



1. Publishable Summary

Introduction

PublicaMundi is a research project originating from everyday problems faced by open data publishers and data consumers alike. Simply stated, open geospatial data are cumbersome to easily *publish* and *consume* for non-GIS experts. While most data publishers and developers are familiar with handling and using typical data (e.g. csv), they are not familiar with the intricacies of geospatial data. Different coordinate reference systems, geospatial databases, map servers, special standardized APIs are some of the tools and know-how required to publish and reuse open geospatial data.

PublicaMundi aims to democratize open geospatial data publishing and reuse, making easier for publishers to share data and for developers to discover and reuse data.

- Project web site: www.publicamundi.eu
- Source code repository: www.github.com/PublicaMundi

The PublicaMundi Consortium comprises “Athena” Research Center (coordinator), rasdaman GmbH, GeoLabs SARL, and Get Ltd.

Motivation

Open data and information provided by the public sector constitute a significant opportunity for transparency, accountability, better governance, and citizen participation. Reuse of open data can also serve as an instrument for growth, leading to innovation through research, better products and services, new jobs and economic advancement. The Vickery report estimates the benefits of extended PSI (Public Sector Information) reuse for the EU27 economy at 140b€/year (1.7% GDP 2008).

These financial benefits will be materialized through the establishment of a Data Economy, led by SMEs providing added value services by repurposing and extending open public data.

Geospatial data account for an estimated 80% of public sector information and are the most significant category of open public data due to their high production, procurement and update costs, as well as their relevance in multiple thematic areas and domains. The availability of such high value open data has the potential to create and sustain a multi-billion market of applications and services. By nullifying the extremely high costs associated with geospatial data production, SMEs can have unprecedented access to a valuable body of knowledge, enabling novel value added applications and services.

Despite the importance of open geospatial data, they are increasingly difficult to discover and reuse, especially in a cross-boundary multilingual context. The vast majority of open data catalogues in the EU have limited support for geospatial information, across all phases of their lifecycle. Geospatial data are treated as *second class citizens*, with insufficient capabilities in publishing methodologies and tools, limited technical foundations to support

value added services, and simplistic non-scalable support for geospatial data visualization. These problems directly influence the capacity of SMEs to reuse open geospatial data and deliver value added applications.

Goals

PublicaMundi will provide reusable tools and technologies for comprehensive, sophisticated and scalable publishing of open geospatial data, with emphasis on streamlining and maximizing their reuse in value added services and applications. Are goals are to deliver:

- A sustainable, efficient, traceable and easy to use publishing methodology, fully supporting the entire *lifecycle* of open geospatial data.
- An open geospatial data catalogue with full support for open geospatial data curation and *management*, integrating the developed data interlinking, multilinguality, processing, analysis, mapping, and visualization software components.
- Technologies and tools to assist in interlinking and multilingual support of open geospatial data in order to increase their value, relevance, and applicability for value added applications.
- Reusable software components implementing mapping and analytics services for demand-aware visualization of open geospatial data, enabling rapid integration in multimodal and value added applications.
- Reusable software components implementing processing and analysis services for demand-aware processing and analysis of open geospatial data, enabling rapid integration in multimodal and value added applications.
- Comprehensive, real-world validation through geodata.gov.gr, of the project's methodologies and software *components* regarding usability, sustainability and purposefulness for developing valued added applications
- A showcase of EU innovation for open geospatial data, mobilizing EU members states, SMEs and individuals towards realizing the EU data economy

Results of the First Year

PublicaMundi aims to make open geospatial data easier to discover, reuse, and share by fully supporting their complete publishing lifecycle in open data catalogues. To achieve this, we are extending and integrating leading open source software for open data publishing and geospatial data management. In this manner, the PublicaMundi technologies will be applied as a whole, or through its individual components, by thousands of developers and real world systems worldwide.

During the first year of the project, we have focused on extending CKAN, the leading data catalogue, into treating geospatial data as first-class citizens and providing OGC and INSPIRE compliant access to geospatial services. A first integrated prototype is already available on labs.geodata.gov.gr (*alias web.dev.publicamundi.eu*), providing beta access to data publishers and developers for Greek open geospatial data.

- *Data Publishing.* We have extended CKAN with an easy to use and flexible data publishing workflow for geospatial data, supporting vector and raster data, in addition to standardized metadata schemata. Data publishers are guided step-by-step into creating metadata in their schema of choice, or importing existing metadata. The metadata they provide can then be transformed on-the-fly in any supported schema. If an existing data catalogue or Spatial Data Infrastructure is available (e.g. INSPIRE-compatible SDI), the data publisher only needs to provide a simple entry point, and all available metadata and data resources are automatically harvested.

Our data publishing workflow significantly lowers the entry barrier for data publishers, while also accommodating the data publishing needs of organizations with existing data catalogues and SDIs, maintaining full compatibility with INSPIRE and OGC web services.

- *Data Discovery.* We have integrated pycsw in CKAN for providing OGC compliant catalogue services across all published geospatial data. pycsw has been extended and improved in terms of scalability, harvesting, and metadata schema support. Our contributions have been integrated in its codebase and power the geospatial catalogue services of data.gov. Also, pycsw has been extended to support the new OGC OpenSearch Geo/Time specification, towards implementation of CSW 3.0.
- *Vector data.* We have extended CKAN to natively support geospatial vector data management, by integrating PostGIS, the leading open source geospatial database. Data publishers can upload geospatial data in any format and coordinate reference system. The system automatically stores the dataset (*ingestion*) and can provide it in another data format (on-demand) or through OGC compatible services. As such, data publishers can provide any data they have at hand, without additional effort into transforming their data in specific-purpose formats. Further, as soon as the data is published, they are automatically available for querying and visualization with no extra effort.
- *Raster data.* We have extended CKAN to natively support raster data management, by integrating rasdaman, the leading big raster data analytics server. Data publishers can publish any raster data they have available, ranging from a single satellite image, to thousands of orthomaps. The raster data are efficiently managed and are available for querying and visualization with no extra effort.
- *Geospatial processing.* We have extended CKAN to support complex and OGC-compliant geospatial processing capabilities, by extending ZOO-WPS, the leading open source OGC WPS server. As soon as vector and raster data have been uploaded by the data publishers in the catalogue, these can be automatically reused through standards-compliant WPS services.

Our work over the next year will focus on further improving all software components with the ultimate goal of introducing PublicaMundi technologies in a production setting. Geodata.gov.gr, the Greek open geospatial data catalogue will migrate to PublicaMundi software and provide value added services to data publishers, users and developers in Greece.

The PublicaMundi Consortium

- The “**Athena**” Research Centre is a research and technology body, founded under the auspices of the Greek Ministry of Development in 2001. Athena RC participates in PublicaMundi through the Institute for the Management of Information Systems (www.imis.athena-innovation.gr). IMIS was founded in 2007, with the mission to conduct research in the area of data management and large-scale information systems. In this small time-frame IMIS has been established as one of the leading EU research institutions in Big Data and Semantic Web technologies, i.e. the foundations of the Data Economy. IMIS is actively promoting the open data agenda in Greece, having founded geodata.gov.gr (the first open data portal in Greece) and by regularly organizing Open Data days and hackathons in order to democratize open data technologies and tools. IMIS is at the forefront of open innovation and active participation in Greece through several other technical and policy initiatives.
- **RASDAMAN** (www.rasdaman.com) is a high-tech SME (rasdaman.com) and research spin-off specialized on Big Raster Analytics, centered around the first and leading Array Database, rasdaman (an OSGeo incubation project). The rasdaman technology, developed over a series of EU FPx and ESA projects, provides fast, scalable, flexible and open standards-based analytics on multi-dimensional raster data in Earth, Life, Space science and engineering. In close cooperation with Jacobs University it transforms leading-edge research results into marketable innovative products.
- **GEOLABS** (www.geolabs.fr) is a high-tech SME providing geospatial management solutions world-wide based on open source technologies. GeoLabs is leading the development of numerous open source geospatial projects and provides SaaS/PaaS

services for geospatial data. GeoLabs has developed MapMint (mapmint.com), a complete web based web mapping infrastructure, participated in the development of geoportail.fr, the development of the Senegal's National Cadastre. Further, it provides commercial cloud-based GIS services, and contributes in numerous open source projects.

- **GET** (www.getmap.gr) is a high-tech SME active in the entire lifecycle of geospatial data. It produces, maintains and curates geospatial data and develops high-value geospatial applications. GET has significant market experience, networking ties, and exploitation channels, with emphasis in cost effective open-source based services. Further, GET is the first Greek company of the GIS sector actively promoting and monetizing open source technologies in large-scale ICT projects.