PUBLICAMUNDI

SCALABLE
REUSABLE
OPEN
GEOSPATIAL
DATA

DELIVERABLE 5.1

VALIDATION RESULTS

Dissemination Level	Public
Due Date of Deliverable	Month 24, 31/10/2015
Actual Submission Date	31/10/2015
Work Package	WP5, Validation, Dissemination & Exploitation
Task	T5.1
Type	Report
Approval Status	Submitted for approval
Version	1.0
Number of Pages	20
Filename	D1.1_Validation_Results.pdf

ABSTRACT

This report analyzes the validation results for all system aspects of PublicaMundi, based on actual stakeholder feedback received during our 2nd GeoDataCamp (Athens, Bremen). We first provide background information for the events and the participating stakeholders. In the following, we present and analyze received feedback grouped around the major functionality clusters of PublicaMundi. In this analysis, we also provide information on how we exploited stakeholder feedback to improve the PublicaMundi system. Where relevant, we also discuss areas that fall outside the scope the project, but may be of use to the open data community.

HISTORY

version	date	reason	revised by
0.1	8 Sep 2015	Initial Draft	Spiros Athanasiou
0.2	3 Oct 2015	Improvements on draft	Alex Dumitru, Angelos Tzotsos
0.3	15 Oct 2015	Version prepared for internal peer-review	Spiros Athanasiou
0.6	18 Oct 2015	Peer Review	Thodoris Vakkas, Angelos Tzotsos
1.0	31 Oct 2015	Final Version	Spiros Athanasiou

AUTHOR LIST

organization	name	contact
IMIS	Spiros Athanasiou	spathan@imis.athena-innovation.gr
IMIS	Angelos Tzotsos	tzotsos@imis.athena-innovation.gr
Rasdaman	Alex Dumitru	dumitru@rasdaman.com
GET	Thodoris Vakkas	tvakkas@getmap.gr

EXECUTIVE SUMMARY

This report analyzes the validation results for all system aspects of PublicaMundi, based on actual stakeholder feedback received during our 2nd PublicaMundi GeoDataCamp (Athens, Bremen).

The first PublicaMundi GeoDataCamp was organized in Athens (June 2014) and served as a first introduction of the project to stakeholders (*data publishers, developers, private sector, citizens*), as well as an opportunity to collect user requirements and receive early feedback for the available PublicaMundi prototypes.

The second PublicaMundi GeoDataCamps were organized in Athens and Bremen (September/October 2015) with a focus on validating all software deliverables of the project. During the events project members were available to shortly present the available services/APIs, guide the participants, respond to questions, and document feedback.

Feedback received from the participants was analyzed and presented in this report, grouped around the major functionality clusters of PublicaMundi. Overall, we have achieved great coverage for all system aspects, and received extremely positive comments regarding the provided functionalities and ease of use. However, in this report we focus only on *critical feedback*, i.e. areas that required improvements in our software. We present how we exploited stakeholder feedback to improve the PublicaMundi software across all of its aspects (*e.g. technical, UI/UX, documentation*). When relevant, we also discuss feedback that fall outside the scope the project, but may be of use to the open data community.

Finally, we would like to emphasize that this report presents validation results based *only on feedback received during our 2nd GeoDataCamps*. With PublicaMundi technologies being in a public beta for more than a year (*labs.geodata.gov.gr*), and with our source code being publicly available (*github.com/PublicaMundi*), we have received valuable insights, critique, and suggestions for improvement from numerous users, developers and data publishers. We would like to express our sincere acknowledgments to all these community members.

ABBREVIATIONS AND ACRONYMS

OGC	Open Geospatial Consortium
WMS	Web Map Service
WFS	Web Feature Service
CSW	Catalogue Service for Web
WPS	Web Processing Service
WCS	Web Coverage Service
WCPS	Web Coverage Processing Service
SQL	Structured Query Language
API	Application Programming Interface
SDI	Spatial Data Infrastructure
FOSS	Free and Open Source Software
OSGeo	Open Source Geospatial Foundation
CRS	Coordinate Reference System
GIS	Geographical Information Systems

TABLE OF CONTENTS

1 PUBLICAMUNDI GEODATACAMPS

The first PublicaMundi GeoDataCamp was organized in Athens (June 2014) and served as a first introduction of the project to stakeholders (*data publishers, developers, private sector, citizens*), as well as an opportunity to collect user requirements and receive early feedback for the available PublicaMundi prototypes.

The second PublicaMundi GeoDataCamps were organized in Athens and Bremen (September/October 2015) with a focus on validating all software deliverables of the project. During the events project members were available to shortly present the available services/APIs, guide the participants, respond to questions, and document feedback.

In the following, we provide background information regarding the events (*organization, participants*).

1.1 ORGANIZATION DETAILS

The 2nd PublicaMundi GeoDataCamps were organized in September/October 2015 in Athens and Bremen, by partners Athena RC and rasdaman respectively. Their focus was on validating all software deliverables of the project (from methodology to data analysis APIs). For these events, we have targeted broader user participation, ensuring good coverage across all potential stakeholders (*public sector, private sector, citizens*). A public invitation was available in the project's web site and distributed through public email lists, social networks, and the partner's existing dissemination channels.

Our initial goal was to limit participation to *at most 40 individuals* per location, in order to ensure a productive collaboration with the project's staff. However, due to increased interest, we raised the limit of participants to 50, while also accommodating requests for demonstration/training events. This was the case for the GeoDataCamp organized in Athens, as the production deployment of PublicaMundi software in geodata.gov.gr gained publicity. To address these requests, we took the following actions:

- *Teleconferences*. We organized five (5) teleconference sessions (Google Talk) to accommodate participation requests from public sector bodies situated outside Athens and/or with a large number of interested participants (>5). These sessions were mostly focused on presenting the data publishing methodologies, as well as on exploring

how PublicaMundi software can be applied to also address organizational needs beyond open data publishing (e.g. transit maps DSS for primary/secondary education).

- *Physical meetings*. We organized three (3) physical meetings in our premises with private sector representatives (*ICT SMEs*, *large integrators*, *individual developers*). These meetings focused on presenting the PublicaMundi APIs, with particular emphasis in our Data and Mapping APIs. We received excellent feedback, and several requests for providing functionalities beyond the current scope of the project (e.g. rich infographics, paid APIs, private deployments), which will be the focus of our future exploitation actions.

- *Community Events*. We participated in several community events (*e.g. Hackathons, competitions, open days*) with a role extending beyond our dissemination actions. In particular, we received inquiries for exploiting PublicaMundi APIs during these events from participating developers (*short training, support, evaluation*). We exploited these opportunities as an intense and critical beta-testing, offering on-site support and close interaction with the participants.

- *Extra documentation*. We prepared training material targeting Greek public sector publishers due to increased demand for internal training purposes. This activity provided us with critical feedback for the case of *novice data publishers*, leading us to introduce several improvements in the UI and publishing processes.

1.2 PARTICIPANTS

We intentionally assembled a representative sample of all PublicaMundi stakeholders, giving special emphasis in *data publishers and developers*, which are the core target group of the project. In particular:

- Data Publishers (*experienced*). This group comprised public sector representatives experienced in the policy and technology aspects open data publishing (*i.e. >3 years experience, >10 data sets published*).

- Data Publishers (*novices*). This group comprised public sector representatives with minimal experience in policy and technical aspects of open data publishing. Due to the recent local reinvigoration of the PSI Directive, they were motivated but highly inexperienced even in core terminology and concepts.

- SDI and Web GIS owners (*public sector*). This group comprised public sector representatives that operate and existing SDI and/or Web GIS for internal or external purposes. These were mostly focused on examining how their existing investments can be repurposed and extended by PublicaMundi (*e.g. harvesting, interactive maps, APIs*), as well as researching the application of PublicaMundi for replacing their existing systems. Their reasoning was that the provided open source software exceeds the functionality of their existing systems, while also ensuring INSPIRE/PSI compliance.

- Public Sector (*policy-monitoring/making*). This group comprised public sector representatives responsible for monitoring and proposing geospatial-related policy issues (*ranging from education, to energy, and transport*). They were motivated by two individual requirements. First, examine how open geospatial data and PublicaMundi services can support their work. Second, explore the application of PublicaMundi for their specific domains.

- Developers (*established SMEs, integrators*). This group comprised developers from SMEs and large ICT integrators. Two specific sub-groups were observed. The first was targeting public sector contracts and wanted to evaluate PublicaMundi as a Web GIS/geospatial DSS application that offers de facto INSPIRE conformance. The second was interested in the technology of PublicaMundi and the means for integrating our output in their commercial offerings.

- Developers (*start-ups, individual*). This group comprised mobile/web developers building applications with open data. They focus was on understanding and using the various APIs provided by PublicaMundi in commercial applications. Most of them already provided a product (e.g. tourist apps) and explored the application of our interactive maps and data APIs to complement and/or simplify their server backends.

- Developers (*open source-oriented*). This group comprised well-experience developers active in the application of open source technologies, with specific emphasis in OSGeo projects. These were focused on issues relating to the PublicaMundi deployment and integration with open source software, as well as in comparing the provided functionalities with open and proprietary software.

- Academia and Researchers (*postgraduate, senior*). This group comprised postgraduate students (*MSc, PhD candidates*) and researchers (*Senior, Professors*) working with geospatial data in their

respective domains. Most were proficient in GIS technologies (*especially visualization/analysis*) and their focus was the application of PublicaMundi services as research-support instruments. They were motivated by the inherent availability of high quality data, the lack of deployment/administration costs, and the use of complex WPS services for their domain-specific analysis needs.

- Citizens (*NGO, Citizen Associations*). This group comprised citizens with an established background of participation in NGOs and open participatory events. These users were motivated by the vision and impact of PublicaMundi. We engaged them due to their role as active open data users, but also as multipliers for our dissemination actions.

1.3   FEEDBACK COLLECTION

Participants in the GeoDataCamps were provided with several opportunities for gaining hands-on experience in all PublicaMundi aspects, with constant interaction with Consortium members. The majority of feedback was received during the vis-à-vis interactions of PublicaMundi members with participants in the form of explicit suggestions for improvement, as well as frequently asked questions related to the provided services. Our members were instructed to lead discussions towards technical aspects, without however avoiding more generic questions relating to open data publishing. For example, comments from novice publishers related to the necessity for completing all metadata fields were followed-up by examples on why this helps data discovery from citizens.

In particular, we performed the following sessions for soliciting feedback from participants. All participants used labs.geodata.gov.gr as a testing sandbox, with complete freedom for experimenting across its various services.

- Data Publishing (assisted). Participants were guided by Consortium members in the entire data publishing workflow with real-world examples of increasing complexity. We presented the process on a step-by-step basis, allowing the participants to interrupt us at any point in time.

- Data Publishing (solo). Participants were asked to publish five datasets of their choosing in labs.geodata.gov.gr. Consortium members provided limited assistance during this process and documenting common challenges/issues.

- Data discovery (solo). Participants were given four different data discovery retrieval scenarios of increased complexity: search for a data set, download in custom format/CRS, find the corresponding WMS, provide an embedded map for the data set. During this session PublicaMundi members did not provide any assistance.

- Interactive maps (solo). Participants were give four different usage scenarios for interactive maps, of increased complexity: find a data set, create a custom map, print a map, upload a custom geospatial data set. Again, during this session PublicaMundi members did not provide any assistance.

- Data and Mapping APIs (assisted). Participants were guided by Consortium members in several scenarios of how PublicaMundi APIs can be discovered and invoked (OGC, Data API, Mapping API). We presented the process on a step-by-step basis, allowing the participants to interrupt us at any point in time

- Data and Mapping APIs (solo). Participants were given four usage scenarios for invoking our APIs, of increased complexity: retrieve a list of beaches, query for beaches near cities, create an interactive map with these results, analyze a satellite image. They were instructed to complete these assignments using our Data and Mapping API interactive authoring tools.

- Grouped discussions. Participants were grouped based on their group/session and a short open discussion was initiated by PublicaMundi members with the goal of explicitly requesting comments and feedback. A plenary session followed where the major comments were presented and discussed by all participants.

2 VALIDATION RESULTS

In this section we provide an analysis of feedback received during the PublicaMundi GeoDataCamps, as well as during their complementary actions. We have grouped feedback around the major functionality clusters of PublicaMundi. Overall, we have achieved great coverage for all system aspects, and received extremely positive comments regarding the provided functionalities and ease of use. However, in the following we focus only on *critical feedback*, i.e. areas that required improvements in our software. Further, we present how we exploited stakeholder feedback to improve the PublicaMundi software across all of its aspects (*e.g. technical, UI/UX, documentation*). Finally, and when relevant, we also discuss feedback that fall outside the scope the project, but may be of use to the open data community.

2.1 DATA PUBLISHING

2.1.1. Organization Management

Short Description	Comments	Action
Organization aliases. In case a data publisher changes its official name, this propagates to the URL names of the various resources.	A change in the official name of public bodies is infrequent for most EU countries. However, while moving to lower government levels, it could be more frequent	The URL scheme has been adapted to remove organization short names from URLs. Therefore even when the publisher name changes, the URL is not affected
Structured publisher hierarchy. Assign roles within an organization based on its official departmental hierarchy	After discussion with data publishers, we reached the conclusion that the request implied the transfer of actual chain-of-command (*and thus responsibility*) within each organization.	No action taken. The introduction of hierarchies would complicate management and introduce artificial (not documented) barriers of responsibility. Current role management suffices.

2.1.2. Metadata

Short Description	Comments	Action
Increase visibility of option for importing existing metadata	Data publishers could not intuitively understand the process for creating a dataset by reusing existing metadata (essentially missing a shortcut in the UI)	Implemented; added option under the 'Add dataset' button
Organize metadata editor in Tabs	Data publishers were mostly used to the editor structure of the INSPIRE geoportal, using tabs for metadata groups	Implemented using "accordion" widgets to follow CKAN UI best practices.
Order of options in drop-down lists for metadata creation	To speed-up metadata creation, it was suggested to reorder the options in the drop-down lists in decreasing popularity	Implemented.
Status information for unpublished datasets and resources	Data publishers required a simple visual indication in their Dashboard regarding the various status stats of their datasets and resources	Implemented; Colored tags next to resources or datasets in the Dashboard provide this information in a glance

2.1.3. Data

Short Description	Comments	Action
Styling options for vector datasets	Data publishers needed specific UI elements to help them change the map styling of the published data.	Implemented in the Management page of OGC web services for vector resources. Support for defining a vector style through SLD is provided.
Provide live links for Mapping and Data API	Developers asked for a unified page that will include all the needed resources to access the PublicaMundi APIs	The feature was implemented by adding a "Developers" tab under each dataset, with an accordion widget, providing links and code snippets for accessing

		vector data from the APIs
Add WMS preview to the original resource	Some publishers commented that by default the uploaded resource is not assigned a preview in cases of zip files including shapefiles	A preview feature was implemented using the WMS preview of ingested datasets
Rename WMS layers in vector ingestion process	Some publishers asked for a layer renaming tool during the publication process	A text field was added in the ingestion popup to ask for a display name for the ingested resource for display purposes
Select format from a provided list	Users wanted to avoid typing the uploaded resource format and asked for a drop-down list of supported formats	An auto-complete widget was implemented, including the supported vector formats
Data preparation	A frequent question from data publishers regarded the steps towards preparing geospatial resources for upload (*e.g. format, CRS*).	No action taken; during training activities we encouraged publishers to submit data *as-is*, explaining them how the automatic ingestion and on-demand transformation processes work.
Support remote access to raster data	Since raster data are often massive in size, users requested an option to provide URLs of already uploaded raster data instead of uploading them during publication	This was implemented with a Celery task, which downloads the raster data asynchronously in order to publish to OGC web services at a later stage through the Dashboard.
Support time series for raster data	Some users from the earth observation field requested a way to be able to include time series raster data through Rasdaman	The import process through WCS-T has been improved for importing raster data cubes

2.1.4. Management & Monitoring

Short Description	Comments	Action
Disconnect the catalogue topics from the topic hierarchy in interactive maps	The standard behavior of the interactive maps was to use the resource's assigned topics (from publishers) and use them to populate the layer hierarchy of the interactive maps. However, since a resource can fall in more than one topics, the layer can also appear multiple times in the layer hierarchy	Implemented. We developed functionality integrated in the administrative panel enabling the explicit assignment of resources to a topic hierarchy. Further, the topic hierarchy for the interactive maps and the catalogue can be completely different. This enables better control and curation of the interactive maps and is optional; the catalogue owner can revert to the original behavior
Custom styles for layers	The automatic ingestion process handling all steps after the upload of a resource did not permit the definition of the style to be used in the maps. This was deemed necessary to support novice data publishers, but also to ensure that layers from different publishers did not appear with the same styles	Implemented. We extended the above administrative environment and allowed the definition of custom styles per resource. Once again, this level of curation is optional; the catalogue can revert to the original behavior
Add paging to the administrator dashboard	Publishers requested for paging capabilities in the resources list in the administrator dashboard	This feature was implemented in the dashboard
Add filtering and sorting in the administrator dashboard	Publishers requested sorting, filtering and search functionality in the administrator dashboard to be able to locate	This feature was implemented in the dashboard

	resources faster	

2.1.5. Multilinguality

Short Description	Comments	Action
Language indicators in the dataset page	Any available translations for a resource should be visible regardless of the catalogue's selected language	Implemented; a dynamic list of all available translations is offered for each resource
Color-coding progress	Publishers requested that the translated fields should be marked with specific cell color (e.g. orange for draft, green for translated, red for pending etc)	Implemented; the translation status is color-coded.
Provide API functions for translated resources	Users were interested in accessing the translated vector resources in a machine-friendly way	An API call was implemented to provide access to translated vector resources
Allow un-publishing resource translation for editing	Some publishers were concerned about the need to correct previously published resource translations, so they asked for the ability to revert a previously published translation until they are satisfied with the changes	Vector resource translations are now automatically unpublished when the authorized user chooses to edit the translation in that language, until he selects to publish again
Metadata translations UI tab	Publishers requested for a new "Translations" tab in the Dataset Management area of CKAN instead of a "Translate" Button leading to a separate translation page	This feature was implemented adding a translations tab in the dataset management area in CKAN

2.1.6. Interlinking

Short Description	Comments	Action
Unsuccessful interlinking	When at least one element in the source column cannot be linked with the target column (i.e. no matches), then the complete interlinking processes should be able to successfully complete.	Implemented; this comment is particularly appropriate for geospatial data of low quality, or for historical data (*e.g. city names used decades ago*). The interlinking process can be completed if and only if all elements of the source column have been linked with the target column.
Ranked results	For each element, the ranking of candidate matches with the source column should be more clear	Implemented; we slightly redesigned the UI to improve visibility
Color-coding progress	The user should be further assisted during the process for manually reviewing/editing linking results	Implemented; we slightly redesigned the UI to improve visibility. Publishers can now immediately distinguish links established manually.
Column ordering	The user should be able to reorder columns and thus gain a better overview of each row that needs to be interlinked	No action taken; column reordering was already supported
Indicating rows edited by the publisher	Publishers requested that somehow they should be able to tell in which rows they edited the interlinked results	Implemented; an additional field exists showing a checkmark for rows edited by the publisher

2.2 DATA DISCOVERY

Short Description	Comments	Action
Explore published data based on organization hierarchy	Visitors would like to explore the hierarchy of public sector organizations (*e.g. Ministry, Department*) and discover its published data sets. This functionality has been offered in the previous version of geodata.gov.gr	No action; the hierarchy of public sector organizations is not publicly available in a machine-readable format, while the notion of hierarchy only serves visitors working in the public sector
View in a glance the available formats and services of a dataset	The summary of a dataset (as included in any search result) should provide the list of formats/services available.	Implemented; while visitors can download resources at their format/CRS of choice, it was considered important to highlight the original resource format
Indicator for INSPIRE metadata compliance	When viewing a dataset, visitors wanted a simple indication providing at a glance information regarding INSPIRE metadata compliance	Implemented; a simple check box in grey with the words 'INSPIRE' appears next to Metadata
Single Download button	Two download buttons where provided, one for the original resource and the other for invoking on demand transformations in other formats/CRS.	Implemented; a single button provides this functionality
Simplify on-demand transformation	Users asked for a simpler way of transforming the data, before downloading them from the "Download" dataset button.	Implemented using a JavaScript pop-up window, where the user selects the SRS and the output format of the resource.
Web Services as resources	Visitors could not understand where they could discover the automatically generated OGC services; users had to press the Detail button	Implemented; we have replaced the 'Details' button with 'Endpoint', which directs to the service's GetCapabilities

Dataset availability in interactive maps	Visitors wanted a direct link from the resource to the interactive map	Implemented; the dedicated Map button at a resource open's the interactive map and the corresponding layer

2.3 INTERACTIVE MAPS

Short Description	Comments	Action
Search for placenames and visible layers in the main search box	The top-level search box by default should include placenames and search only on visible layers; the provided functionality for search across all published data was deemed overly complex	Implemented
Search for available layers	The left-pane search box by default should provide results only for the available layers and not for Topics/Publishers	Implemented
Improve the visibility of map tools	The map tools were available in a bottom-left pane; many users could not easily find them	Implemented; the map tools are permanently visible at the top-right pane of the maps
Remove need to explicitly press the 'Select Object' button	For users not familiar with GIS, the need to press 'Select Object' and thus transit from panning to selection was considered unintuitive	Implemented; select functionality is turned-on by default
Upload polygon	A very common use case for the interactive maps is the uploading of a polygon (bounding box or more complex geometry) in order to locate it, cross-check it against areas of environmental interest, and print maps.	Implemented; user can upload a polygon at a format/CRS of their choosing.
Add CAD formats	Many engineers use CAD	Implemented; DXF files

Short Description	Comments	Action
	programs instead of GIS; an option to upload and download CAD files was requested	can be uploaded, while map exports can be downloaded in DXF
Add common WPS services as aliases	Developers requested addition of commonly used services (like buffer, intersect etc) with aliases within the Fluent WPS API	This feature was implemented in the Data API

2.4 MANAGEMENT

Short Description	Comments	Action
Deployment through configuration	Most early testers expressed their concerns regarding the complexity of setup for the system	We automated the installation of the whole system, including the production site, using the popular DevOps framework Ansible. Users can define their configuration through simple text files and trigger the system deployment.
Partial deployment of the system	Early testers asked if there is a way to deploy only parts of the system (e.g. without raster or vector support)	The installation process through Ansible was adapted to be flexible enough to skip parts of the installation depending on the user requirements, although some core features (e.g. ckanext-publicamundi extension) cannot be removed