4.5 - Technical toolset for open data competitions
Apps for Europe

Revision: version 1

Authors:

Pieter Colpaert (MMLab - iMinds - UGent)
Anastasia Dimou (MMLab - iMinds - UGent)
Ghislain Auguste Atemezing (EURECOM)
Raphael Troncy (EURECOM)

Project co-funded by the European Commission within the ICT Policy Support Programme
Dissemination level: Public

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Preface

If you read this, you are probably planning to organise an “Apps for” event and you are looking for the least effort to set up a “Apps for” website. I hope that when reading this document, you will see that you have came to the right place. You will most probably also notice that we are not going to focus on a website alone: you also want a data interface for the hackers, you want to maybe create a new dataset and you want to publish the outcomes of your hackathon in a structured way. Hence the title of this document: technical tools, and not just a website.

The document is a result of organizing “Apps for” events for 3 years in a row and, as a result, feeling the pain of lacking some tools over and over again. We hope that these tools will help you with your hackathon and we hope that it may lead to more open tools for co-creation.

Pieter Colpaert and Anastasia Dimou

Executive Summary

This is the deliverable D4.5, “Technical toolset for Open Data competitions” of the Apps for Europe project. The deliverable is split into 4 parts:

1. A customizable website containing the necessary features for information, subscriptions and sponsors.
2. An API on top of the datasets, powered by The DataTank and the Datalift
3. A collection of tools for cleaning, modeling and interlinking data
4. Documentation about how to use the software Cataloguing Open Data applications

List of Abbreviations and acronyms

apps4x	Apps for X Co-creation Event Vocabulary
DCAT	Data Catalog Vocabulary
DVIA	The Data VIsualization Application Vocabulary
odapps	Open Data Applications Vocabulary
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SPARQL	SPARQL Protocol and RDF Query Language
URI	Uniform Resource Identifier

Introduction

One of the challenges yet to be tackled in the European Open Data Ecosystem is that reuse and co-creation competitions are not handled in the same way as the corresponding Open Data they are used to build on top of. While published Open Data sets are indexed and catalogued, there was no corresponding effort/interest so far to index and catalogue the applications built on top of these Open Data sets. Missing the

opportunity to easily discover and have an overview of the available applications already implemented either in the frame of the Open Data competitions or in general, means that new applications are re-created from scratch instead of reusing existing open sourced solutions and extending them. As the applications are not reused across silos, a lot of low quality applications are implemented offering the same functionality and at the end, they lack the opportunity to cause major impact.

Even though the idea of indexing open datasets and cataloguing them prevailed, corresponding catalogues of the various applications built on top of these datasets were not established to the same extent. Applications implemented in the frame of Open Data competitions, are not only not discovered but also not aggregated.

In the context of this deliverable, a Semantic Web based solution to aggregate such applications based on their data is introduced. A vocabulary is defined to semantically enrich the data around a competition, a concept to be implemented based on the available Open Data and the application implemented at the end. There are different ways to semantically enrich the data around a co-creation event, a concept introduced or an application implemented. In the frame of this deliverable, alternative possibilities are described and documented. Independently of the prefered approach to semantically enrich the co-creation events' data, the result is information enriched with semantics based on the vocabulary defined. Those descriptions are aggregated to a triplestore and an endpoint is available for the end users to search and query through the list of the available data.

Structure

This document suggests a backend and frontend infrastructure which is based on a combination of platforms to support an "Apps for"-event website. There are three platforms used: wordpress, The DataTank and The DataLift. Wordpress is a content management system (CMS) with a focus on blogs. On top of Wordpress we implemented a plug-in which allows one to publish the apps developed at their event as if they were posts. The data about the apps and ideas published by this plug-in will be well structured and will be reusable by other websites, so that for instance a future implementation of appsforeurope.eu will be able to aggregate all the apps and their business potential on one portal. This way, potential investors can scroll through the website and choose the best team for a certain idea and also hackathon participants can consult this list first to check whether code is reusable for their idea, or whether their idea is original enough at all.

Organizing a hackathon takes a lot of effort. Every effort should reach its full potential, thus the apps and ideas produced in a hackathon need to be disseminated to the fullest and need to be sustainable. In the Apps for Europe project, we are addressing this issue by providing business lounges: there is no better way to make apps sustainable than by providing them with a functional business model. This deliverable aims to present a technology stack with alternative options for each level to support the uppermost goal of the business lounge.

State of the Art

Open Data

Open data[1] is data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike.

Linked Open Data

Linked Open Data is a way of publishing structured data that allows data to be connected and enriched, so that different representations of the same content can be found, and links made between related resources[2]. According to the 5 stars of Linked Open Data[3] [4], advancing from Open Data to Linked Open Data is only a two-step effort. [ref-heath]

RDF Vocabularies

The Resource Description Framework (RDF) is a general-purpose language for representing information in the Web[5]. RDF is particularly intended for representing data about Web resources and represents information about things that can be identified on the Web. RDF supports the identification of things using Web identifiers (Uniform Resource Identifiers, URIs), and the description of resources in terms of simple properties and property values. RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "triple"). RDF represents resources as a graph of nodes and arcs representing the resources, and their properties and values. RDF has an

[1] http://okfn.org/opendata/
[2] http://pro.europeana.eu/web/guest/glossary
[3] http://www.w3.org/DesignIssues/LinkedData.html
[4] http://5stardata.info/
[5] http://www.w3.org/TR/rdf-primer/

XML-based syntax, called RDF/XML.

RDF Schema is RDF's vocabulary description language[6]. The RDF Vocabulary Description language provides a mechanism for describing this information, but does not say whether or how an application should use it. RDF Schema defines classes and properties that may be used to describe classes, properties and other resources. All things described by RDF are called *resources;* classes group resources. The members of a class are known as *instances* of the class. RDF properties represent attributes of resources and relationships between resources.

Triplestores and SPARQL

A triplestore is a purpose-built database for the storage and retrieval of RDF (Resource Description Framework) data in the form of triples.

The SPARQL Protocol and RDF Query Language (SPARQL)[7] [8], is a query language and protocol for RDF. SPARQL queries are executed against RDF datasets. A SPARQL endpoint accepts queries and returns results via HTTP.

A SPARQL query comprises from:

- the prefix declarations (PREFIX ...)
- a result clause (e.g. SELECT ...)
- the dataset definition (FROM ...)
- the query pattern (WHERE …)
- the query modifiers (GROUP BY/ ORDER BY etc.)

Publishing the co-creation event's data in an API

Using The DataTank

The DataTank is an open source tool that publishes data [ref-datatank]. The DataTank can read the data and publish it on the web using a URI as an identifier. It can then provide these data in any of the supported formats (CSV, XML, JSON, SHP, KML), no matter what the original data structure was. As not all datasets can be published as Linked Open Data since, firstly, there is no one-size-fits-all formula available, and secondly building linked data cycles at organization-level involves a long process, The

[6] http://www.w3.org/TR/rdf-schema/
[7] http://www.w3.org/TR/rdf-sparql-protocol/
[8] http://www.w3.org/TR/rdf-sparql-query/

DataTank publishes Open Data on three levels which are in line with the five stars of Linked Open Data:

1. **Non-machine readable data**. In this case, only data about the data can be stored. The metadata are indexed and published online so that the dataset becomes discoverable. The DataTank is integrated with CKAN to maintain a list of all the available datasets.

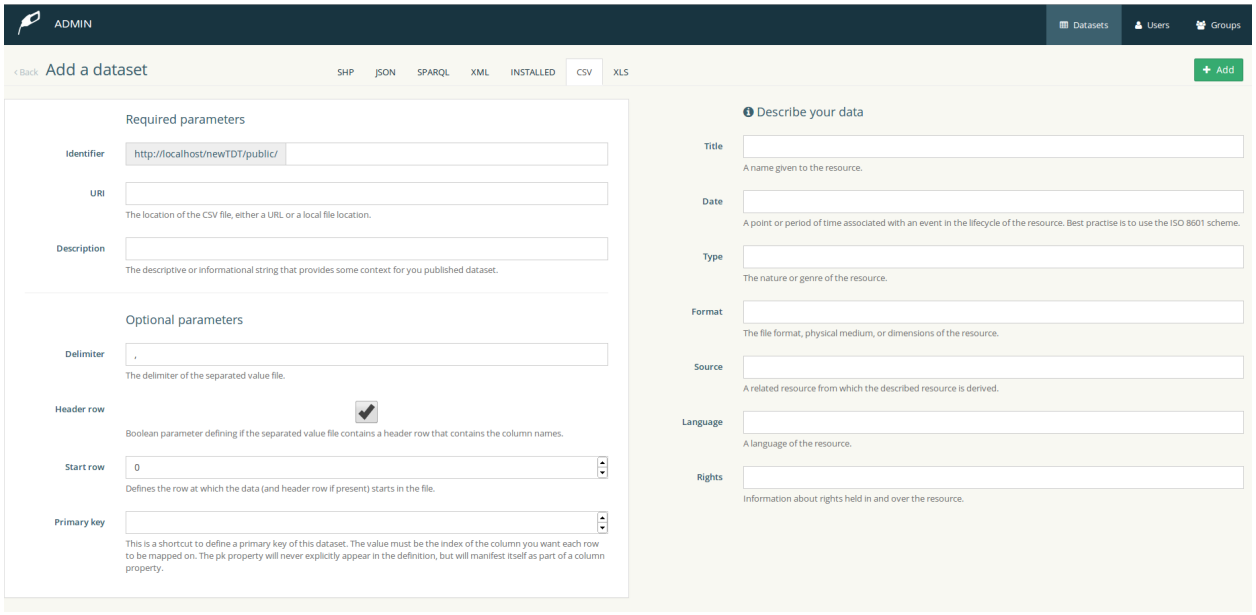
2. **Data about which only the serialization format is known**. This type refers to data available in a machine readable format. These data can be stored in text based files such as CSV, XML and JSON or in binary structures such as SHP files and relational databases. Here the DataTank, can provide extra features. It can adapt a machine readable file over an HTTP interface (URIs, HTTP caching, HTTP errors, HTTP headers, etc.) towards other machine readable formats.

3. **Data about which both the serialization format is known as well as the model**. In this case, an extra intermediate step will be taken. As long as the model is of interest to be captured, the model is interpreted in RDF and the data is semantically enriched. The generated RDF data is fed in triplesstores, accompanied by endpoints that serve as interfaces for querying the underlying datasets.

Publish Open Data to be used during the co-creation event. The DataTank is used to publish the data to be used at the co-creation event.
The former is described below, the latter is explained in more details at section. More details regarding The DataTank can be found at http://TheDataTank.com and regarding its installation at http://docs.thedatatank.com/4.0/create_definition.

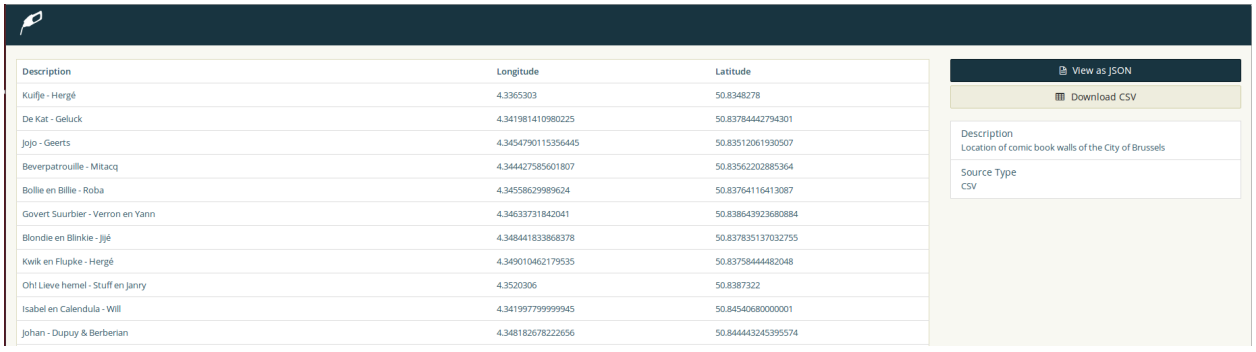
Publishing data with The DataTank

The DataTank publishes data on the web in a RESTful way. Its function is not based on copying or storing the data, nor will it act as if it owns that data. This means that the owner of the data stays in full control. For example if a data publisher wants to publish a CSV file, he needs to pass along the uri of where the file is, and what character is used to delimit the values in order to have a basic set of information that allows extraction out of the file. Details regarding how one can publish data using The DataTank can be found at http://docs.thedatatank.com/4.0/create_definition.



Requesting data with The DataTank

The data consumers can provide enough information to read the data and can extract the data in different formats using the DataTank. When a user requests some data, The DataTank will read the data structure on the fly and return the data in the requested format (e.g. JSON, XML, PHP, ..). Furthermore, The DataTank deals successfully with the burdens posed by Large datasets that can't be fully maintained in memory, or are not meant to return in one HTTP response. To this end, paging has been implemented to allow the user to browse through the data, while still getting manageable HTTP responses.



Using The DataTank to get the co-creation event's data in an API

In order to further support the co-creation event organisers to easily publish the data to be used at the co-creation event competitions, the co-creation event organisers can

use again the DataTank or any other mean that allows them to refer to the data afterwards. Therefore, the DataTank has twofold utilitarian value for the co-creation event organisers. The two alternative annotation scenarios and the Open Data publication support offered by the DataTank are discussed in details in this section.

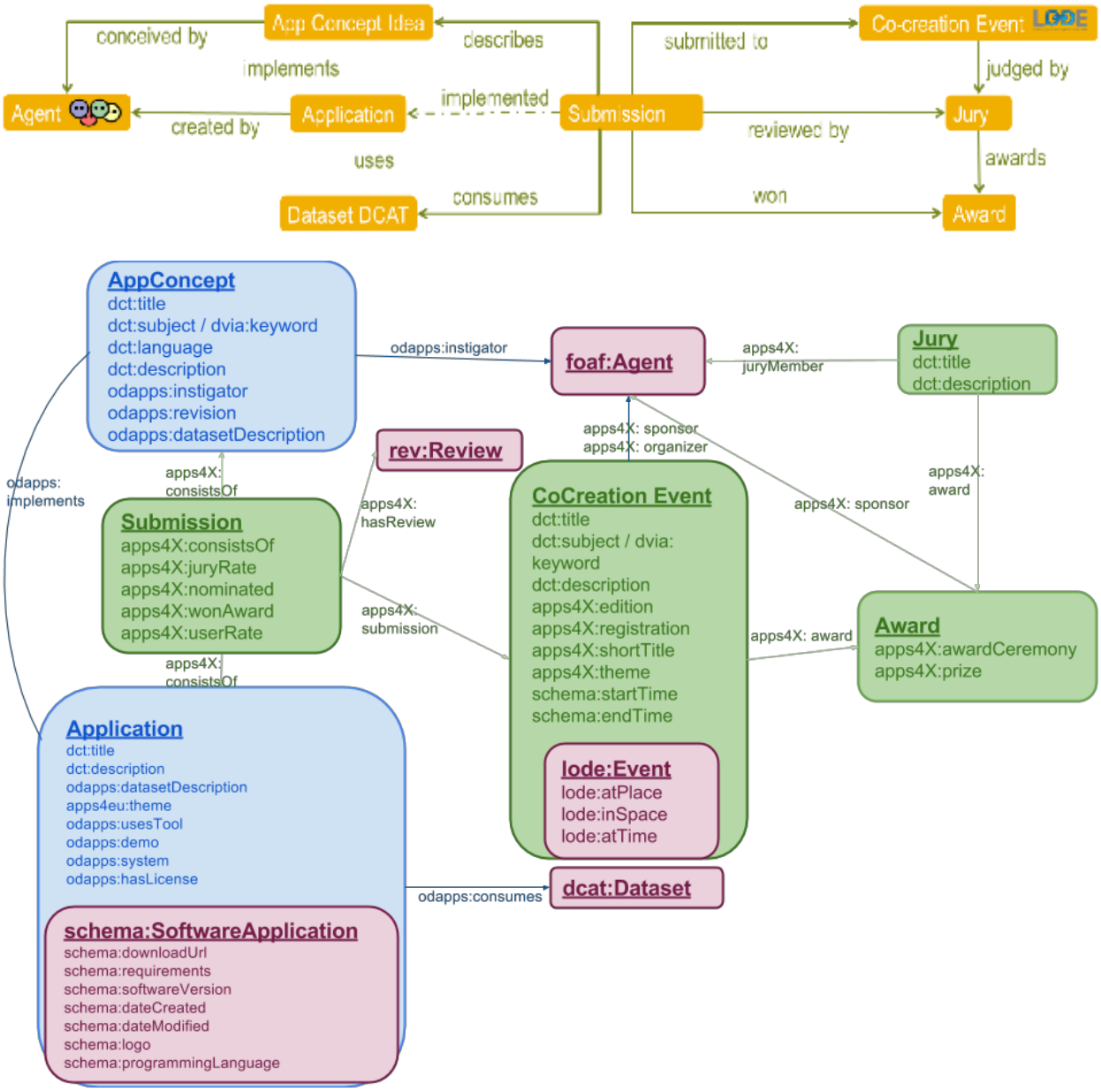
At a hackathon, participants need to get access to the data as fast as possible as time is limited. If the co-creation event organisers only provide proprietary formats or data dumps in certain formats, they slow down their participants, as they need to adapt to the data's format and pay more effort on parsing and processing them. Therefore they will have a less functional application at the end of the co-creation event.
When the data that is going to be used for the hackathon has been cleaned up, it can easily be added as a resource in a RESTful API using The DataTank. This way, developers do not have to write a lot of server-side code anymore as they can write their apps directly on top of the interface.

Apps for Europe vocabularies

Open Data applications are a particular kind of applications as they are built on top of data which are publicly available. Their description includes data about the data which is used for each application. Nevertheless, the same applications could be re-used with data of a similar (or not) context. Therefore their discoverability is of crucial importance, as they can be reused leading to more stable and richer applications that can cause a major impact compared to having several of a lower quality. On the other way around, it is equally important to be able one to find applications that were built on top of Data that s/he is interested in and find out how s/he can use them.
In order to support the aforementioned, we offer a solution based on Semantic Web technologies. We introduce new vocabularies, which can be used to describe the data of the applications and the links between them and the data they use. The new vocabularies are based on existing ones and are defined to describe Applications implemented using Open Data and *"Apps for X*" co-creation events.
There were two new vocabularies introduced. A vocabulary describing the different app concepts is defined (odapps), as Data Catalogue Vocabulary (DCAT)[9] describes datasets, with all necessary properties, which datasets it uses, what technology it has been written in etc. Another vocabulary is defined to describe the "Apps for X" co-creation events (apps4x) with all necessary properties, e.g. sponsors, criteria, app contest it was part of etc. The choice of having two separate vocabularies was made to assure that the odapps vocabulary will be re-used to describe applications built on top

[9] http://www.w3.org/TR/vocab-dcat/

of Open Data even if they were created out of the scope of an Apps for X competition.



Open Data Concepts and Applications

Open Data Concepts and Applications vocabulary (odapps) is defined to describe the ideas of applications on top of Open Data or the Applications built on top of Open Data. The vocabulary's namespace is http://semweb.mmlab.be/ns/odapps# and the prefered prefix is odapps. In this section, we provide a full description of the vocabulary.

Classes

App Concept (odapps:AppConcept)

This class denotes an idea conceived. The idea is based on Open Data and can be turned into an application.

Application (odapps:Application)

This class denotes an application that uses Open Data in order to implement an App Concept. The class extends the one defined at the visualization vocabulary for LOD applications[10] (DVIA vocabulary) which aims to represent only applications which visualize Open Data.

Tool (odapps:Tool)

This class denotes the tool or library that uses Open Data in order to implement an App Concept. The class extends the one defined at the visualization vocabulary for LOD applications (DVIA, http://purl.org/ontology/dvia# vocabulary) which aims to represent only applications which visualize Open Data.

Properties

conceived (odapps:conceived)

This property denotes the Agent which conceived the idea about the App Concept, therefore the Application too.

consumes (odapps:consumes)

This property is an extension of the property dvia:consumes and is used to describe the Open Data that are related to the Concept or to the Application.

description (odapps:description)

This property denotes the description of the idea.

dataset Description (odapps:datasetDescription)

This property denotes the description of the dataset used. In case one wants to learn more about the underlying dataset, s/he may read the description. This property extends the dataset description property of the DVIA vocabulary.

keyword (odapps:keyword)

[10] http://purl.org/ontology/dvia#

This property denotes the keywords that characterize the Application released and/or the Concept. This property extends the keyword property of the DVIA vocabulary.

license (odapps:license)

This property denotes the license the Application that is released and extends the dvia:hasLicense property.

instigator (odapps:instigator)

The Agent which conceived the idea about the App Concept (the instigator/initiator).

implements (odapps:implements)

This property denotes the App Concept that this application implements.

revision (odapps:revision)

This property denotes an App Concept which was revised and re-posed / extended.

revised (odapps:revised)

This property denotes the Agent who revised and/or extended an App Concept.

uses Tool (odapps:usesTool)

This property links to the tools or libraries used for the application and extends the uses of the tool property of DVIA vocabulary.

Co-creation events

The vocabulary "Co-creation events" is introduced to describe a co-creation event, a particular type of events where people are gathered to conceive and discuss on ideas around Open Data reuse and implement applications on top of this Open Data. The vocabulary's namespace is http://semweb.mmlab.be/ns/apps4X/ and the vocabulary's prefered prefix is apps4X.

Classes

Award

This class denotes the award offered by the jury to a certain App Concept and/or Application.

Co-creation Event

This class defines a co-creation event, namely a competition, a contest, a challenge etc. It is a subproperty of a lode:Event or schema:Event as it describes a certain type of

event.

Jury

This class denotes a jury evaluating the submissions to a co-creation event.

Submission

This class denotes an App Concept idea as it was submitted to a certain competition.

Properties

award (apps4x:award)

This property denotes one or more awards offered at this co-creation event.

award Ceremony (app4x:awardCeremony)

This property denotes the award' s ceremony event.

edition (apps4x:edition)

This property denotes the event's edition.

jury member (app4x:juryMember)

This property denotes the individual people that the jury consists of.

jury rate (app4x:juryRate)

This property denotes the rate given by the jury.

nominated (apps4X:nominated)

This property denotes an App Concept nominated for the award.

organizer (apps4x:organizer)

This property denotes the event's organizer.

prize (apps4x:prize)

This property denotes the prize offered at the co-creation event.

registration (apps4x:registration)

This property denotes a link to the registration for the event.

size (apps4x:size)

This property denotes the event's size in regard to its thematic area and/or locality.

sponsor (apps4x:sponsor)

This property denotes the event's sponsor.

short title (app4x:shortTitle)

This property denotes the event's short title and/or its abbreviation.

theme (apps4x:theme)

This property denotes the theme of the co-creation event.

wonAward (apps4x:wonAward)

This property denotes the award that this App Concept won.

users rate (apps4x:usersRate)

This property denotes the rate given by the users.

Semantic enrichment of concepts' and applications' data

It is important to have the data of the co-creation events, the ideas conceived and the applications implemented annotated with the vocabulary defined and described above. But, as this is a task that requires knowledge of the semantic web technologies and as this is not the case for most of the co-creation event organizers, we aim to help them automate this task. To this end, we propose alternative solutions that can help them with this task. The simplest solution would be to use the Wordpress plugin we developed. This plugin provides forms that the co-creation event organizers can use and applies semantic annotation to any value that is filled in. Then, the co-creation event organizers can submit the website to the Apps for Europe organizers and they, in their turn, can extract the enriched data and submit them to a pan-european triplestore where information about app concepts, applications and different co-creation events organised all over Europe are gathered.

As this solution can limit the co-creation events organizers to use a certain technology we provide alternatives. Besides this they can develop any technology to automate this semantic enrichment, the co-creation event organizers can gather the data in files of different formats and map this data, using one of the proposed alternative tools, to

turn them into RDF using the aforementioned vocabulary. Last, the co-creation event organizers can choose any other means to reach the corresponding RDF representation. As soon as a semantically annotated version of the co-creation event's data exists, the co-creation event organizers may ask the Apps for Europe organizers to include the new information on the pan-european triplestore so as to make this data equally discoverable as any other idea and/or application on top of Open Data.
In the following sub-section, we provide more information about the Wordpress plugin developed. Then, a description of the two alternative tools used to publish co-creation event data follows.

Wordpress plug-in

To minimise the administrative burden for app contest organisers, a Wordpress plugin is implemented. The plugin is available at https://github.com/mmlab/AppsForX for downloading. The plugin aims to facilitate co-creation event organisers to host a website with all the information needed easily populated. Using this Wordpress plugin, the co-creation event organisers can fill in forms that assist them to provide more structured information about their co-creation event. Furthermore, the organisers have all the required forms ready to fill in to describe the concepts and the applications which came of their co-creation event and then publish them on their website. The advantage of using this Wordpress plugin is not limited to assisting the organisers to populate their website with information regarding the concepts and applications.
The plugin automatically adds semantic annotations to the filled-in values using the aforementioned vocabularies and other existing vocabularies. The scope of this Wordpress plugin is to facilitate the co-creation event organisers to host their website for the co-creation event they organise and assist them in publishing the data of the conceived app concepts and the implemented applications on top of Open Data. Besides the Wordpress plugin co-creation event organisers can use any other hosting facility they prefer and they can choose their prefered mean to semantically annotate the data of the event, the concept and the application, before they publish them to the pan-european triplestore. To this end, we propose two alternative Linked Data publishing platforms which the co-creation event organisers can use to publish their data, DataTank and DataLift.

Installation and usage

A more recent version of this chapter can be found at:
https://github.com/mmlab/AppsForX

Requirements

Your server might vaporize if these minimum requirements aren't met:

- PHP >= 5.3
- WordPress >= 3.5

Installation

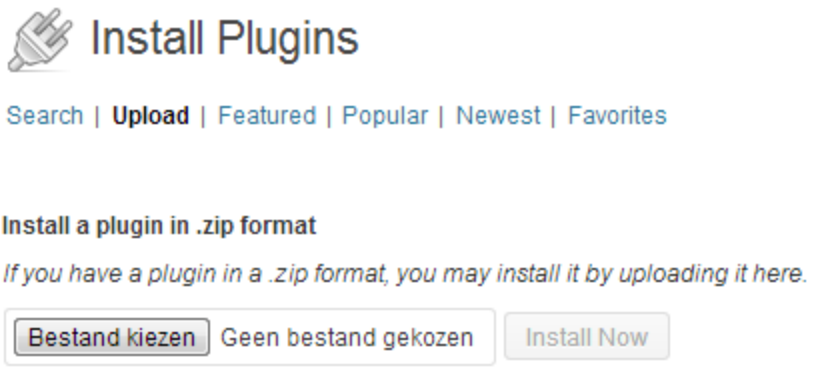
As an end-user download the latest version of the plug-in from this location: https://github.com/mmlab/AppsForX/archive/master.zip

If you are familiar with git and you want to add your own changes:

1. Clone the plug-in using: git clone https://github.com/mmlab/AppsForX.git
2. pull in the submodules bash git submodule init git submodule update

To use the plugin, follow these steps:

1. In the administration area of your WordPress site, go to the Plugins section, upload the zip file, and activate.



Install a plug-in by clicking "browse", selecting the .zip file and press install now

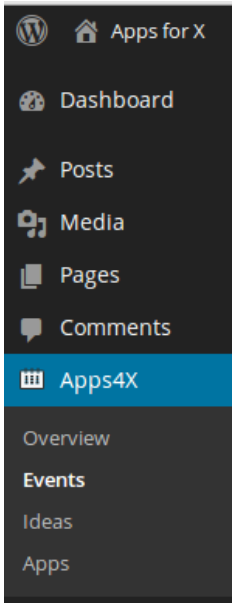
2. You can now add new events, ideas and apps directly from the administration area
3. If you want users to be able to submit ideas/apps, give them at least the "Submitter" role.
4. You may want to enable open user registration as well. You can find these parameters under Settings > General.

Enabling other users to submit apps or ideas

5. You can link from your theme files to the event pages. If you want event/ideas/app archives to show up in your main WordPress navigation menu, go to Appearance > Menus > Add a new menu, choose the wanted archive, then assign the menu to the primary location.

Functionality

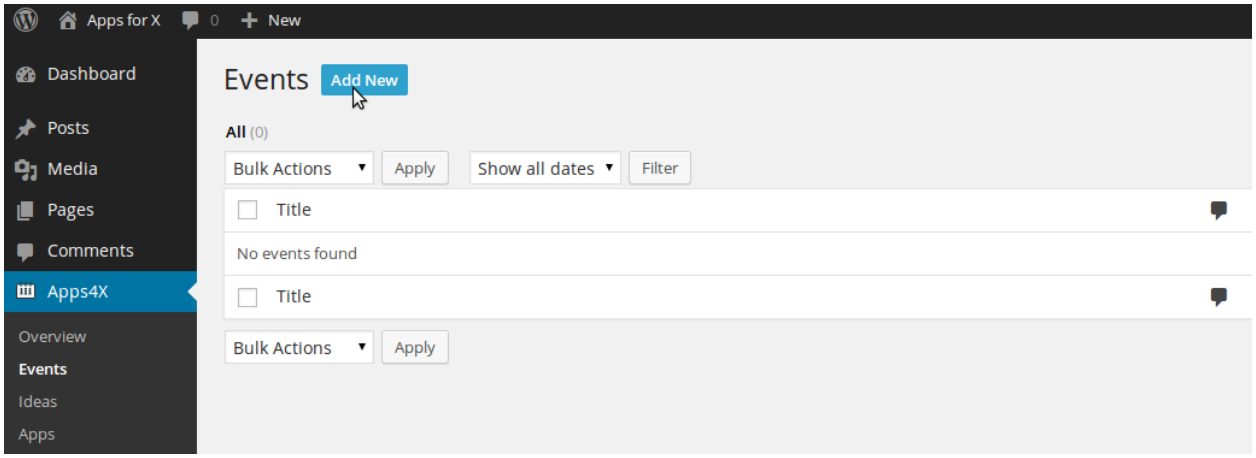
As soon as the plugin is installed, a new menu "Apps4X" will appear on the left side of the Wordpress admin page; there three subsections appear, Events, Ideas and Apps.



Add a new co-creation event

Once a co-creation event's organizer installs the wordpress and the Apps4X plugin, he can choose to add an Event. He can click on the "Events" option of the Apps4X menu. An overview of the events may appear. If it is a new installation, the list of events will be empty. If the same installation was used for a previous co-creation event, the specification of it will appear there. In general, the same installation may be re-used for several different competitions, thus the co-organizers which organize series of

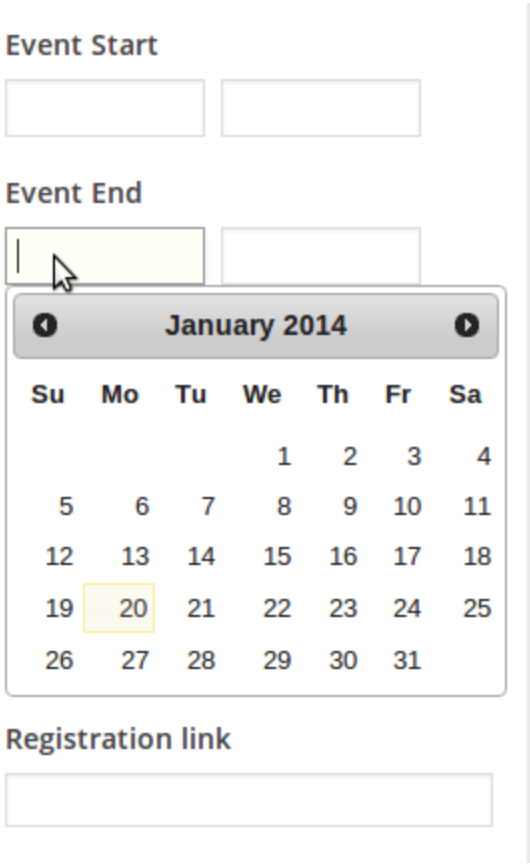
events are further facilitated.
From the overview list of the Events, the co-creation event's organizer can click on the button "New" to add a new co-creation event.



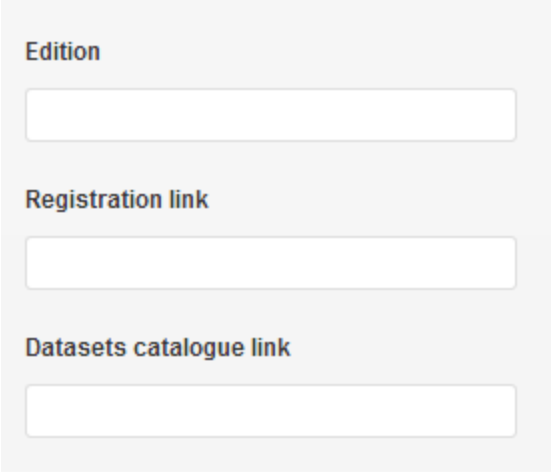
A new form ready to be populated with information about the new co-creation event organized appears.

On top the co-creation event organizer fills in the title of the co-creation event and below a description on any other text.
On the right side, the co-creation event organizer can add more detailed information regarding the co-creation event. S/he can add the event's logo either by selecting a local file or by drag and dropping a file/image in the corresponding box.
Besides the full title, the co-creation event organizer can add an abbreviated title for

shorter reference. If this is provided, it is used to define the URL of the page for the co-creation event.
Next, the organizer can add the duration of the competition by specifying the date the event starts and ends and the exact time if known. In order to support the co-creation event organizer to provide a valid date, a pop up calendar appears.

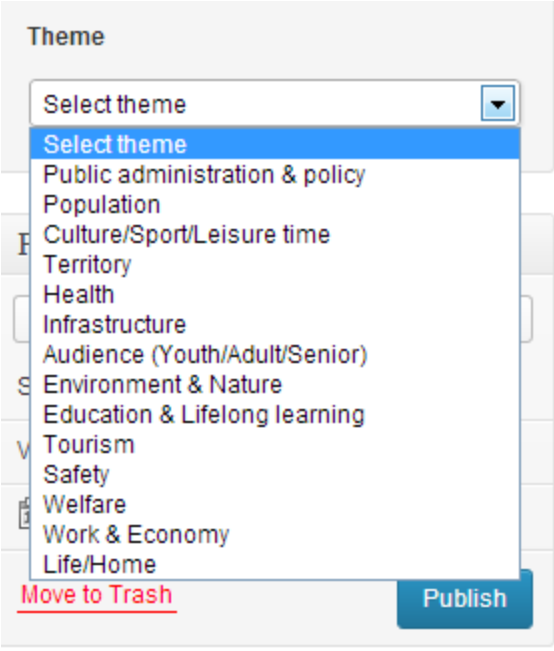


Next the organizer can add the edition of the competition. If the same event is organized occasionally, it has different editions and therefore a certain edition needs to be identified. e.g. Apps for Gent co-creation event is organized on an annual basis; the current edition is 3, but the next time the organizers adds a new page for the 2014's organization of the event, the number 4 will be filled in.



Furthermore, URLs pointing to a registration link can be provided, as well as a link to the web page with a list of datasets available for this co-creation event.

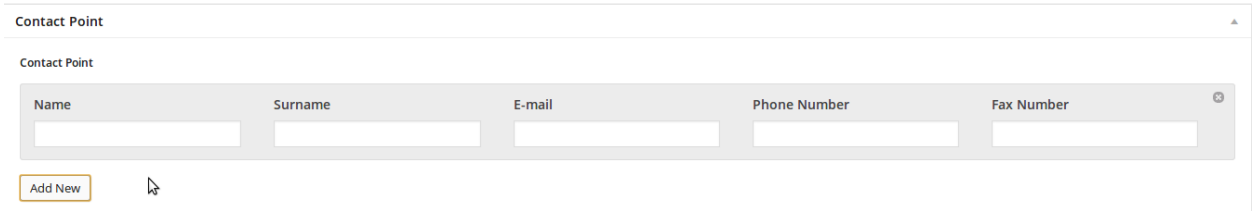
Last, the theme of the competition can be chosen among a list of very broad and generic theme categories. The co-creation event organizer can choose the most relevant theme category from a drop-down list.



More advanced details related to the event's contact points, organizers, sponsors, jury and awards can be added. These have the speciality that multiple entries of each one of them for a single competition can appear. In order to add more of them, the organizer

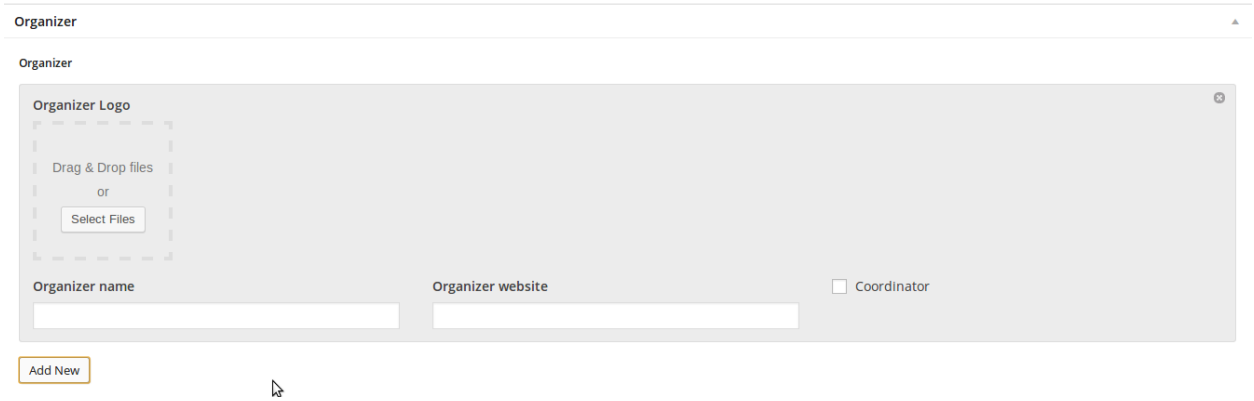
needs to click on the corresponding “Add new” button.

In the beginning, the organizer can add the contact points. He adds as many contact points as exist.



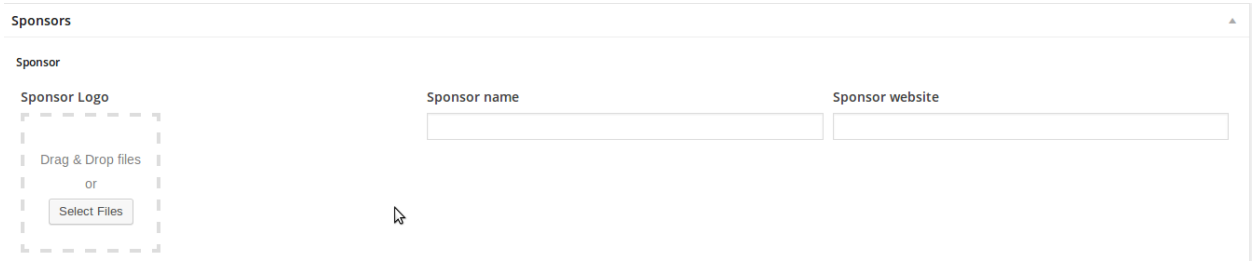
The organizer adds the name, surname, e-mail, the phone number and fax number of each contact point. None of the fields are mandatory to be filled in.

Then, he can add one or more organizers for the event.



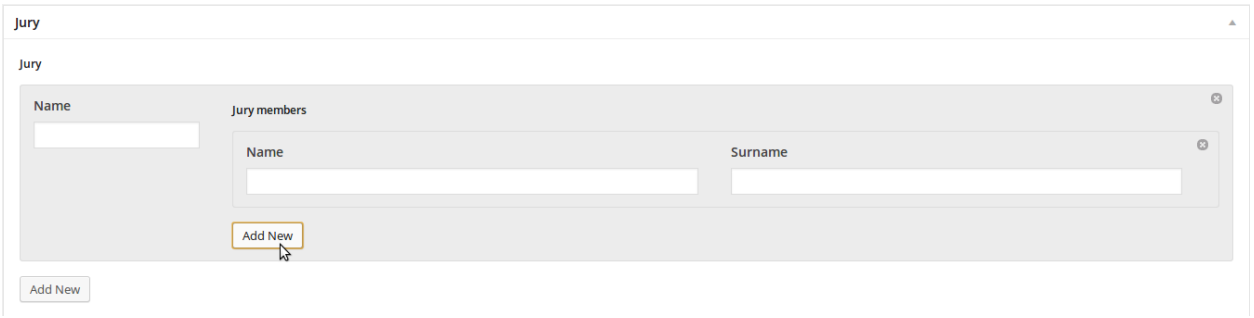
The organizer’s logo, name and website are provided. At the end, the organizer can choose whether this organizer is the coordinator or not.

Then, the different sponsors can be added.

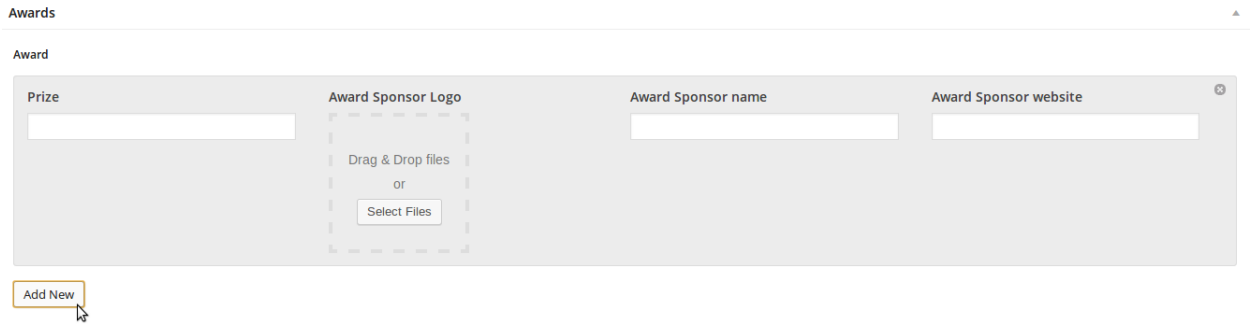


The organizer needs to add the name, the website and the logo of the sponsor. None of the fields are mandatory.

Finally, the information about the jury and the awards can be filled in. Firstly, a name of the jury needs to be added and its members. For each member the co-creation event organizer needs to add a new entry and fill in his/her name and surname.



Finally, the awards details can be filled in. The co-creation event organizer can name the prize and add information about the sponsor of the prize, if there is a specific one.



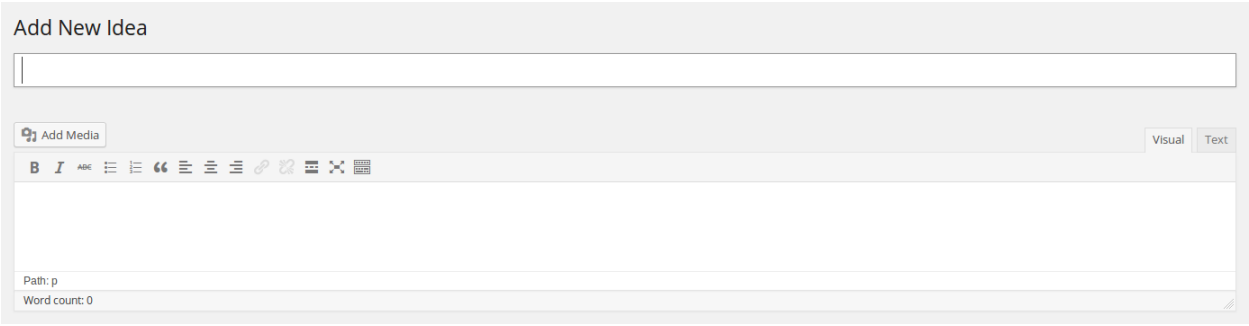
By the time everything is filled in, the user can click on the “Publish” button. A message will appear on the top of the page confirming the post’s publication and a link to view the the post will appear on the top of the page.

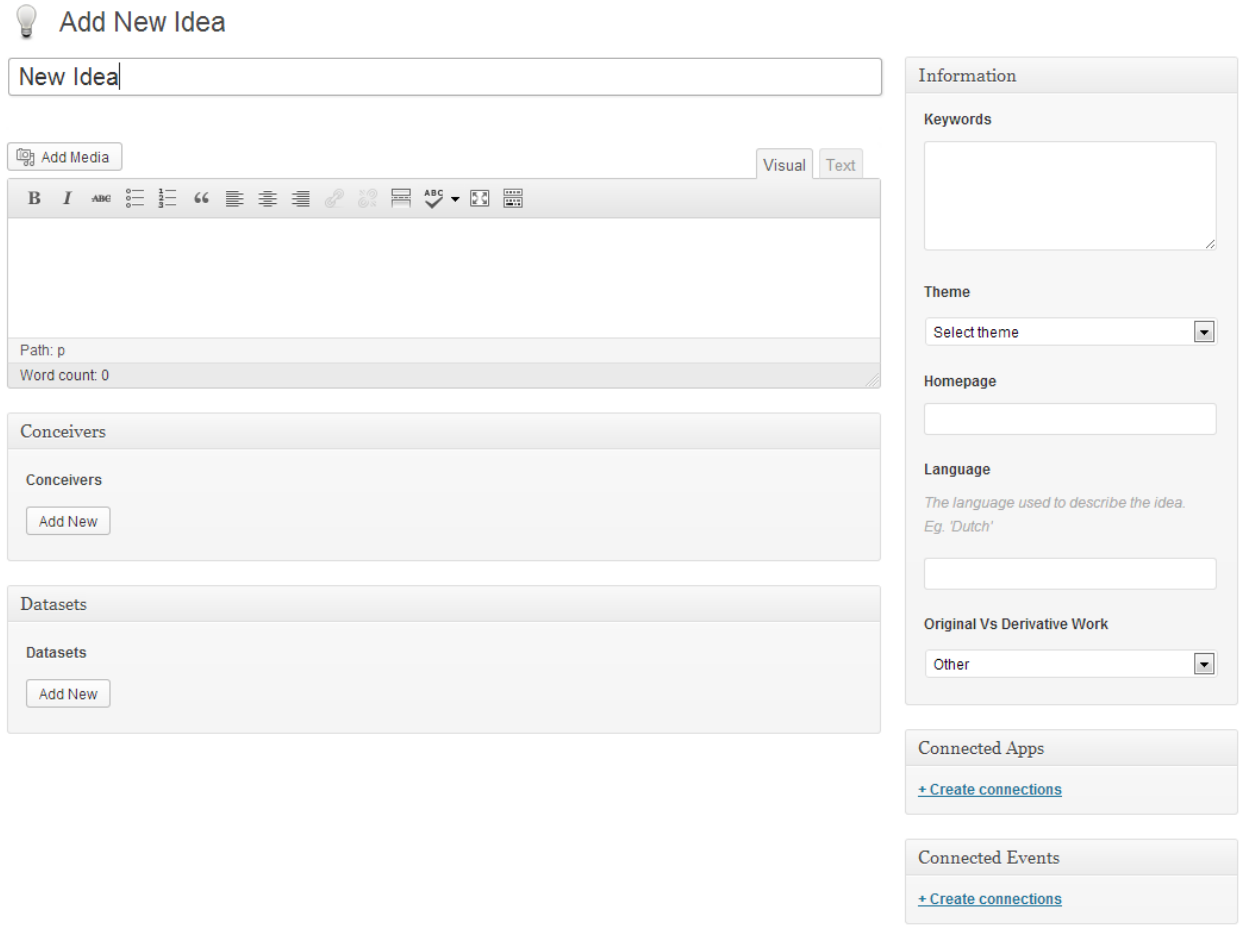
Submit a new App Concept

Besides automating the publication of the co-creation event’s details, the plugin can be used to publish App Concepts and Applications of Open Data either as a result of an Apps for X event or not.

A new App Concept can be added in the same way as a new Event. On the Apps for X specific menu, a new App Concept can be added by clicking on the “Ideas” option. The list of already added ideas will appear and a button to add a new will be present on top, exactly in the same way as for events.

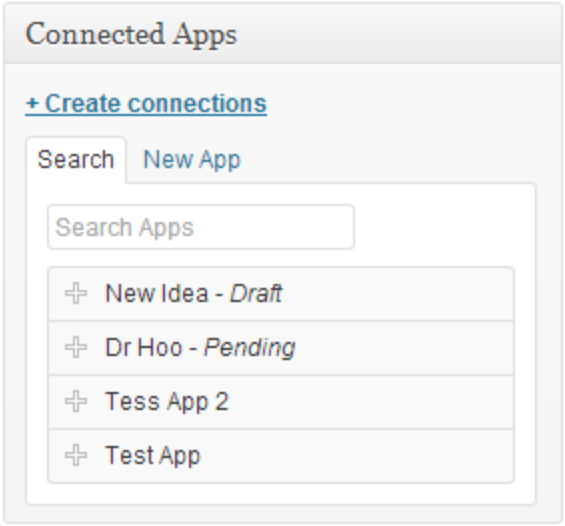
Adding a new App concept is very simple. The user who adds the concept needs only to fill in the title of the idea and the more detailed description of the idea.

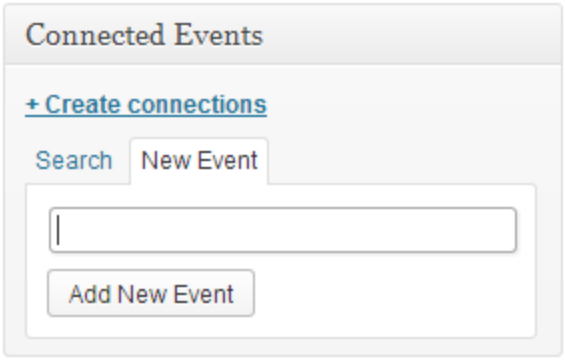




The user can add different keywords relevant to the idea (App Concept) and choose the theme under which the idea can be classified. The list of themes is the same as the list of themes for the co-creation events.

Besides adding the idea, one can link the idea with one or more applications and/or one or more co-creation events. In order to add a connection to a co-creation event, the user needs to click on the “create connections” link on the “Connected Events” menu. S/he can choose among the existing co-creation events and/or add a new one to link the new idea.





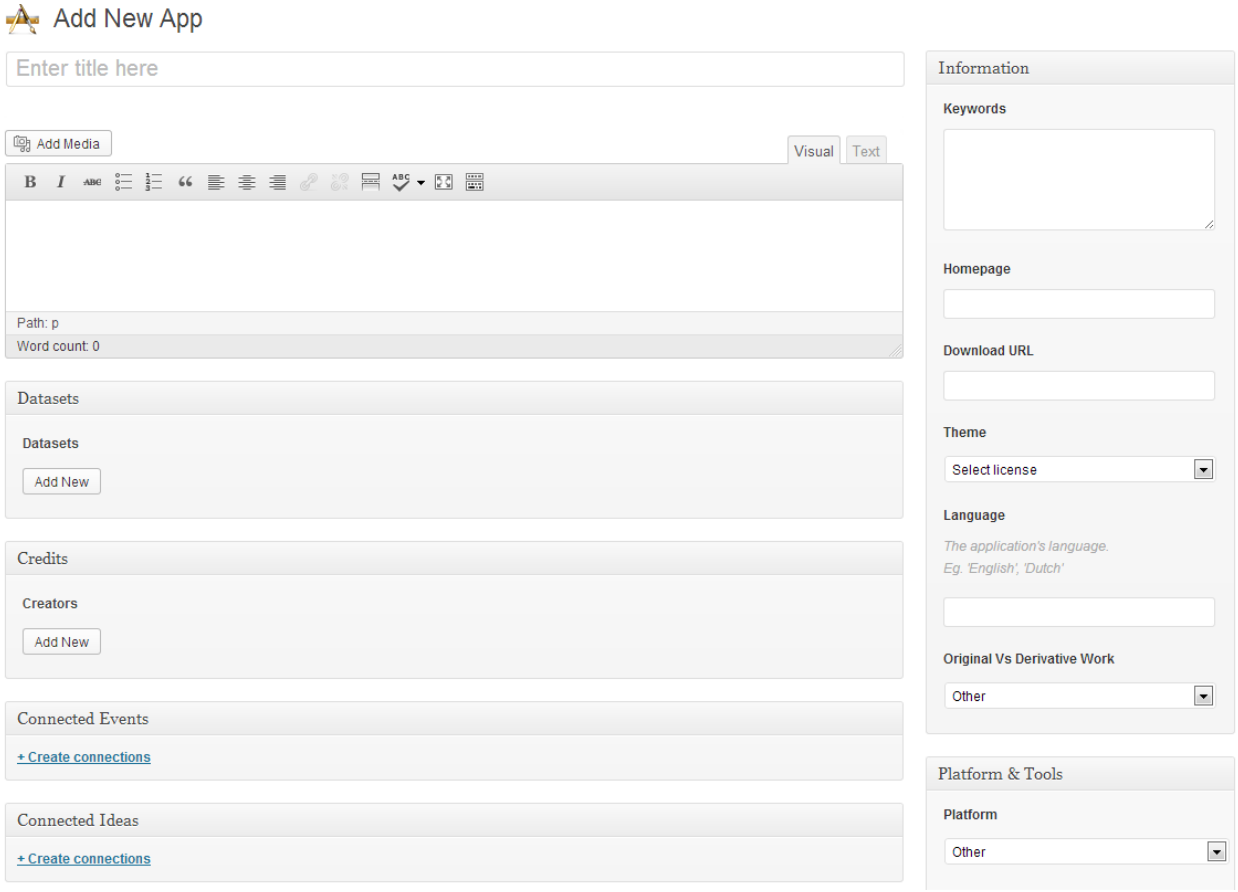
It’s exactly the same procedure to add a link to an existing or new application.

By the time everything is filled in, the user can click on the “Publish” button. A message will appear on the top of the page confirming the post’s publication and a link to view the the post will appear on the top of the page.

Submit a new Application

As an event and an App idea can be published, one can publish the data of an application built using Open Data either in the frame of an Apps for X co-creation event or not.

The user can add a new entry for a new application in the same way s/he adds an event and/or an idea. From the overview page of the "Apps", s\he can click on "Add new" and a new form to add an application appears.



As a user adds a new idea, s/he can add a new application. A new Application can be added in the same way as a new Idea, or Event. On the Apps for X specific menu, a new Application can be added by clicking on the "Apps" option. The list of already added applications will appear and a button to add a new will be present on top, exactly in the same way as for ideas and events.

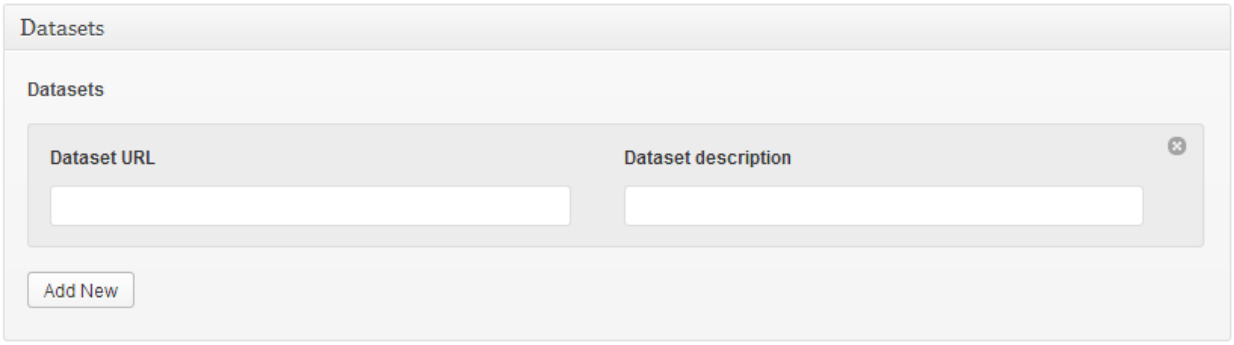
Adding a new Application is similar to adding a new Idea. The user who adds the application's data needs to fill in the title of the application and the more detailed description.

Then, s/he can add the keywords that described the application and the choose the theme under which the application can be classified. The theme classification is exactly the same as the one used for the events and the ideas.

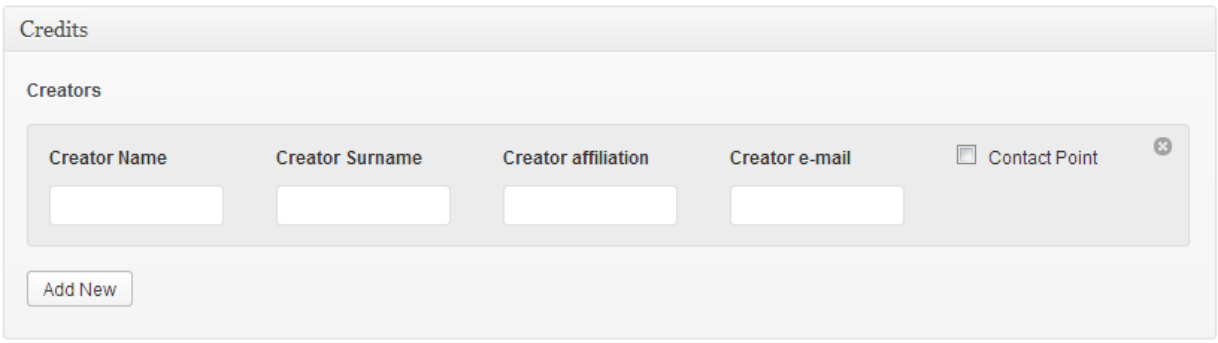
Furthermore, s/he can add the homepage of the application and a URL where a user can download the application.

Lastly the user who adds the application can indicate whether it is an original work or if it is a derivative work and the language that he used to describe his application.

Besides the aforementioned, the user can link the application to the datasets that were used for this application. Therefore, s/he needs to click on “Add new dataset” and then provide the URL and the description for each one of the datasets used.



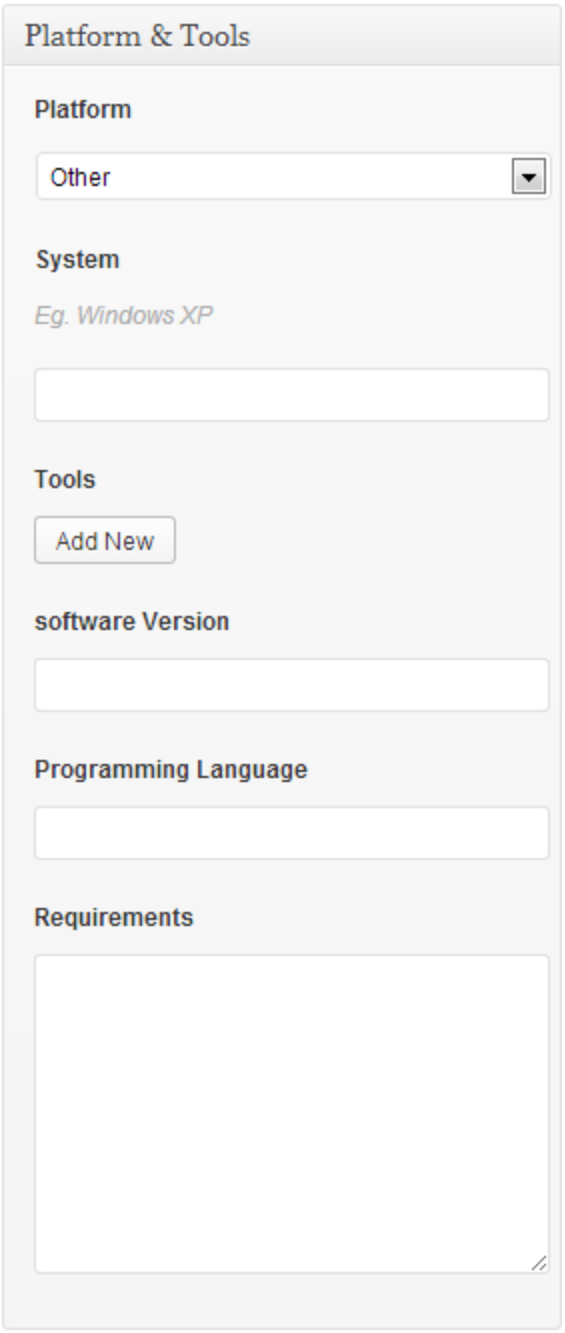
Further, s/he can add each one of the persons who were involved in its development. The user can add their names, surname, affiliation and e-mail and choose which ones are the contact points.



As in the case of ideas, the user can link the application to an idea and/or a co-creation event. In order to add a connection to a co-creation event, the user needs to click on the “create connections” link on the “Connected Events” menu. S/he can choose among the existing co-creation events to link the new application. Similarly, in the case of

ideas, the user can link the application to an existing idea or add an idea to link the application.

Finally, as an application is a piece of software, some more technical details can be provided. The user can take advantage of the “Platforms and Tools” section to do that.



The user can add the platform that the application works at, the system, the different tools used to have this application, the version of the application, the requirements to install the application and the programming language used to develop this application.

Publishing co-creation events' data as Linked Open Data

What to do if your website doesn't use Wordpress? If the co-creation event organizers can have a programmer, they can create their own plug-in (or hard code it). Otherwise they can use any technology they want to publish the data of the co-creation event and the ideas and the applications generated and afterwards semantically annotate them. Yet, then they still need to follow the vocabulary that we have defined and can be found at: https://github.com/mmlab/apps4eu-vocabulary. The vocabulary is described in details above.
In the case that the co-creation event organizers want to semantically enrich their data after they are published, we propose two tools: The DataTank and the Datalift. Details regarding each one can be found below.

Using The DataTank

The DataTank is extended not only to be a data adapter for publishing Open Data but to publish data as Linked Open Data too. A dataset and its content are identified by a unique URI, exploiting the REST interface and can be requested as RDF in multiple notations.

The DataTank executes the mapping from different file formats to RDF in four steps: extracting, transforming, mapping, loading and publishing. These four steps are formalised in tdt/input. For each of the steps, an action will be executed upon a chunk of data:

1. **Extraction**
   The DataTank can extract JSON files, XML files and CSV files in chunks of data. Those actions are fed to the stream and for each chunk the following actions (mapping and publishing) are executed consecutively and on a "per-chunk" basi.

2. **Mapping**
   If the model of a chunk is known (e.g., it's a GeoJSON point chunk), the chunk can be mapped towards another model. This is typically an OWL ontology and the output of the mapping phase are triples in RDF. At this phase, transformations can occur to process and customize certain fields: e.g. merge 2 fields, transform the units of a field (e.g., from yards to meter), executing a regular expression, and so on. The mapping rules are formalized using the mapping language Vertere which was initially defined to declare mappings from data in CSV files to

their RDF representation.

3. **Loading**
   A loader ingests each chunk of data which is mapped into a certain store.

4. **Publishing**
   After the data has been put in the store, The DataTank still has to be configured with the right parameters to provide an API on top of this store.

A semantifying sequence consists of an extract, a map and a load part. These parts need to be configured in a *job*. This job thus contains all the information that is needed to extract data from a certain data source, map the extracted data based on a mapping file, creating triples and loading these triples into a triple store.
Similar to the core, where the user would add resource definitions to the tdtadmin/resources package, s/he will have to use the input endpoint to create new jobs which is located at *input* (e.g. http://foo/input). Below are some examples of job configuration creations.
Ex. 1: Adding a CLI loader
A command line interface loader, without mapping, can be used to check what the output of a job will be. This can be used to create the mapping file for the next example.

Using The Datalift

Datalift is an open source platform [ref-datalift] helping to lift raw data sources or legacy data to semantic interlinked data sources.
The ambition of DataLift is to act as a catalyst for the emergence of the Web of Data by providing a complete path from raw data to fully interlinked, identified, and qualified linked datasets. The Datalift platform supports the following stages in lifting the data:

1. Selection of ontologies for publishing data;
2. Conversion of data to the appropriate format (e.g., from CSV to RDF);
3. Interlinking of data with other data sources;
4. Publication of linked data ;
5. Access control and licence management.

Functionalities of the Datalift platform

The architecture of Datalift is modular. Several levels of abstraction allow decoupling between the different stages from raw data to semantic data. The dataset selection

allows us to identify the data to be published and migrate them to a first RDF version. The ontologies selection step asks the user to input a set of vocabularies' terms that will be used to describe the lifted data. Once the terms are selected, they can be mapped to the raw RDF and then converted to properly formatted RDF. The data is then published on the DataLift SPARQL endpoint. Finally, the process aims at providing links from the newly published data to other datasets already published as Linked Data on the Web. Figure XXX below shows the workflow of converting raw data into "structured" RDF data.

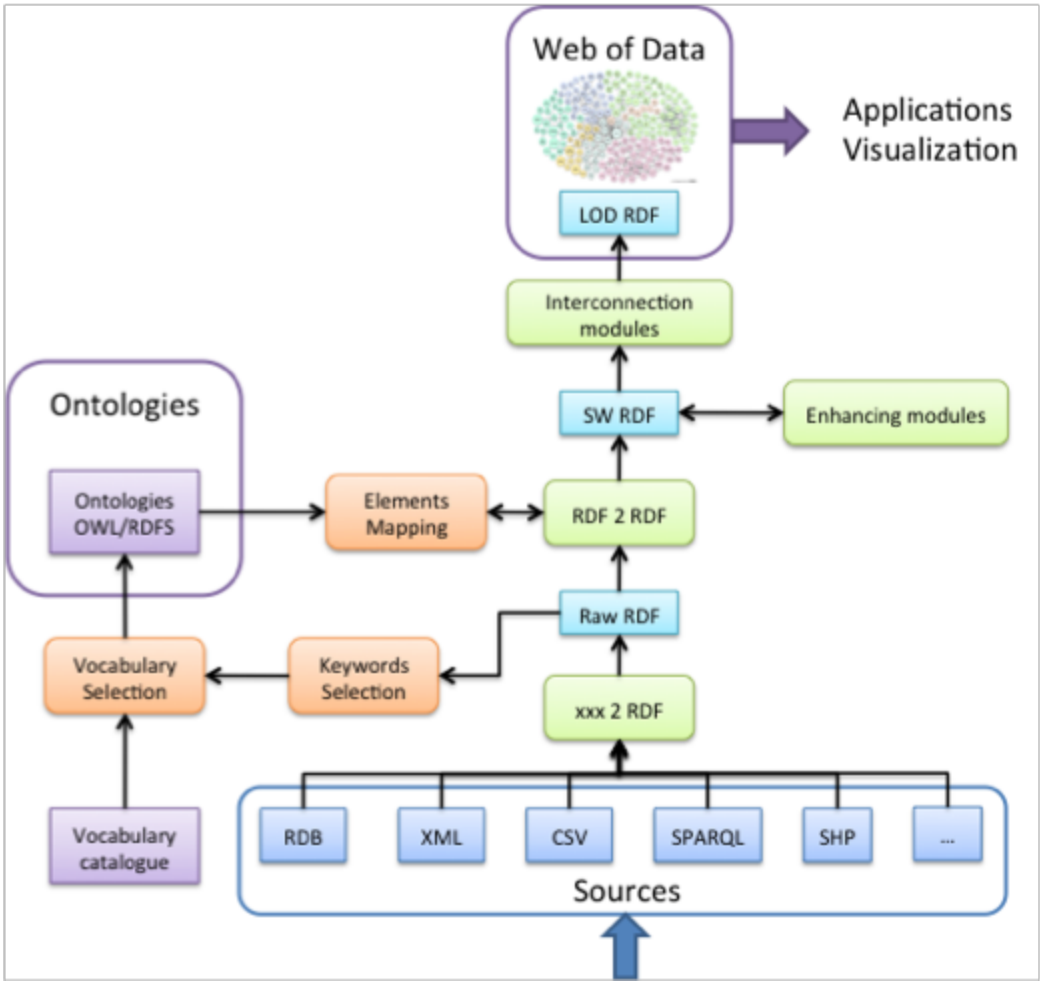


Figure xx: Datalift workflow, aka "lifting " raw source to RDF.

1. *Dataset Selection*

The first step of the data lifting process is to identify and access the datasets to be processed. A dataset is either a file or the result of a query to retrieve data from a datastore. The kinds of files currently considered are CSV, RDF, XML, GML and Shape

files. Queries are SQL queries sent to an RDBMS or SPARQL queries on a triple store.

2. Ontologies Selection

The publisher of a dataset should be able to select the vocabularies that are the most suitable to describe the data, and the least possible terms should be created specifically for a dataset publication task. The Linked Open Vocabularies[1] (LOV) developed in Datalift provides easy access methods to this ecosystem of vocabularies, and in particular by making explicit the ways they link to each other and providing metrics on how they are used in the linked data cloud. LOV is integrated as module in the DataLift platform to assist the ontology selection.

3. Data Conversion

Once URIs are created and a set of vocabulary terms able to represent the data is selected, it is time to convert the source dataset into a more precise RDF representation. Many tools exist to convert various structured data sources to RDF. The major source of structured data on the Web comes from spreadsheets, relational databases and XML files. We propose a two steps approach. First, a conversion from the source format to raw RDF is performed. Second, a conversion of the raw RDF into *"well-formed"* RDF using selected vocabularies is performed using SPARQL Construct queries. Most tools provide spreadsheet conversion to CSV, and CSV to RDF is straightforward, each line becoming a resource, and columns becoming RDF properties. The W3C RDB2RDF WG[3] proposes the Direct Mapping to automatically generate RDF from the tables but without using any vocabulary, and R2RML [http://www.w3.org/TR/r2rml/] to assign vocabulary terms to the database schema. In the case of XML, a generic XSLT transformation is performed to produce RDF from a wide range of XML documents. The DataLift platform provides a graphical interface to help mapping the data to selected vocabulary terms.

4-Data Protection

This module is linked to Apache Shiro for obtaining the information, *i.e.*, username and password, about the user who is accessing the platform. The module[4] checks which are the data targeted by the user's query and then verifies whether the user can access the requested data. This verification leads to three kinds of possible answers, depending on the access privileges of the user: some of the requested data is returned, all the requested data is returned, or no data is returned. This means that the user's query is filtered in such a way that she is allowed to access only the data she is granted access to. The access policies are expressed using RDF and SPARQL 1.1[ref-sparql11] Semantic Web languages thus providing a completely standard way of

expressing and enforcing access control rules.

5- Data Interlinking

The interlinking step provides means to link datasets published through the Datalift platform with other datasets available on the Web of Data. Technically, the module helps to find equivalence links in the form of "*owl:sameAs*" relations. An analysis of the vocabulary terms used by the published data set and a potential data set to be interlinked is performed. When the vocabulary terms are different, the module checks if alignments between the terms used by the two data sets are available. We use the alignment server provided with the Alignment API[5] for that purpose. The correspondences are translated into SPARQL graph patterns and transformation functions are combined into a SILK [6] script.

6- Data Publication

This module aims at publishing the data obtained from the previous steps to a triple store, either public or private. The providers can restrict which graphs can be accessible, they could decide whether to provide just a "Linked Data" or a "Linked Open Data". Datalift comes by default with Sesame[11], but provides API for connecting to Allegrograph, OWLIM, and Virtuoso triple stores as well.

Installation

Please find the documentation for installing Datalift at http://datalift.org/wiki/index.php/Platform_installation_(english). The latest version of the platform is available at http://datalift.org/en/node/24, which is still a work in progress until the mature and stable version by the end of March.

Usage

The data lifting has several distinct steps. DataLift makes it possible to replay each step in producing different results for each step. To facilitate access to all the different treatments and their results, they are grouped as one project. The project gathers together the various sources used and the results of all treatments done.
Each module has its own way to be used within the lifting process in DataLift. For more details, the readers are encouraged to read this resource at http://datalift.org/wiki/index.php/How_to_use_the_Datalift_platform_to_publish_a_data set_on_the_Web#The_lifting_project.

[11] http://www.openrdf.org/

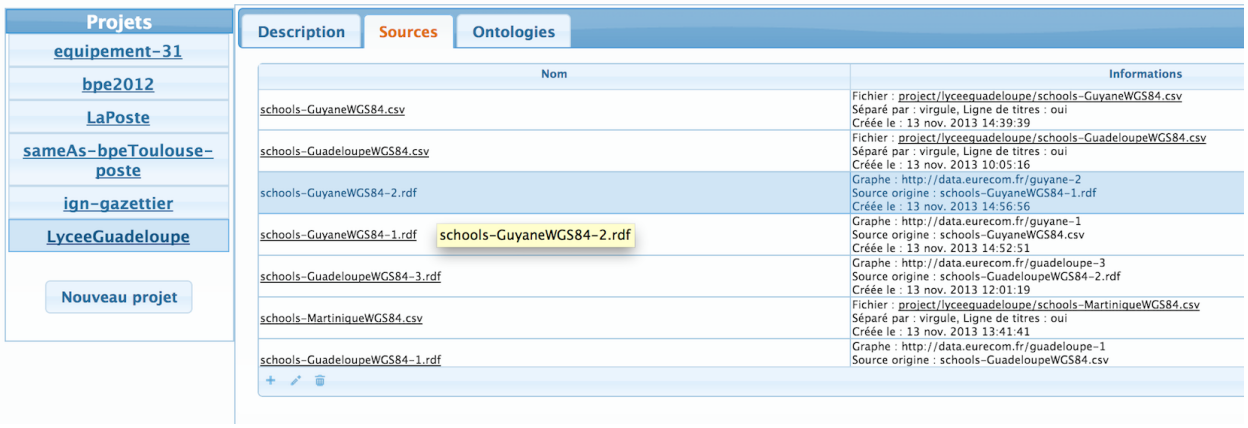


Figure xx: View of different projects on the left-side and the tabular views of a given project, consisting of the description, the sources and the ontologies used / or the be used in the project.

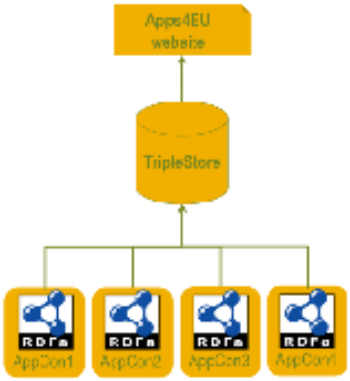
Figure XX: View of different tasks available for a project in Datalift with csv file as input.

Future Work

Until the end of the Apps for Europe project, EURECOM and iMinds will continue developing the started projects. Currently EURECOM is hosting the SPARQL endpoint

at http://apps4europe.eurecom.fr/sparql and the facetted browser at http://apps4europe.eurecom.fr/fct. This endpoint will be populated using an RDFa parser to convert websites into RDF using ontologies published within this project.

Harvesting mechanism and a pan-European triplestore



As all the Apps for X contests in Europe will use the same vocabulary, the published concepts will be aggregated and indexed. This triplestore will feed a pan European Apps for Europe triplestore, enabling to scroll through all the concepts submitted to any Apps for X event.

Publishing previous competitions data & enrichment with Apps4EU vocab

Automatic content enrichment with the Apps4EU vocabs

Automatically apply semantic annotations to the Events, App Concepts and Applications → Web pages for Events, App Concepts and Applications with automatic RDFa annotations

Automatic annotation of the app concept using RDFa,
ease the production of semantic data that describe the different App Concepts, Applications and Apps for X Events → Automatic Aggregation and Indexing of the applications based on their semantic annotations

Using the DataTank to publish the competitions' data

Example

At the second Apps for Ghent ever organized, we have installed one of the first versions of The DataTank. This is still functional and can be seen at http://appsforghent.be/api-tdt/.

Installation and usage

Please find the installation and usage documentation at http://docs.thedatatank.com.

If you want to integrate The DataTank with wordpress, we suggest installing The DataTank at http://data.{yourcontest}.com and link The DataTank from your main page.

References

[1] http://lov.okfn.org/dataset/lov/
[2] http://www.w3.org/wiki/ConverterToRdf
[3] http://www.w3.org/2001/sw/rdb2rdf/
[4] http://wimmics.inria.fr/projects/shi3ld/
[5]http://alignapi.gforge.inria.fr/
[6] https://www.assembla.com/wiki/show/silk/

[ref-datatank] Vander Sande, M., Colpaert, P., Van Deursen, D., Mannens, E., & Van de Walle, R. (2012). The DataTank: an open data adapter with semantic output. 21st International Conference on World Wide Web, Proceedings. Presented at the 21st International Conference on World Wide Web.

[ref-datalift] F. Scharffe, G. Atemezing, R. Troncy, F. Gandon, S. Villata, B. Bucher, F. Hamdi, L. Bihanic, G. K´ep´eklian, F. Cotton, J. Euzenat, Z. Fan, P.-Y. Vandenbussche, and B. Vatant. Enabling linked-data publication with the datalift platform. In 26th Conference on Artificial Intelligence (AAAI-12), 2012.

[ref-heath] Tom Heath and Christian Bizer (2011) *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

[ref-sparql11] Steve Harris, Andy Seaborne (eds) : W3C SPARQL 1.1 Query Language, W3C Recommendation 21 March 2013. Url: http://www.w3.org/TR/2013/REC-sparql11-query-20130321/