

# proDataMarket: A Data Marketplace for Monetizing Linked Data

Dumitru Roman<sup>1</sup>, Javier Paniagua<sup>2</sup>, Tatiana Tarasova<sup>2</sup>, Georgi Georgiev<sup>3</sup>,  
Dina Sukhobok<sup>1</sup>, Nikolay Nikolov<sup>1</sup> and Till Christopher Lech<sup>1</sup>

<sup>1</sup>SINTEF, Pb. 124 Blindern, 0314 Oslo, Norway  
{firstname.lastname}@sintef.no

<sup>2</sup>SpazioDati S.r.l., Via A. Olivetti 13, 38122, Trento (TN), Italy  
{lastname}@spaziodati.eu

<sup>3</sup>Ototext AD, 47A Tsarigradsko Shosse, Sofia 1124, Bulgaria  
{firstname.lastname}@ototext.com

**Abstract.** Linked data has emerged as an interesting technology for publishing structured data on the Web but also as a powerful mechanism for integrating disparate data sources. Various tools and approaches have been developed in the semantic Web community to produce and consume linked data, however little attention has been paid to monetization of linked data. In this paper we introduce a data marketplace – proDataMarket – that enables data providers to generate, advertise, and sell linked data, and data consumers to purchase linked data on the marketplace. The marketplace was originally designed with a focus on geospatial linked data (targeting property-related data providers and consumers) but its capabilities are generic and can be used for data in various domains. This demo will highlight the capabilities offered to the providers and consumers of the data made available on the marketplace.

**Keywords:** data marketplace, data publishing, data consumption, data monetization, linked data

## 1 Marketplace Overview

The proDataMarket marketplace is a virtual space that connects providers of open and proprietary data. It was originally designed as a platform for sharing and monetizing linked property-related data (e.g., real-estate and related contextual data), though its software components are generic and can be used for data in various domains.

On one hand, the marketplace aims at making it easier for data providers to publish, distribute and eventually reach out to potential consumers of their data. On the other hand, it helps data consumers discover and easily access data published at the marketplace. Consequently, the technical platform of the marketplace is composed of the tools, services and infrastructure developed to support two types of users: producers and consumers, each of which has a dedicated area on the marketplace. Fig. 1 gives an overview of the marketplace services it provides to data producers and data consumers.

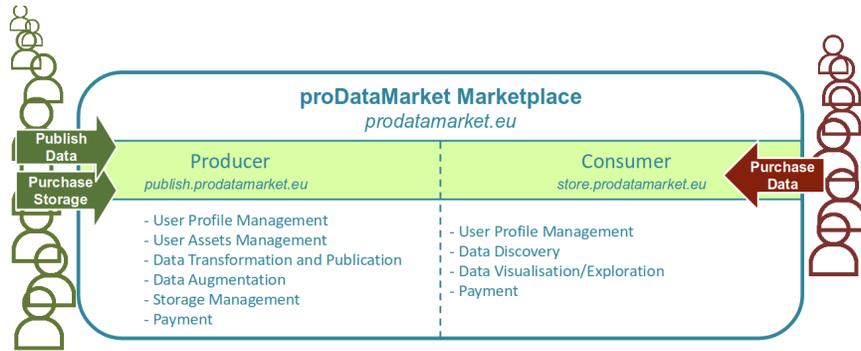


Fig. 1. Marketplace services overview

A high level overview of the marketplace architecture is presented in Fig. 2. Software components developed in the project were grouped either into Producer or Consumer areas in the marketplace, depending on whether they realize services for data producers or data consumers.

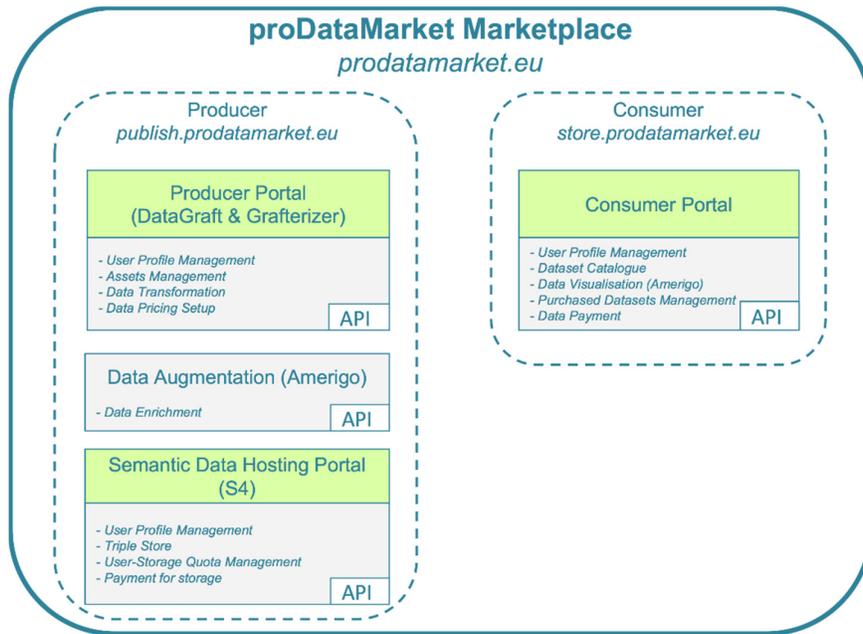


Fig. 2. Marketplace architecture – high level overview

User interfaces of the components (whenever present) are highlighted in light-green, while grey boxes summarize all the important user operations enabled through the components. Whenever the components were built on top of the existing products or services, the names of the latter are given in parentheses. All the components communicate with each other via a RESTful API. In the followings we briefly discuss the marketplace services offered to the producers and consumers respectively, and end with an overview of the planned demonstration.

## 2 Producer Services

The Producer Services are available via a user interface of the DataGraft platform [1][2].<sup>1</sup> DataGraft implements User Profile Management and Assets Management operations, where assets can be data files, queries, transformations or SPARQL endpoints. Data Transformation and Publication operations are provided via Grafterizer [3] – a framework for data cleaning and knowledge graph generation.

Data Augmentation allows data producers to enrich their data with contextual indicators. This functionality is currently available via the API implemented as part of the Amerigo Augmentation Engine developed by SpazioDati and deployed as a service. This service can be used to enrich a dataset that contains geographical entities with indicators that describe certain phenomena in the given area. The indicators are computed from contextual databases, such as OpenStreetMap<sup>2</sup> used by default or a custom data source provided by the user.

The data hosting payment services and associated user interfaces belong to the area in the marketplace where the data producer can “reserve a place” in the market. In particular, the system asks the data producer to provision a hosting space by, first, requesting and authorizing payments for it and then paying for it on a subscription basis. The data hosting component is currently based on Ontotext S4 triplestore as a service solution.<sup>3</sup>

## 3 Consumer Services

The Consumer Services are exposed to the end users through the Consumer Portal<sup>4</sup>. The Portal implements User Profile Management that regulates access to the data available in the marketplace. Not registered users have access to free Open Data and preview of proprietary datasets. Registration is required to purchase subscriptions and get access to parts of or full proprietary datasets. Data catalogue enables search on datasets and provides access to datasets’ metadata and available subscription options.

---

<sup>1</sup> <https://datagraft.io/>

<sup>2</sup> <https://www.openstreetmap.org/>

<sup>3</sup> <https://console.s4.ontotext.com/>

<sup>4</sup> <https://store.prodatamarket.eu/>

Amerigo Data Visualisation Service<sup>5</sup> allows users to explore data on a map through available visualisations prepared by data producers (e.g., choropleth or category map). The maps offer different types of interactive data filtering widgets, to facilitate exploration of different types of data.

Finally, the Data Payment component enables users to purchase data on the marketplace. The component implements subscription-based data access and supports various business models of different data producers. It communicates with the Data Pricing Setup component on the Producer side, to obtain vendor-specific configurations for each dataset.

## 4 Demonstration Outline

The demonstration will focus on an end-to-end scenario covering data provisioning and consumption on the marketplace, and will consist of two parts, one for the data provider, and one for the data consumer:

- *Data provider*: Set up a database in the cloud (configuration, payment), populate the database with data and create the queries through which the data will be served to the marketplace; configure data visualization to be advertised on the marketplace; configure payment/subscription options for the data; configure access to the dataset page on the marketplace;
- *Data consumer*: Search for data on the marketplace; metadata browsing, visual data exploration; data purchase.

The demo scenario will be related to selling/buying the mass transportation score in a given city, calculated per census cell (used as input to estimating value of real estate properties in the given city).

As of September 2017, the marketplace is publicly available via <http://prodatamarket.eu/>. Some of the components of the marketplace (e.g., DataGraft, S4) are also publicly available as separate components.

**Acknowledgements.** The work in this paper is partly supported by the EC funded project proDataMarket (Grant number: 644497).

## References

1. Roman, D., et al. DataGraft: One-Stop-Shop for Open Data Management. To appear in the *Semantic Web Journal (SWJ) – Interoperability, Usability, Applicability* (published and printed by IOS Press, ISSN: 1570-0844), 2017, DOI: 10.3233/SW-170263.
2. Roman, D., et al. DataGraft: Simplifying Open Data Publishing. *ESWC (Satellite Events) 2016: 101-106*.
3. Sukhobok, D., et al. Tabular data cleaning and linked data generation with Grafterizer. *ESWC (Satellite Events) 2016: 134-139*.

---

<sup>5</sup> Powered by CartoDB, <https://github.com/CartoDB/cartodb>.