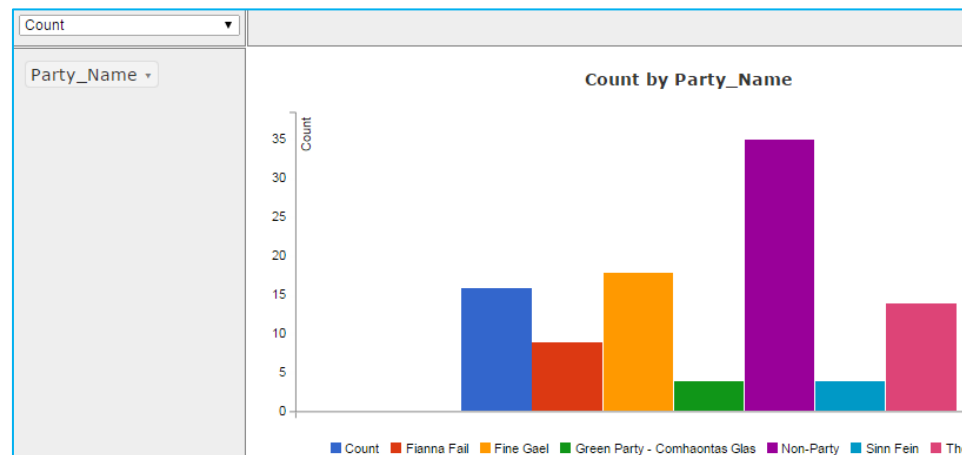


Figure 35: Visualisation of quality of the content of the Party_Name column in the dataset

Further analysis can reveal number of votes counted per count (1, 10, etc.) for each electoral candidate for Dublin City Council Local Election Results. To do this, open the dataset (Local Election Count Results 2014 – Balbriggan) as shown above and move to the Pivot table. In place of the **Count** select **Sum** and in the additional menu displayed, select **Total Votes**, then drag **Firstname** into the vertical box and **Count_Number** into the horizontal box. View the table that is displayed.

Table

Candidate_First_Pref_Votes

Status

Party_Name

Local_Electoral_Area_Number

Surname

Occurred_On_Count

Candidate_Id

_id

Total_Votes

Transfers

Sum

Total_Votes

Firstname

Count_Number

Firstname	1	10	11	12	13	Totals
BARRY	840.00	937.00	945.00	1,031.00		3,753.00
BRIAN	1,130.00	1,177.00	1,306.00	1,520.00	1,581.00	6,714.00
CATHAL	694.00	780.00	785.00	813.00		3,072.00
CIARAN	595.00	791.00	797.00	838.00	890.00	3,911.00
DARRELL	467.00	553.00				1,020.00
DAVID JAMES	1,327.00	1,521.00	1,541.00	1,577.00		5,966.00
DENISE	309.00					309.00
FRANK	363.00					363.00
GARRETT	286.00					286.00
GRÁINNE	868.00	1,023.00	1,080.00	1,095.00		4,066.00
J. P.	742.00	931.00	940.00	1,072.00	1,180.00	4,865.00
JANETTE	596.00	719.00	725.00			2,040.00
JEAN	125.00					125.00
JOE	804.00	927.00	942.00	983.00		3,656.00
KEN	2,073.00	1,819.00	1,819.00	1,819.00	1,819.00	9,349.00
MALACHY	1,575.00	1,721.00	1,762.00	1,776.00		6,834.00
MAURICE	139.00					139.00
NIALL	167.00					167.00
PEADAR	161.00					161.00
RICHARD	676.00	792.00	838.00	844.00	955.00	4,105.00
TERRY	571.00	873.00	890.00	895.00		3,229.00
TOM	450.00					450.00
TONY	1,404.00	1,663.00	1,800.00	1,814.00		6,681.00
Totals	16,362.00	16,227.00	16,170.00	16,077.00	6,425.00	71,261.00

Figure 36: Votes counted per count (1-13) for each electoral candidate of a local election results

Visualise the table graphically by selecting suitable visualisation type:

(a) Select Line chart:

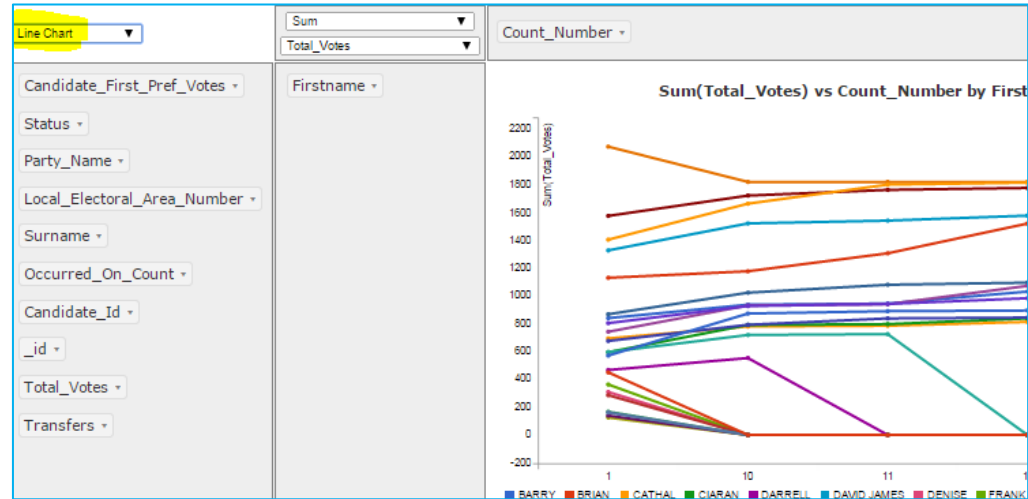


Figure 37: Visualisation of total votes per Candidate using line graph

(b) Select Stacked Bar chart:

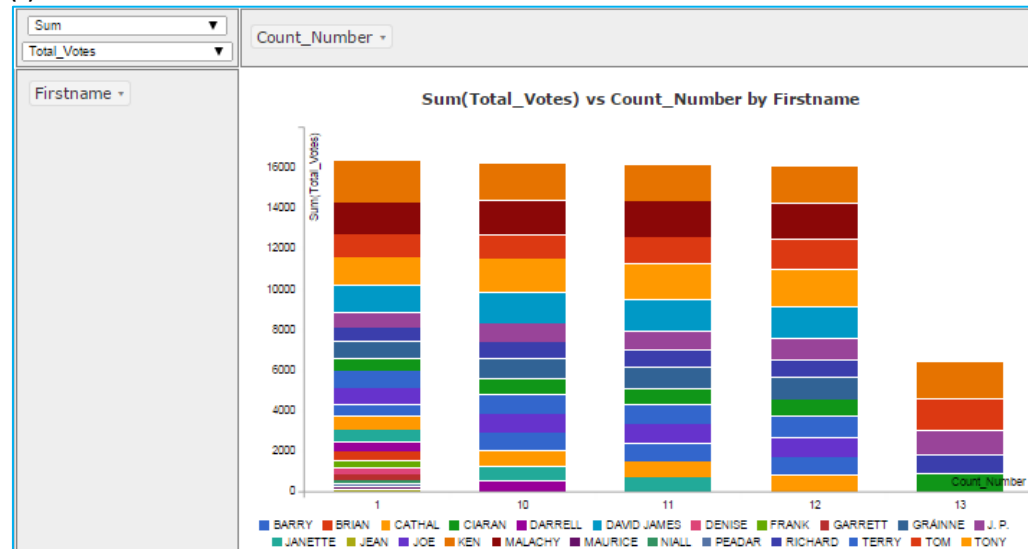


Figure 38: Visualisation of total votes per Candidate using Stacked Bar chart

(c) Select Bar chart:

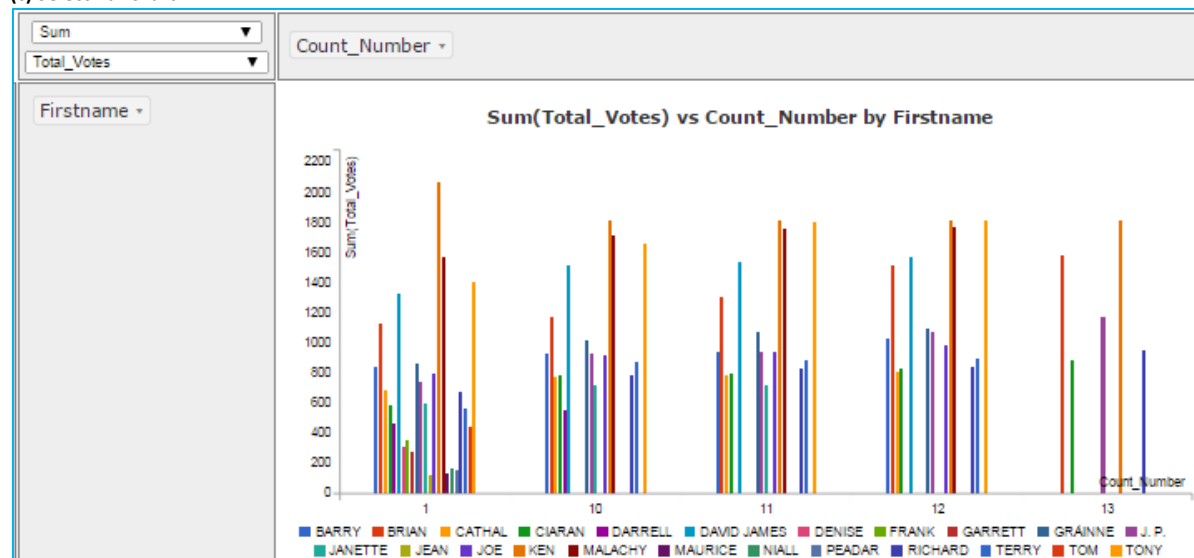


Figure 39: Visualisation of total votes per Candidate using Bar chart

Instance 5: Analyse Dataset Components to Better Understand its Meaning:

In the above visualisation instance, if you replace **Count_Number** with **Firstname** and **Surname** by dragging back the **Count_Number** and dragging in the **Firstname** and **Surname**; by selecting **Stacked Bar chart** as the visualisation chart type; then by leaving every other parameters as they are, you will have the visualisation graphics as below:

(a) Select Bar chart as visualisation chart type:

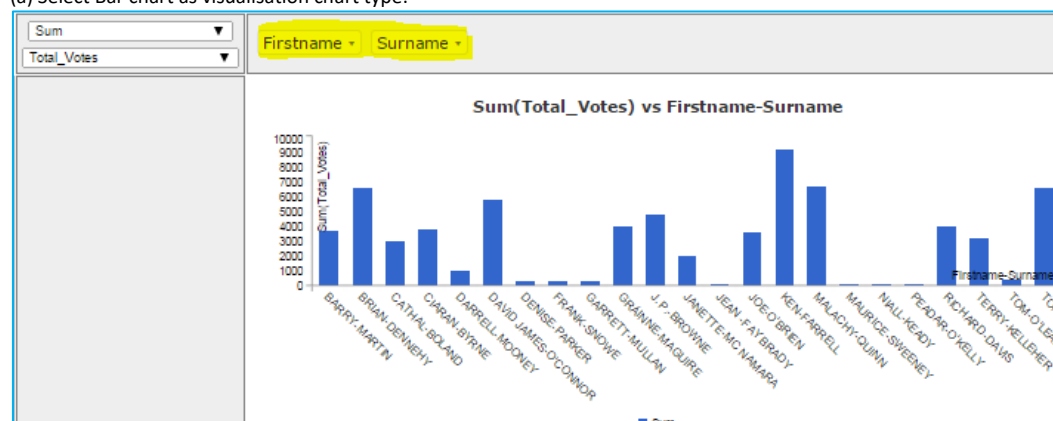


Figure 40: Bar chart visualisation of votes received by individual candidates

(b) Area Chart – Select Area Chart in place of Bar chart:

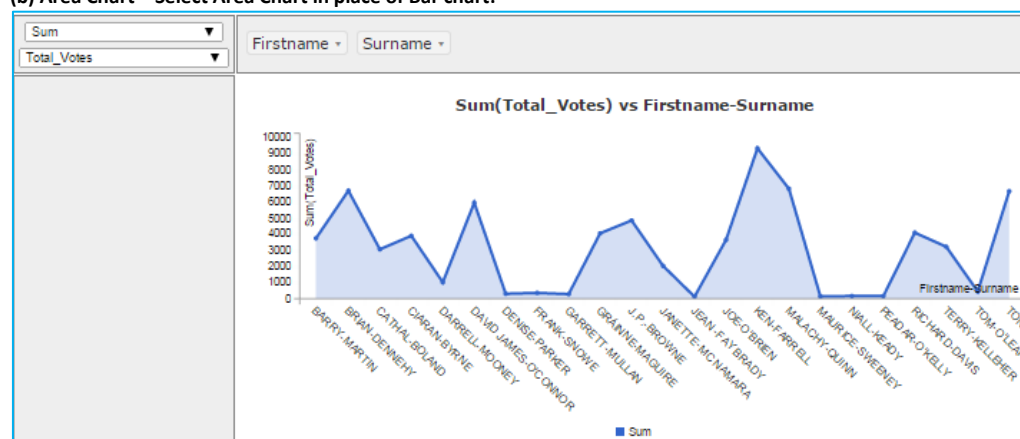


Figure 41: Area Chart visualisation of candidate votes

Table 11: TET User Survey - Evaluation 8b: Analyse and visualise a dataset to reveal its quality & meaning

Questions (This survey table covers evaluation instances 2 – 5 of Solution Evaluation 12 - Analyse and Visualise a Dataset to Understand its Meaning)		Responses
1. The visualisation of dataset quality attempts to further simplify the understanding of the quality of a given dataset. How easy do you find deciding on the right data visualisation graphic type?		Good enough
2. Comment on the learnability or operability of this tools		Good visualisation tool for learning datasets
3. Comment on the performance in terms of speed of this tools		better
4. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)		10
5. Additional Comments	3.8.3: bar chart not working time: 5.36pm to 5.55pm	

APPENDIX 1B: DETAILED RESULTS FROM PRE-USABILITY EVALUATION TESTING MAR, 2016

Pictures from Pre-Usability Evaluation Testing – Dublin 3 Mar 2016



Evaluation 1: Check Metadata Quality and Metadata Completeness Rating

1. Metadata helps in searching out the dataset and in ensuring users of the appropriateness of the dataset for their uses. Does this functionality seem good enough to improve accessibility and usability of datasets?

Yes	No	Other comments by relevance
4/8 (50%)	0	<p>1. Missing some metadata fields required in Dublinlinked (spatial projection, specific temporal fields – date creates, updated, released) also different terminology for similar fields may be confusing. Help prompts for data publisher. Also need to align with Dublinlinked 9 categories</p> <p>2. The naming of the file is the more important guide to the appropriate file for the end user</p> <p>3. It would be useful to have some help text for some of the headings. Eg what does 85% completeness actually mean</p>

2. Assume that metadata provided for a dataset correctly describes the nature of the dataset. In your opinion, will this situation improve retrieval of relevant datasets to match user searches (improve data searchability)?

Yes	No	Other comments by relevance
5/8 (63%)	0	<p>1. It depends if you can search based on every meta-data heading. For example 'Spatial Coverage' would bring back many datasets in an area that may not be at all related or relevant.</p> <p>2.</p>

3. By compelling data suppliers to complete metadata form for all datasets they supply, open data may become more available and accessible to potential consumers. Do you agree or disagree and why?

Yes	No	Other comments by relevance
6/8 (75%)	0	<p>1. Agree but need to give options that reflect the data – need option for temporal coverage that reflects dynamic data e.g. autofill based on update frequency</p> <p>2. Yes I agree. If date fields for dynamic data could be updated automatically, it would make more sense.</p>

		For static data, metadata templates should be provided to limit data entry for data publishers.
		3. Yes, so long as it does not become a hindrance to suppliers actually supplying data.
		4. Manner of implementation

5. What value does this system capability add to the existing open data platforms?

Relevant Responses:

1. More complete metadata;
2. Quick decision on relevance of datasets, help for searching.
3. Too much metadata may put off users.

6. Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction): **Average rating = 6**

Evaluation 2: Personalized Search (User with Admin Login)

1. Do you consider this functionality to be well personalised and user-friendly?

Yes	Reasonably	No	Other comments by relevance
2/8	3/8	1	<p>1. There should be more options for 'Role' to include e.g. GIS Officer, GIS/Data Analyst, etc. Perhaps list the roles alphabetically.</p> <p>2. Not clear how to get into personalised search. Need input into 'roles' as they do not reflect actual Dublinked user groups. Also areas of interest need to align with Dublinked categories.</p>

Evaluation 3: Add Dataset (Admin login)

1. In your opinion, the process of adding a dataset is: difficult, not so difficult; easy, very easy or can't properly rate the process (learnability/simplicity). Please select applicable option.

Difficult	Not so Difficult	Easy	Very Easy	Can't rate	Other comments by relevance
2/8	2/8	2/8	0	1	<p>1. There are lots of inputs that are not adequately explained. There are also some bugs. I got an internal server error at one stage.</p> <p>2. Terminology can be unclear. URLs not always available. Needs to align better with Dublinked metadata</p> <p>3. Not intuitive to a user needs guidance</p> <p>4. Rate manner of implementation: Average = 33%</p>

Summary of Solution Evaluation Survey - TET Alpha - End Users

Evaluation 1: check metadata quality and metadata completeness rating

1. Metadata helps in searching out the dataset and in ensuring users of the appropriateness of the dataset for their uses. Does this functionality seem good enough to improve *accessibility* and *usability* of datasets?

Yes	No	Other comments by relevance
3/3		<ul style="list-style-type: none"> It improves accessibility but some of the metadata headings may need explanation.

4. Will the possibility for you to see the metadata of the dataset beforehand affect your decision to use or not to use the dataset and also save you time for searching for the relevant datasets (improved platform *experience & usability*)?

Yes	No	Other comments by relevance
3/3		<ul style="list-style-type: none"> Seeing the metadata beforehand would affect a user's decision to pick a dataset.

5. What value does this system capability add to the existing open data platforms?

Yes	No	Other comments by relevance
		<ul style="list-style-type: none"> I think an end user would check a dataset to see if it is useful to them regardless if the metadata completeness was at 80-90% or not.

7. Additional comment

Yes	No	Other comments by relevance
		Overall idea design good. Need some issue fixing. Such as – after clicking the links provided in metadata fields always goes in different links. (Spatial Coverage)

		Time (5.06pm to 5.49pm)
--	--	-------------------------

Evaluation 2: Profile-based Personalization (user, not logged in)

1. Does this functionality appear good enough to improve data *accessibility* to the users and more importantly searching out *relevant* datasets through responsive searching based on pre-defined search criteria enhanced with filtering tags?

Yes	No	Other comments by relevance
1/3		I think there are too many options in for role. I would search for dataset category more so than role in case whomever uploaded the dataset didn't think to assign it to a role that I would have picked.

2. How *relevant* and *reliable* are the outcomes produced by using this feature?

Yes	No	Other comments by relevance
		<ul style="list-style-type: none"> Outcomes are relevant but too many roles to search through. Good, but numbers of dataset are less. So need more dataset to evaluate this.

3. Comment on the *interface friendliness* and *learnability* qualities of this functionality

Yes	No	Other comments by relevance
		<p>There should include the autosuggestions while searching. Updated by including the filter options: Relevance, Name and Date</p> <p>Results: 148 Relevance Name Date</p> <p>Luas_Network_2012_Stops_ITM</p> <p>Luas stops data The points represent the location of Luas stops at the centre of the stop platform and were created using Luas As-Built data. Stop names are in English and Irish. Note that Racecourse and Brennanstown stops are passive stops. These are future stops that are not currently open. The data is avail in ...</p> <p>Full Map Download Share Print Zoom Layers Data</p> <p>Population</p> <p>Contains information about population in Fingal. Number of people per Electoral Division from Censuses 2011, 2006, 2002, 1996, 1991, 1986, 1981</p> <p>Full Map Download Share Print Zoom Layers Data</p>

5. Additional Comments

Yes	No	Other comments by relevance
		<ul style="list-style-type: none"> Relevant if the data is input correctly at the start i.e. category The UI has changed. It was not expected according the description. But search is working fine, Time: 12.11 pm to 12.27pm

Evaluation 3: Profile-based Personalization (User not logged in) - Alternative

1. Does this functionality appear good enough and does it improve *performance* (speed, efficiency) and *simplicity* of searching for *relevant* datasets?

Yes	No	Other comments by relevance
		<ul style="list-style-type: none"> I wouldn't use this feature as I couldn't be sure that whomever uploaded the dataset selected the correct role type. As a 'resident' I might think that I would be interested in a particular dataset but whomever entered the metadata might not have thought it would be relevant to 'residents'. Friendly and simple UI and Performance good

2. How about the *accuracy* of fetching the *relevant* datasets on the some of the categories tested by you. (Note: *relevance of searches depends on the accuracy of the metadata and tags used by the supplier in describing the dataset*)

Yes	No	Other comments by relevance
		<ul style="list-style-type: none"> Residents had 23 datasets but unemployed person (3) and persons with disabilities (2). Unemployed and disabled are still residents and some might take offence that some datasets were listed a relevant to some residents but not to unemployed persons.

Evaluation 4: Personalized Search (User is logged in)

1. Is the searching based on personalised profile account fast, seamless and user-friendly?

Yes	No	Other comments by relevance
-----	----	-----------------------------

3/3	Yes, features works well
-----	--------------------------

2. How accurate are the results of the searches based on your profile details and interests which the system uses to carry out data searches?

Yes	No	Other comments by relevance
		<ul style="list-style-type: none"> All are related to my interest Seems correct based on the profile.

3. How relevant are the results of the personalised searches to you?

Yes	No	Other comments by relevance
		<ul style="list-style-type: none"> Would have to go through all datasets to be sure but the ones displayed seem relevant. → Provided for by including description of dataset, filter options and specific filter options for relevance of search: Relevance, Name and Date <p>Results: 148 Relevance Name Date </p> <p>Luas_Network_2012_Stops_ITM</p> <p>Luas stops data The points represent the location of Luas stops at the centre of the stop platform and were created using Luas As-Built data. Stop names are in English and Irish. Note that Racecourse and Brennanstown stops are passive stops. These are future stops that are not currently open. The data is avail in ...</p> <p> </p> <p>Population</p> <p>Contains information about population in Fingal. Number of people per Electoral Division from Censuses 2011, 2006, 2002, 1996, 1991, 1986, 1981</p> <p> </p>

5. Additional Comments

Yes	No	Other comments by relevance
		I can't select more that one interest or category. I think system should give users to select multiple are of interest while registration or after login

Evaluation 5: Personalized Recommendations

7. Additional Comments

Yes	No	Other comments by relevance
		Can not reproduced the scenario. There is no Datasets Recommended for Me button Time: 5.03 pm to 5.07

Evaluation 6: Recommend Related Datasets

4. Additional Comments

Yes	No	Other comments by relevance
		<ul style="list-style-type: none"> Can't find Actions and Expectations/Results/Screen displays Can't reproduce according to the description, Time: 5.07 to 5.14

Evaluation 7: Analyse and Visualise a Dataset to Determine its Quality

5. Comment on how this functionality will support you data usability?

Yes	No	Other comments by relevance
		I can decide the quality of the dataset by visualising it easily

Evaluation 8: Analyse and Visualise a Dataset to Understand its Meaning
(Instance 1: Analyse and Visualise Dataset Components to Better Understand their Meaning)

2. Comment on the learnability or operability of this tool

Yes	No	Other comments by relevance
		Good enough

Evaluation 8b: Analyse and Visualise a Dataset to Understand its Meaning (in 5 instance)
Survey question focused on 5 areas shown on the table below. Note that only 3 evaluators evaluated 8b. The other did not:

#	Focus Areas	Responses (one of the 3 respondents)	Responses (2 of the 3 respondents)
1	How easy do you find deciding on the right data visualisation graphic type?	Good enough	Blank
2	Comment on the learnability or operability of this tools	Good visualisation tool for learning datasets	Blank
3	Comment on the performance in terms of speed of this tools	better	Blank
4	Rate your satisfaction with the manner of implementation of this functionality? Choose from 1-10. (1 = no satisfaction, 5=good, 10 = excellent satisfaction)	10	Blank
5	Additional Comments	3.8.3: bar chart not working time: 5.36pm to 5.55pm	Blank

APPENDIX 2: USABILITY EVALUATION – INSTRUCTION MANUAL

Pictures from ROUTE-TO-PA Usability Workshop, Galway, 3 Aug., 2016



The scenario

Assume you belong to a small group of students living in Dublin and the group is interested in ways of reducing overall traffic flow challenges in Dublin while also thinking about ways to get older adults and disabled people more mobile in city life. Your group has been looking at relatively cheap and efficient 3-wheeled water proof scooters that take small parking spaces like bicycles. You are considering if making them available in addition to the Dublin Bikes might be a clever way of mobilizing more of the population and reducing overall traffic problems. You were amazed on seeing thousands of mopeds on city streets in Taipei and wondered about enterprise ideas in this space for Dublin. Your group has chosen to review available datasets in Dublinked to determine how feasible the idea of introducing the three-wheeled moped into Dublin traffic system might seem. Use the instruction provided below:

- To view the available datasets in Dublinked data portal via the Transparency Enhancing Toolset (TET) infrastructure (**search, filter, view** of datasets, charts, their **descriptions** and related Social Platform for Open Data (SPOD) **discussions** if any) – **TET activities**;
- From the SPOD platform, create your data artefacts and store them in your private room (**My space**) and create '**Agora**' (public discussion) room to collect feedback from the public on how to improve Dublin traffic – **SPOD activities**.
- **Import** and **export** data artefacts between TET and SPOD interfaces or move over from TET to SPOD to make data-supported comments on the SPOD forum – these are **integrated activities between TET and SPOD**.

TET usability tasks - Instructions

Search for A dataset on TET

The first step is to examine the "Dublin City Council budget and spending" data to understand the amount allocated to various sectors and the actual spending on each in the previous year (assumed).








1. Dataset name: Dublin City Council Spending and Revenue budgets.
2. Search keywords: 'Dublin spending budgets'
3. Where to search in this link to TET platform : <http://vmegov01.deri.ie>
4. Instructions:

Type search keywords on the search bar of TET platform, press enter.

Observe & perceive: Take a look at the screen contents, noting down your observation and perceptions.








Data Resource Views (1)

At this stage you have refined your search result, and you can see the dataset *Dublin City Council Spending and Revenue budgets* on the list.

1. Instructions: Click on the *Dublin City Council Spending and Revenue budgets* table icon  to open it.
2. Observe and perceive:
 - I. Observe that "*Traffic Improvement Measures*" has some incomplete values against.
 - II. Note down your observations and perceptions.
3. Copy the API link under the 'Metadata' section (just click on the copy icon .
4. Go to Dublin SPOD: click the link <http://dublin-spod.routetopa.eu> or type it in your browser bar.
5. Click on 'My space' then on the plus  icon to create a data visualisation,
6. Then select  to create a graph and on the 'Selected url' line at the bottom of the page, paste the API Link you copied from TET window.
7. Click on the right arrow icon () on the top right corner of the screen.
8. Under the 'FIELDS', select all parameter except 'Code' and click on () again.
9. Click on the a Vertical bar chart:
 - I. Complete the Title and Description fields.
 - II. For X-axis: select 'ExpenditureType'.
 - III. For the first & second Y-axes: select: 'Adopted by Council' and 'Estimated by Manager' respectively.
 - IV. Click on the 'ADD' button on the top right side of the screen to save the graph to your 'My space'.
10. Inside 'My space':
 - I. Click on the icon  to display the graph created.
 - II. Note down your observations and perceptions.






Data Resource Views (2)


Search for a second dataset on TET.

1. Dataset name: *Modes of Travel in Dublin Region*.
2. Search keywords: 'Modes Travel'
3. Instructions: Click on the *Modes of Travel in Dublin Region* chart icon  to open it.
4. Observe and perceive:
 - I. Observe the most popular and least popular modes of travel.
 - II. Note down your observations and perceptions.
5. Copy the API link under the 'Metadata' section (just click on the copy icon .
6. Go to Dublin SPOD: Go to the Dublin SPOD window.
7. Click on 'My space' then on the plus  icon to create a data visualisation.
8. Then select  to create a graph and on the 'Selected url' line at the bottom of the page, paste the API Link you copied from TET window.
9. Click on the right arrow icon () on the top right corner of the screen.
10. Under the 'FIELDS', select all parameters and click on () again.
11. Click on the Line chart:
 - I. Complete the Title and Description fields.
 - II. for X-axis: select 'Mode'.
 - III. For the Y-axes: select: 'Dublin City'; up to 'South Dublin' one for each Y-axis.
 - IV. Click on the 'ADD' button on the top right side of the screen to save the graph to your 'My space'.
12. Inside 'My space':
 - I. Click on the icon  to display the graph created.
 - II. Note down your observations and perceptions.

Data Resource Views (3)






Search for a third dataset on TET.

1. Dataset name: *Population*.
2. Search keywords: 'Population'.
3. Instructions: Click on the *Population* chart icon  to open it.
4. Observe and perceive:
 - I. Observe the high population density areas.
 - II. Note down your observations and perceptions.
5. Copy the API link under the 'Metadata' section (just click on the copy icon .
6. Go to Dublin SPOD window.
7. Click on 'My space' then on the plus  icon to create a data visualisation.
8. Then select  to create a graph and on the 'Selected url' line at the bottom of the page, paste the API Link you copied from TET window.
9. Click on the right arrow icon () on the top right corner of the screen.

10. Under the 'FIELDS', select all parameter except 'CSO_Code' and click on (→) again.
11. Click on the a Vertical bar chart:
 - I. Complete the Title and Description fields
 - II. for X-axis: select 'Name'
 - III. For the first & second Y-axes: select: '1991' and
 - IV. Repeat for the remaining years – each on the Y-axis.
 - V. Click on the 'ADD' button on the top right side of the screen to save the graph to your 'My space'.
12. Inside 'My space'
 - I. click on the icon  to display the graph created.
 - II. Note down your observations and perceptions.






Data Resource Views (4)

Search for a fourth dataset on TET.

1. Dataset name: *Luas Network 2012 Stops ITM*.
2. Search keywords: '*Luas Networks*'.
1. Instructions: Click on the *Luas Network 2012 Stops ITM* table icon  to open it.
2. Copy the API link under the 'Metadata' section (just click on the copy icon ).
3. Go to Dublin SPOD: click the link <http://dublin-spod.routetopa.eu> or type on type on your browser bar.
4. Click on 'My space' then on the plus  icon to create a data visualisation.
5. Then select  to create a graph and on the 'Selected url' line at the bottom of the page, paste the API Link you copied from TET window.
6. Click on the right arrow icon (→) on the top right corner of the screen.
7. Under the 'FIELDS', select all parameter and click on (→) again.
8. Click on the 'Map' chart:
 - I. Complete the Title and Description fields.
 - II. For Latitude: select 'Latitude'.
 - III. For Longitude: select 'Longitude'.
 - IV. For each of the balloon content: select 'Stop English' and 'Stop Irish' respectively.
 - V. Click on the 'ADD' button on the top right side of the screen to save the graph to your 'My space'.
9. Inside 'My space'
 - I. Click on the icon  to display the graph created.
 - II. Observe the Luas stops which are potential stops for Moped parks.
 - III. Note down your observations and perceptions.

Data Resource Views (5)

Search for a fifth dataset on TET.

1. Dataset name: *Dublinbikes*.
2. Search keywords: '*Dublin Bikes*'.
3. Instructions: Click on the *dublinbikes* chart icon  to open it.
4. Copy the API link under the 'Metadata' section (just click on the copy icon ).
5. Go to Dublin SPOD: click the link <http://dublin-spod.routetopa.eu> or type it into your browser bar.
6. Click on 'My space' then on the plus  icon to create a data visualisation.
7. Then select  to create a graph and on the 'Selected url' line at the bottom of the page, paste the API Link you copied from TET window.
8. Click on the right arrow icon (→) on the top right corner of the screen.
9. Under the 'FIELDS', select all parameter and click on (→) again.
10. Click on the 'Map' chart:
 - I. Complete the Title and description fields.
 - II. For Latitude: select 'Latitude'.
 - III. For Longitude: select 'Longitude'.
 - IV. For each of the balloon content: select 'Name' and 'Stn No' respectively.
 - V. Click on the 'ADD' button on the top right side of the screen to save the graph to your 'My space'.
11. Inside 'My space'
 - I. Click on the icon  to display the graph created.
 - II. Observe the distribution of the Dublinbike stations which could give clues for moped parks.
 - III. Note down your observations and perceptions.

Description of data resource

At this stage you have viewed many datasets on the TET platform and visualised them also both in TET and SPOD as seen in the activities in sub section 1.2 to subsection 1.6. You were asked to focus on tables, charts & descriptions in each case including focus on the scenarios to improve Dublin traffic with Mopeds. Perhaps you also saw some previous user

comments (from the SPOD forum) about any of the datasets you viewed in sub section 1.2. These will be necessary to create forums for discussions on the idea development to improve Dublin commuting.

SPOD evaluation tasks - instructions

Social discussions on data resource



At this stage you have viewed many datasets on the TET platform and visualised them also both in TET and SPOD as seen in the activities in sub section 1.2 to subsection 1.6. You were asked to **focus on tables, charts & descriptions in each case including focus on the scenarios to improve Dublin traffic with Mopeds**. Perhaps you also saw some previous user comments (from the SPOD forum) about any of the datasets you viewed in sub section 1.2. These will be necessary to create forums for discussions on the idea development to improve Dublin commuting.


You believe at the back of your mind that scooters are faster and better because you can wear your work clothes when using them. The new mopeds allow for protection from the rain as you have seen them before in Taiwan. Before proposing the introduction of the Moped tricycles to Dublin commuting life, you wanted to discuss the idea with the members of the Dublin student group each of whom is also researching this project. Your choice is to use the ROUTE-TO-PA SPOD platform and this demands the use of SPOD social platform for discussions. The moment you login to the SPOD interface (<http://dublin-spod.routetopa.eu/>), you find the top menu items – ‘what’s new’, ‘Members’, ‘My space’, ‘Agora’, and ‘CoCreation’. In this evaluation exercise, we will be concerned with the use of the last three functionalities.

CoCreation




This is the room where you can collaborate with group members to generate data-driven ideas and produce documents together; invite other people to join the collaborative work. Here you can invoke the commands to create the graphs you need to visualise a dataset. You can reach a range of datasets on the system to use for your idea generation or create a dataset right inside the room using tools available.

Note: Our focus is to use datasets available in the platform to create various visualisation graphs rather than create a dataset in the Co-Creation room.

After you logged into <http://dublin-spod.routetopa.eu>, click on the *Co-Creation* button on the top menu list, then click on the  icon at the button right of the page to add a new *Co-Creation data* room. Next, click on the  icon and follow on-screen instructions to complete the form fields:


1. Name: Name of the room.
2. Subject: Topic of discussion.
3. Description: Describe what you hope to discuss.
4. Period: Insert length of the discussion period.
5. Goal: What you hope to achieve from the Co-Creation collaboration and discussions.
6. Invitation text: Add member’s invitation message.
7. Members: Invite members to join Type names of members and select email address.
8. When done click on ‘Submit’ button.
9. Use the  icon to invite more people

Note:

1. All the data visualisation you did in the ‘My space’ can be done here in Co-Creation room.
2. You can import data artefacts from My space to support discussions that are open only to members of the group.
3. Use the right side of the screen to ‘Explore’ datasets, and ‘Develop’ ideas. Read the info on the screen.
4. Use these icons    to write notes, develop graphs and import visualisations from ‘My space’ for discussions and ideas development inside the Co-Creation room.
5. You can chat with group members also.
6. Observations and Perceptions: Observe all the various uses of the functions of the Co-Creation room, perceive the ease and simplicity of use, clarity of text as well as minimal action in term of creation of a forum.

Agora

Agora is a public space for discussions relating to a specific topic created by one of the users. Everyone who is registered on the ROUTE-TO-PA platform and users of SPOD will see the various discussion forums going on in SPOD and so can click in and participate without a specific invitation to do so because it is a public environment unlike the *Co-Creation* room.

1. Instructions: To create a forum on the landing page of SPOD when you log in:
 - a. Click on the Agora button on the top menu to open Agora where you can see many boxes of different colours, each representing a discussion room about a particular topic.
 - b. Click on the  icon at the button right of the page to add a new public discussion room.
 - c. Complete the form that opens with:
 - i. SUBJECT: Type in the discussion topic e.g. *Mopeds in Dublin; Dublin traffic problems; or Mobility improvement in Dublin*, etc. Different topic for each group
 - ii. click on the greyed ‘BODY’ just below the ‘SUBJECT’

iii. BODY: Type a short description of what this topic is about, the discussion expected, and the goals of the forum.

d. When done, click on the Submit button to create the public forum.

Note: that the forum created in Agora is automatically opened to members of the public who are users of SPOD.

e. Members of the group should discuss issues about the topic very briefly.

f. Other group members who are in other forums should participate very briefly on forums created by another group just to demo the fact that each forum on Agora is a public forum. You can just write *"Hello I like the discussions going on in this forum"* then submit your contribution to that forum.

Note: As the emphasis is not on the meaning of the discussion or its goals of improving Dublin traffic but on the demo of usability of the SPOD functions. Therefore, do not spend much time on discussions.

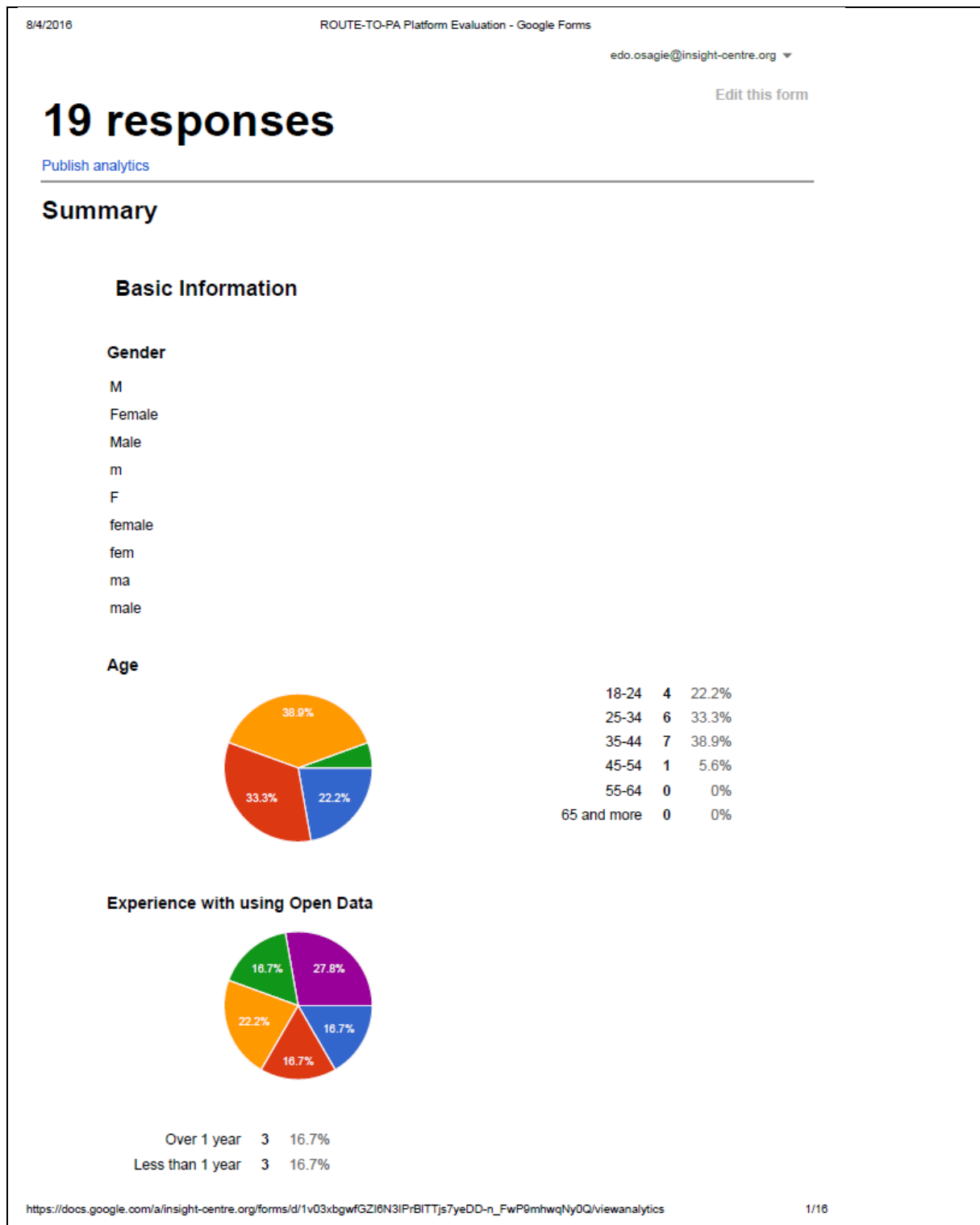
2. Observations and Perceptions: Observe all the various uses of the functions of Agora, perceive the ease and simplicity of use, clarity of text as well as minimal action in terms of creation of a forum.
3. Note down all your observations.

EVALUATION SURVEY

1. Please go to the google survey by clicking [here](http://bit.ly/2aimdmy). If this does not work, copy this link to your browser: <http://bit.ly/2aimdmy>.
2. Complete the all sections based on your observations and perceptions while carrying out the tasks in Sections 1 and 2.
3. You may repeat these activities if necessary to re-confirm your observations and perceptions.

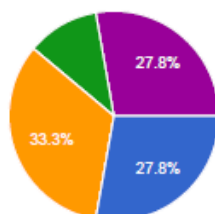
APPENDIX 3: USABILITY EVALUATION – GOOGLE SURVEY REPORT

In this appendix, we present comments of users and the score for the various features of only the TET aspect of ROUTE-TO-PA usability Evaluation workshop held in Galway, on Aug 3rd, 2016



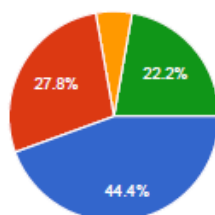
2 - 3 years	4	22.2%
4 years and more	3	16.7%
No major experience	5	27.8%

Frequency of using Open Data



Every day and more often	5	27.8%
Few times a week	0	0%
Once a week	6	33.3%
Once per month	2	11.1%
Very rarely to none	5	27.8%

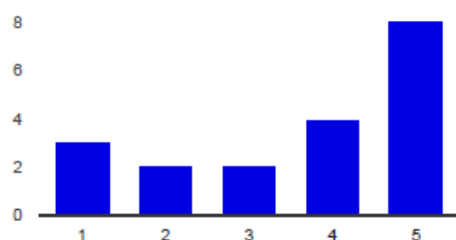
Function



Student	8	44.4%
Researcher	5	27.8%
Developer	1	5.6%
Public Servant	4	22.2%
Other	0	0%

1. Search for Datasets

The result list of datasets returned is relevant to the search keywords entered



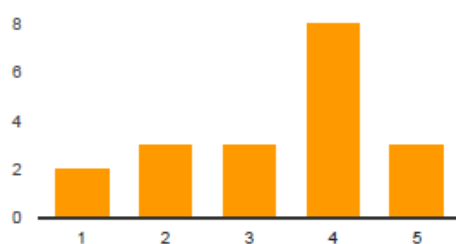
Strongly disagree: 1 3 15.8%

	2	2	10.5%
	3	2	10.5%
	4	4	21.1%
Strongly agree:	5	8	42.1%

If you disagree, please explain:

There are similar data set names/local authority confusion

Searching for Dublin wouldn't get Finglas etc. You must match the name of the data set for it to appear/

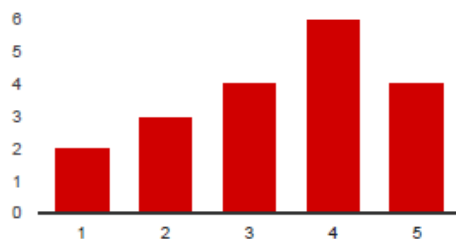
The filters options enhance the data search accuracy

Strongly disagree:	1	2	10.5%
	2	3	15.8%
	3	3	15.8%
	4	8	42.1%
Strongly agree:	5	3	15.8%

If you disagree, please explain:

guess so - didn't have to use them

not relevant to current exercise

The presentation of the results is clear and easy to read

Strongly disagree: 1 2 10.5%
 2 3 15.8%
 3 4 21.1%
 4 6 31.6%
 Strongly agree: 5 4 21.1%

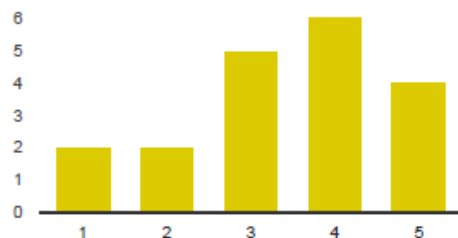
If you disagree, please explain:

It's not easy to find the way for the first time user once the user got t know the platform better
 it was more clear.

It's easy if you view the info in chart form first

2. Data resource views

The data tables and charts are simple and easy to read



Strongly disagree: 1 2 10.5%
 2 2 10.5%
 3 5 26.3%
 4 6 31.6%
 Strongly agree: 5 4 21.1%

If you disagree, please explain:

some of the data labels are unclear

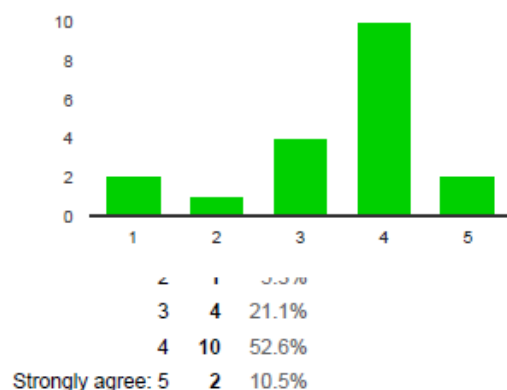
if charts can suggested based on data set features, would be great

Difficult to read long named titles from Database

Some large graphs have data bars over-lapping etc.

Charts are easier than tables

The datasets presented as data tables are understandable

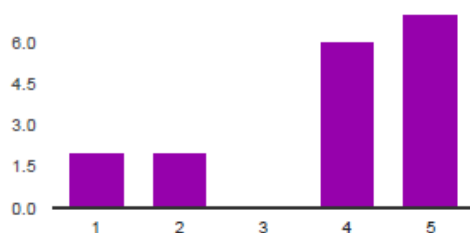


If you disagree, please explain:

Data in a table is never as clear as a good visualisation.

Charts are easier than tables

The datasets presented as charts are understandable



Strongly disagree: 1 2 11.8%

2 2 11.8%

3 0 0%

4 6 35.3%

Strongly agree: 5 7 41.2%

If you disagree, please explain:

i would probably limit the options

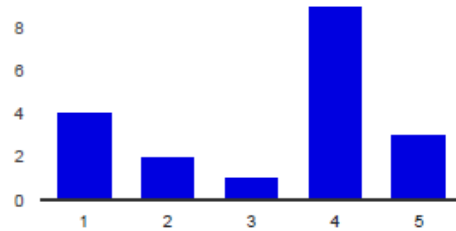
Sometimes axes are just numbers rather than saying what they are.

3. Description of data resource

I understand the description of the dataset.

https://docs.google.com/a/insight-centre.org/forms/d/1v03xbgwGZl6N3lPrBITTs7yeDD-n_FwP9mhwqNy0Q/viewanalytics

5/16



Strongly disagree: 1 4 21.1%

2 2 10.5%

3 1 5.3%

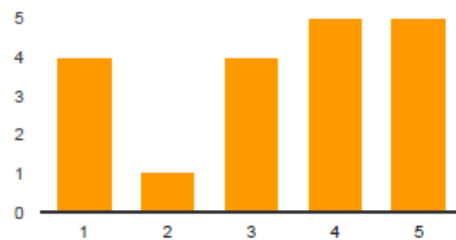
4 9 47.4%

Strongly agree: 5 3 15.8%

If you disagree, please explain:

I have no idea what you mean by this!

Data description provides sufficient information about the content and meaning of the dataset I viewed



Strongly disagree: 1 4 21.1%

2 1 5.3%

3 4 21.1%

4 5 26.3%

Strongly agree: 5 5 26.3%

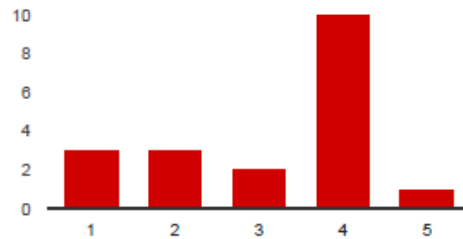
If you disagree, please explain:

see above - are you still on TET or SPOD

it depends on the datasets, some descriptions are easier than others

4. Social discussions on data resource

I find the discussions from SPOD on my dataset relevant to my need about the dataset

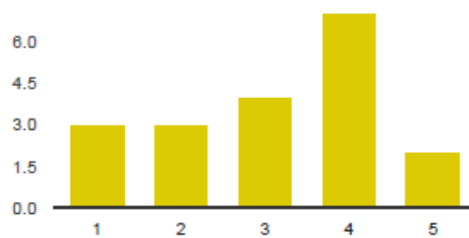


Strongly disagree: 1	3	15.8%
2	3	15.8%
3	2	10.5%
4	10	52.6%
Strongly agree: 5	1	5.3%

If you disagree, please explain:

we had difficulties connecting with each other so didn't get to try it out
The discussions from SPOD were not visible on Route to PA
It was more about using the discussion than the dataset itself

The discussions provide some sense about the content of the dataset



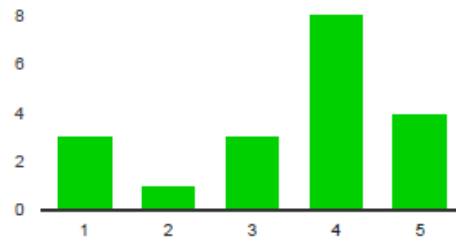
Strongly disagree: 1	3	15.8%
2	3	15.8%
3	4	21.1%
4	7	36.8%
Strongly agree: 5	2	10.5%

If you disagree, please explain:

see above

The comments appear to be just test comments at the moment and do not reflect an understanding or reflection of the contents of the datasets.

It made sense of the discussion space

The discussions about data on SPOD are clear and understandable to me

Strongly disagree: 1 3 15.8%

2 1 5.3%

3 3 15.8%

4 8 42.1%

Strongly agree: 5 4 21.1%

If you disagree, please explain:

see above

Additional Comments**Please provide any additional comments and feedback**

overall it was a good experience

what does 'few steps' mean - positive (only a few) negative (quite a few)

1) Copy pasting the API Url from TET to my space in SPOD is a bit clumsy, perhaps a single button in TET (export to myspace in SPOD) can do it. 2) If we click (Discuss the dataset) in TET, it takes you to the new feed in SPOD rather than a discussion room in AGORA or Co-creation

Not obvious when I click on SPOD Discussion tab in the data resource view how to start a discussion, as Start the discussion link is on the bottom right. Also goes to login screen even though I am logged onto SPOD already. It would be useful to have a select all fields option. When I go to edit a graph it loses my x-axis and y-axis already selected. The maps graph is not zooming to the extent and it is centred in the ocean.

• TET - Like the auto generated graphs- allows the user to explore the data set • TET - Like the fact you can go into the table from the auto generated graphs • TET- Like the ability to go to the classic view • Dislike - table headings (at Data Creation Stage) difficult to interpret multiply variations • TET searching: similarly named data sets • TET incomplete data - When uploading data should Data Officer be warned that data is missing - Error Correction • Like the Comments Graph in Agora • Want - Name of data set in Tabular mode • Want- Process or method used to determine Data Sets • Want - Pop-up tool tips • Social Media Like link in TET to transfer to SPOD - My Space • SPOD - My Space - Select DataSet - (Showing 1 to 20)

APPENDIX 4: GITHUB ISSUES LIST

Github Issues List

<input type="checkbox"/> 14 Open <input checked="" type="checkbox"/> 5 Closed		Author ▾	Labels ▾	Milestones ▾	Assignee ▾	Sort ▾
<input type="checkbox"/>	RTPA: Some Issues from Usability Evaluation #20 opened 11 hours ago by eosagie					
<input type="checkbox"/>	TET Beta Issues #19 opened 12 hours ago by eosagie					
<input type="checkbox"/>	dataset search result page, title of each file in it's tooltip #17 opened 3 days ago by mogaio					
<input type="checkbox"/>	Files section, number 3 is always added at the bottom #16 opened 3 days ago by mogaio					
<input type="checkbox"/>	Related Datasets slide bar needs a brief description #15 opened 3 days ago by mogaio					
<input type="checkbox"/>	[Error] [High Priority] Links on related datasets do not work #14 opened 3 days ago by callmealien					
<input type="checkbox"/>	Problems in visualising dataset #13 opened 6 days ago by elenapalmisano					1
<input type="checkbox"/>	Summary table not dispalyed properly #12 opened 9 days ago by waqarini					
<input type="checkbox"/>	Update Bootstrap enhancement Medium priority #11 opened on 15 Nov 2016 by arekstasiewicz					
<input type="checkbox"/>	Discussion widget enhancement High priority #10 opened on 15 Nov 2016 by arekstasiewicz					
<input type="checkbox"/>	Dataset details page enhancement High priority #9 opened on 15 Nov 2016 by arekstasiewicz					
<input type="checkbox"/>	Share icons enhancement Low priority #8 opened on 15 Nov 2016 by arekstasiewicz					
<input type="checkbox"/>	TET logo enhancement #7 by arekstasiewicz was closed on 15 Nov 2016					1
<input type="checkbox"/>	Search box enhancement #6 by arekstasiewicz was closed on 15 Nov 2016					1
<input type="checkbox"/>	TET Interface - Response to Language Translations #5 by eosagie was closed on 15 Nov 2016					2
<input type="checkbox"/>	TET on Internet Explorer #4 opened on 3 Nov 2016 by eosagie					
<input type="checkbox"/>	TET Review on IE environment #3 opened on 3 Nov 2016 by eosagie					
<input type="checkbox"/>	TET platform #2 by eosagie was closed on 15 Nov 2016					1
<input type="checkbox"/>	[Bug] Chart View not Working bug #1 by callmealien was closed on 15 Nov 2016					7

ProTip! What's not been updated in a month: [updated:<2016-12-19](#).

BIBLIOGRAPHY

- [1] A Control Theory Perspective on Agile Methodology Use and Changing User Requirements, Likoebe M. Maruping, Viswanath Venkatesh, <http://pubsonline.informs.org/doi/pdf/10.1287/isre.1090.0238>
- [2] U-SCRUM: An Agile Methodology for Promoting Usability, Mona Singh, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4599538>
- [3] The Scrum Guide™ The Definitive Guide to Scrum: The Rules of the Game
Ken Schwaber and Jeff Sutherland, <http://www.scrumguides.org/docs/scrumguide/v1/scrum-guide-us.pdf>
- [4] Joomla. Drupal and WordPress - A Statistical Comparison of Open Source CMS. Savan K Patel , V.R. Rathod, Satyen Parikh. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6169111>
- [5] Quote from the CKAN 'README' file <https://github.com/okfn/ckan> (22/05/2013)
- [6] Open Data and the Academy: An Evaluation of CKAN for Research Data Management, Joss Winn, University of Lincoln, UK.1, <http://eprints.lincoln.ac.uk/9778/1/CKANEvaluation.pdf>
- [7] Beck, Kent; et al. (2001). *Manifesto for Agile Software Development*. Agile Alliance. Retrieved 14 June 2010.
- [8] Jen, L. and Lee, Y. (2000) IEEE recommended practice for architectural description of software-intensive systems, IEEE Architecture.
- [9] Tutorialspoint Online (n.d.) SDLC – The V-Model. http://www.tutorialspoint.com/sdlc/sdlc_v_model.htm
- [10] Easterbrook Steve (2010). The Difference Between Verification and Validation. Serendipity blogpost. <http://www.easterbrook.ca/steve/2010/11/the-difference-between-verification-and-validation/> Viewed 09 Oct., 2015.
- [11] Tutorialpoint (Online) http://www.tutorialspoint.com/sdlc/sdlc_v_model.htm. Viewed 12 Oct., 2015
- [12] Vihari, Chaitanya (2014) Software Development Lifecycle Models. <http://www.testnbug.com/2014/12/software-development-life-cycle-models/>. Viewed: 13 Oct., 2015
- [13] Firesmith, Donald (2013) Using V Model for Testing. https://insights.sei.cmu.edu/sei_blog/2013/11/using-v-models-for-testing.html. Viewed 12 Oct., 2015
- [14] ISTQB (online) What is Agile Model – Advantages and Disadvantages. <http://istqbexamcertification.com/what-is-agile-model-advantages-disadvantages-and-when-to-use-it/>. Viewed 12 Oct., 2015
- [15] Institute of Electrical and Electronics Engineers. (IEEE 829) Software Testing Template.
- [16] Eric Bissonnette 2015. Test Plan Part 1: Why Create a Test Plan. Source: <https://crowdsourcedtesting.com/resources/why-create-a-test-plan/> Viewed: 07 Oct., 2015
- [17] ISTQB (n.d) What is the Purpose and Importance of Test Plans in Software Testing? <http://istqbexamcertification.com/what-is-the-purpose-and-importance-of-test-plans/> viewed 07 Oct., 2015
- [18] Purpose & Advantages of Test Plan (n.d) <http://www.softwaretestingtimes.com/2010/05/purpose-advantages-of-test-plan.html>
- [19] Easterbrook Steve, Professor, Department of Computer Science, University of Toronto, Canada.