3.1 Publishable summary

DaPaaS facts

Project:
Data Publishing through the Cloud: A Data- and Platform-as-a-Service Approach for Efficient Open Data Publication and Consumption

Project coordinator
Dr. Dumitru Roman
SINTEF
Forskningsveien 1, Blindern
0314 OSLO, NORWAY
email: Dumitru.Roman@sintef.no
Phone : +47 98283943

Partners from:


SINTEF

Ontotext AD

Swirrl IT Limited

Sirma Mobile JSC

Saltlux Inc.

The Open Data Institute

Duration: 24 months from 1.12.13

Total cost: ~2.1 M Euro

Programme: Objective ICT-2013.4.3 SME initiative on analytics

Further information: http://project.dapaas.eu/

While in recent years a large number of datasets have been published as Open (and often Linked) Data, applications utilizing these open and distributed datasets have been rather few. Reasons include, amongst others: the technical complexity and economical cost of publishing, interlinking and providing reliable access to the data; insufficient monetization incentives on the data provider side; lack of simplified and unified solutions for data consumption in a multi-platform way; lack of an Open Data "marketplace" where datasets and 3rd party components are available to application developers to reuse, combine and develop novel data applications.

At present Open Data publishers and application developers need to rely on generic Cloud platforms like Amazon Web Services (AWS), Microsoft Azure and Google App Engine, and build, deploy and maintain a complex Open/Linked Data software and data stacks from scratch. On the consumption side, efficient access to data is hindered by the lack of customizable user friendly interfaces to datasets and data-driven applications, interfaces that can easily be put together for

example for generation of data-drive portals that enable better access to data (e.g. for presentation, summarization, etc.) on various channels such as web and mobile.

To overcome the aforementioned issues, a unifying approach for a software infrastructure is needed, combining Data-as-a-Service (DaaS) and Platform-as-a-Service (PaaS) for open data and applications, complemented by novel mechanisms for cross-platform data access and consumption and an overall methodology, in order to optimize the publication of data and creation of data applications, eventually simplifying the reuse of Open Data.

The DaPaaS project[1] directly addresses such challenges by developing a methodology and an open software infrastructure providing data and application hosting capabilities for Open Data, with the aim of optimizing publication and consumption of Open Data as well as the development of data driven applications.

The DaPaaS platform aims to serve as an intermediary between:

- *Data publishers*, such as government agencies, who need an easy and economically feasible way of publishing and hosting Open Data;
- *Application developers and system integrators*, who need simple services and widgets for cross-platform data consumption (e.g. SMEs who need a reliable access to Open Data sets and services that can be integrated in various enterprise or consumer oriented scenarios);
- *End users* data consumers such as citizens, interested in new services and insights into open data.

At the same time, the feasibility and sustainability of the DaPaaS approach aims to be verified by several use cases in the hyperlocal domain, by utilizing and aggregating open data about local areas to offer new services to the public/locals.

The key target users and their relationships to the DaPaaS Platform are illustrated in Figure 1 below.

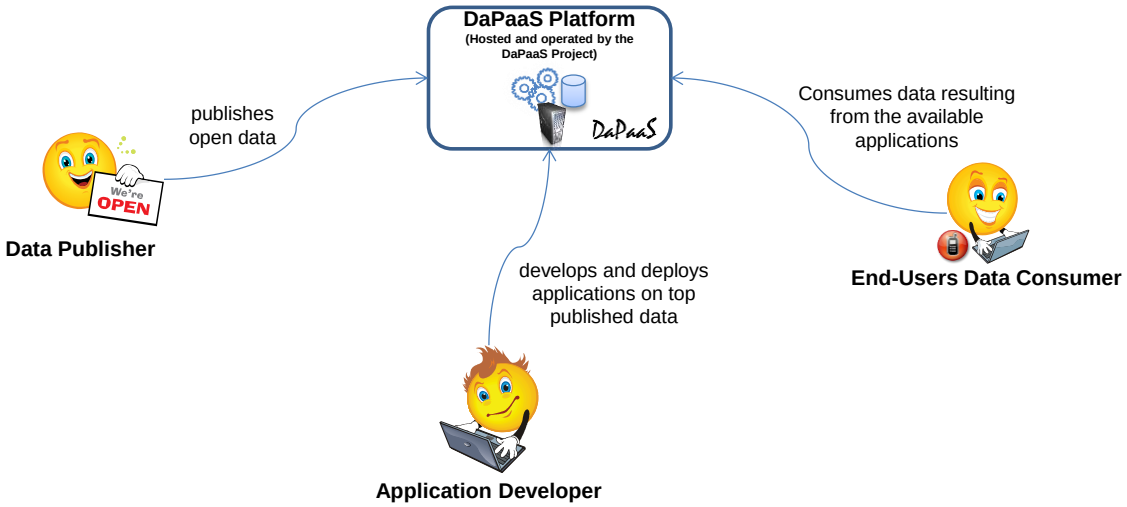

Figure 1. Key Roles in DaPaaS

[1] DaPaaS is funded by the European Commission under FP7 (610988); http://dapaas.eu/

The Data Publisher publishes data on the DaPaaS Platform which stores the data and makes it available for 3rd party application developers and end user data consumers. The Application Developer develops data-driven applications that use the data made available via the DaPaaS Platform. The applications are deployed and hosted in the DaPaaS Platform. End Users Data Consumers consume data resulting from the deployed applications.

The DaPaaS project is currently developing the DaPaaS Platform and will be responsible for the platform provisioning, operations and maintenance.

The platform consists of an open stack of components, which provide capabilities covering various aspects of the data management lifecycle, as depicted in Figure 2.

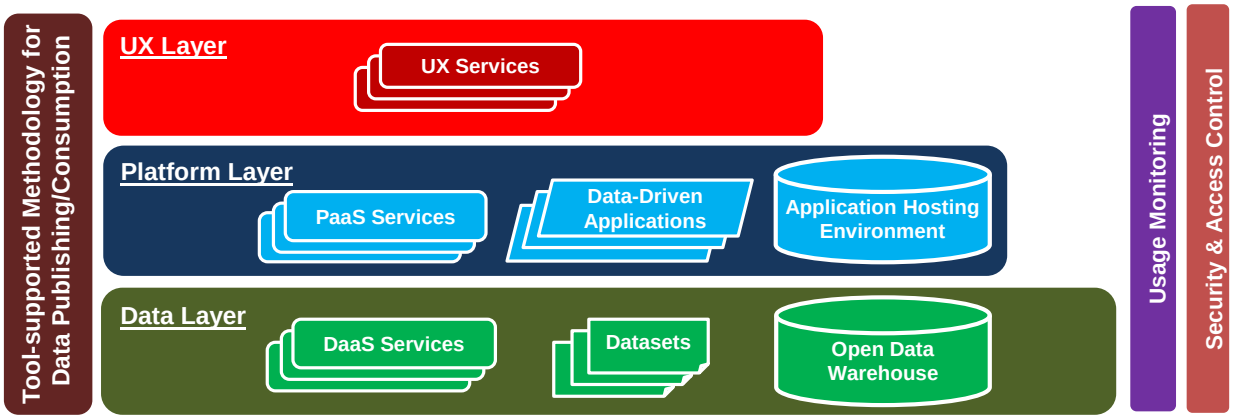

Figure 2. DaPaaS Layers

The **Data** layer provides capabilities for structured Open and Linked Data to be imported, stored, interlinked, and queried by 3rd party applications and services.

The **Platform** layer provides various infrastructure services for data transformation and cleanup, search and data analysis to application developers. The platform layer will be open for 3rd party application developers who want to co-locate their data analysis and processing components together with the data hosted on the platform.

Data-driven portals and mobile interfaces (**UX** layer) aim to simplify the task of developing a web portal or a mobile application for querying, exploration and consumption of datasets hosted on the platform. A hosting environment on a public Cloud will be provided, where data providers and application developers can publish their datasets and services for free.

During **Year 1** the project has developed and released the *first integrated prototype* addressing the open data publication process including support for data import, transformation, storage, querying, as well as application hosting. At the same time, the *PLUQI use case* has been defined in details and a set of requirements for the infrastructure have been provided.

During the first year, DaPaaS has taken a pro-active attitude towards dissemination and exploitation activities. We have formally presented the project and run tutorials at various European data/web focused events including the European Data Forum, Semantic Days, ESWC. In addition to formal presentations at events, members of the DaPaaS consortium have attended various open data events to raise awareness of the project. These have included OKFestival, ODI summit, Share PSI conferences/workshops. To engage with data users at a local level, we have presented the DaPaaS project at existing open data meet ups, such as the South West Data and Lotico London Semantic

Web communities. These two groups included attendees from the developers, academic and SME community. Furthermore, the ODI have organised a number of local meet-up groups, often coinciding with the location of project meetings. To date, these have included: setting up the first Sofia Open Data meet-up, hosting a meet-up in Manchester and hosting a sponsored formal evening reception at WWW2014, which attracted 60 high level attendees from government, industry and academia, with commercial sponsorship from Saltlux, ODI Seoul and Ontotext.

During this time, an open source collaboration has been initialized at an early stage of the project for some components of the DaPaaS platform, and is actively used by the members of the project. The open source components can be reached at: https://github.com/dapaas/.

http://dapaas.eu
@dapaasproject

DaPaaS

Vision

While a large number of datasets have been published as open (and often linked) data in the recent years, applications utilising these open and distributed data have been rather few, which can be explained by the challenges encountered in the area of data publication and consumption.

The DaPaaS (Data-and Platform-as-a-Service) project will deliver an environment where developers can both publish and host data-sets and data-intensive applications, which can then be accessed by end-user applications in a cross-platform manner. The resulting platform will reduce the complexity of open data publishing and consumption, thus making these activities easier and cheaper for SMEs and small public bodies which otherwise may not have sufficient expertise and resources to do so.

Data as a Service

The data layer (DaaS) will offer a publishing infrastructure that features components for large-scale data replication, versioning, analysis, and data access APIs, with primary focus on Linked Data (RDF).

Platform as a Service

The platform layer (PaaS) will provide efficient datastore access, data import and transformation services, reporting services, data enrichment and linking, search and indexing, and caching.

The DaPaaS UX layer will provide the means for consumers to access, navigate and explore the data both through open data portals and mobile services. Publishers will have access to an additional interface for data integration, transformation, enrichment and clean up. To create the DaPaaS platform, the project will address various components and activities dealing with the life-cycle for open data publishing and consumption.

Impact and means of achievement

The DaPaaS platform aims at overcoming some of the main current challenges regarding uptake of Open Data publishing and consumption. DaPaaS will achieve:

– A reduction in the cost for SMEs and public organisations which lack sufficient expertise and resources to publish open data.

– A reduction on the dependency of open data publishers on generic Cloud platforms to build, deploy and maintain their open/linked data from scratch.

– An increase in the speed of publishing new datasets and updating existing datasets through the provision of a sound methodology and integrated toolset.

– A reduction in the cost of developing applications that use open data by providing an integrated platform where infrastructure and 3rd party value added services and components can be reused.

– A reduction in the complexity of developing applications that use open data by creating a set of cross- platform and mobile widgets and components utilizing the open data sets on the platform which can be used by application developers.

– An increase in the reuse of open data by providing fast and seamless access to numerous open data sets to the applications hosted on the DaPaaS platform.

DaPaaS at a glance

Project title:
"Data Publishing through the Cloud: A Data- and Platform-as-a-Service Approach for Efficient Data Publication and Consumption" (DaPaaS)

Project reference:
610988

Project co-ordinator:
SINTEF

Partners:
Stiftelsen Sintef (Norway)
Ontotext AD (Bulgaria)
Swirrl IT (UK)
Sirma Mobile AD (Bulgaria)
Saltlux Incorporated (Republic of Korea)
Open Data Institute (UK)

Duration:
24 months (start date 01/11/2013)

Cost:
€2,173,746 (EU contribution: €1,499,000)

Website:
dapaas.eu / project.dapaas.eu