Open Data collection using mobile phones based on CKAN platform

Katarzyna Oświecińska (1,2)

(1) Orange Labs
CBR
ul. Obrzeżna 7
02-691 Warsaw, Poland
(2) Polsko-Japońska Akademia Technik
Komputerowych
ul. Koszykowa 86
02-008Warsaw, Poland
Email: k.oswiecinska@gmail.com

Jarosław Legierski (1,3)

(1) Orange Labs
CBR
ul. Obrzeżna 7
02-691 Warsaw, Poland
Email: jaroslaw.legierski@orange.com

(3) Warsaw University of Technology, Faculty of
Mathematics and Information Science,
ul. Koszykowa 75, 00-662 Warsaw, Poland

Abstract — **This paper presents concept and prototype of universal easy-to-use and self-configurable mobile application dedicated for open data sets collection and sending them to the central database storage based on CKAN platform. This article describes the concept of open data collection in crowdsourcing model using mobile phones and also an architecture of end user application (CKAN client) dedicated for Android operating system.**

I. INTRODUCTION

The concept of Open Data has been known in science and business environments for many years. The main definition of such kind of information describes that "certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control" [1], [2].

In the world there are many installed platforms, exposed open data sets produced by: cities, government institutions, scientific, and private companies. Large sets of this information being offered under open licenses such as mentioned in [3] are exposed in up to bottom model. Main part of such data comes from government institutions collecting and processing a large number of information about: their citizens, financial data or statistical information. Another idea regarding open data collection is concentrated on up to button approach. This model is based on idea of crowdsourcing and allows citizens to collect data.

Because of growing popularity of mobile devices (e.g. smart phones or smart watches) in recent years this kind of devices becomes the most convenient tool for collecting and aggregating open data sets. Therefore, dedicated, easy-to-use mobile application use mobile phone functions such as GPS receiver and photo camera. In many cases it becomes the best tool for the data acquisition using crowdsourcing.

II. EXISTING SOLUTIONS

On the Internet we can find large set of mobile applications dedicated to open data collecting. For example Yanosik [4] – the most popular, based on crowdsourcing model, mobile application in Poland. This service allows drivers to report information about dangerous situations on the roads: accidents, construction zones, police patrols locations etc. Other application users, basing on the data collected and exposed in this system, can adapt their behavior to local conditions and thus improve the safety on the roads. Another exemplary system based on crowdsourcing concept is mobile application 19115 [5] developed by City of Warsaw. This service (mobile application is only one component of the 19115 system) is dedicated to report non-emergency incidents in Warsaw. Residents using this service can report any information about the city (road damage, problems with garbage, problems which needs the intervention of the municipal guard, etc.). It could be noticed that applications presented above are dedicated only to specific use cases (road traffic reporting or processing interventions in City of Warsaw) and it is not possible to use them to collect any kind of open data. These solutions are dedicated to specific usage, not configurable and closed from systems integration point of view.

In literature we can find some open solutions dedicated to open data collection. Another type of mobile applications are those, mostly concentrated on processing and transferring open data from end user in universal form in which the scope and type of collected open data is configurable by the system administrators. The very good example is Open Data Kit [6]- set of tools which manages mobile data collection. Open Data Kit allows to build a data collection in survey form, collect the data

using mobile device and send data to the server. Tool supports also aggregation of data on the server side and visualize them.

In literature, there can be found also applications dedicated for CKAN platform [7] - open source framework de'facto standard in open data exposition. First one is Ukansearch - mobile web application to improve access and availability to open data. Unfortunately actually Ukansearch allows only to search using tags datasets on the www.data.ug portal [8] and doesn't support open data collection process. The second solution is sample CKAN Android client [9]. This tool was developed about 3 years ago and still exists only in beta version.

III. CKAN

CKAN (Comprehensive Knowledge Archive Network) – is open source data portal and open data exposition platform. This project was started in March 2006 [7] and is maintained by The Open Knowledge Foundation [10]. After 9 years of development CKAN is world-leading open data platform and there are more than 60 instances [11] of the system installed worldwide such as: data.gov - a portal of the United States government [12] British data.gov.uk [13], or the publicdata.eu European Union open data portal [14].
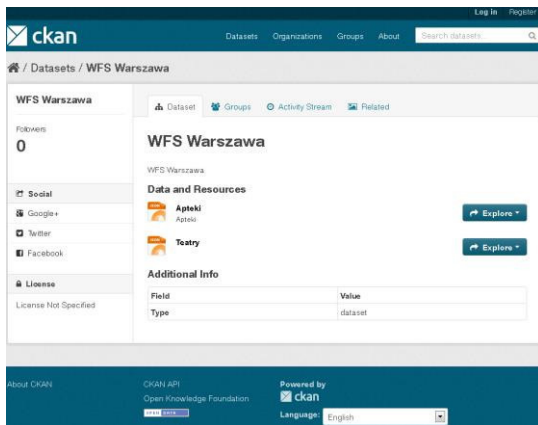

Fig. 1. Example data set in CKAN platform

CKAN platform allows for storage and management of open data repository, publication, search and visualize of data sets. System also allows developers to download and upload data sets via rich JSON based on API (Application Programming Interface). CKAN is open source software distributed under open license Affero GNU GPL v3.0.

IV. IDEA OF CROWDENABLER

Because currently there is no universal mobile application dedicated to open data collection using

CKAN portal the main goal was to build the concept, development and to test this type of application.

Crowd Enabler is an application dedicated to the collection of open data. With it, users of smartphones with the Android operating system can complement the individual datasets hosted on the CKAN platform. Native application function allows to add individual records to CKAN using geolocation information based on implemented in the phone GPS receiver. The main feature of the Crowd Enabler application allows to add objects containing text data, image files and the geographical coordinates by the end-user of the mobile phone. A user with the application, may, at any place (with enabled internet access and GPS location), take a photo of any object, add descriptions, and add its geolocation.
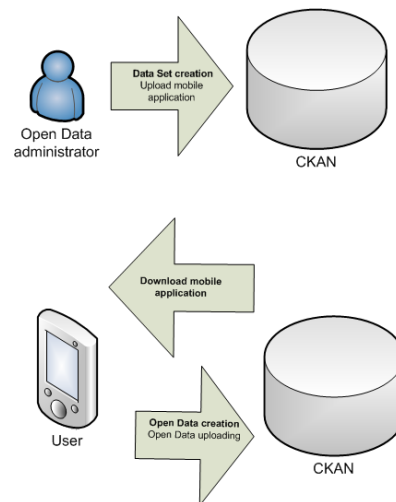

Fig. 2. Idea of Crowd Enabler

From the CKAN administrator and mobile application end user point of view the application usage looks as follows:

1) Data Set creation and upload Crowd Enabler.
In this step system administrator creates in CKAN an empty data set, prepares and uploads it as an CKAN resource. Then installs version of Crowd Enabler mobile application. The url to the Crowd Enabler is distributed (by e-mail or web page) to the end application users.
2) Download Crowd Enabler mobile application.
In this step application users responsible for open data collection download and install mobile application on their phones. It should be noticed that because of not publishing Crowd Enabler in Google Play application shop the option install from "Unknown sources" in mobile devices during this step must be turned on.
3) Data collection.
In this step application user can collect and store

data sets using mobile application. Application after start retrieves all information from CKAN platform using RDF metadata extension and reconfigures (changes names of the buttons and hints for editable fields). Application process open data in three following steps:

a) collected geographic location based on GPS receiver,

b) collected object description (in text form),

c) collected photo,

d) sending data to the repository (CKAN).

V. RDF ONTOLOGY USAGE

Resource Description Framework (RDF) [15] is a standard for describing web resources. RDF using some of the extension such as Data Catalog Vocabulary (DCAT) can be used for description of any resources on the Internet and used for building semantic Web concept. Semantic internet is the vision of Tim Berners-Lee and in data layer it's architecture is based on Linked Data principles. These principles are as follows [16], [17]:

• Usage URIs as names for things.

• Usage HTTP URIs so that people can look up those names.

• When someone looks up a URI, provides useful information, using the standards (RDF, SPARQL).

• Exposition links to other URIs, so that they can discover more things.

In presented in this paper application nonstandard RDF usage was proposed and implemented – this additional elements describing web resources were used for exposition of configuration data for mobile application.

In Crowd Enabler RDF extensions were used to define some application parameters. (e.g. Text box description etc.). Application uses RDF extension in Data Catalog Vocabulary (DCAT) standard implemented in CKAN platform.

Fig. 3. DCAT RDF extension in CKAN platform

In CKAN in format key-value there were 4 parameters defined and described in the table below:

Table 1. DCAT RDF parameters defined in CKAN dedicated for Crowd Enabler

	Parameter	description
1	LocationDescription	Task name / name of collected object
2	Hint	Collected object detailed description
3	PhotoButton	Task name (photo)
4	AppNAme	Name of application

Based on RDF XML extension in CKAN parameters presented above are accessible for developers during access to the data set via API using additional header in HTTP request **Accept: application/rdf+xml.**

```
<rdf:RDF      xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#"      xmlns:dcat="http://www.w3.org/ns/dcat#"
xmlns:dct="http://purl.org/dc/terms/">
  <dcat:Dataset
rdf:about="http://demo.ckan.org/dataset/monuments">
    <owl:sameAs      rdf:resource="urn:uuid:3aecf676-
e1cd-4eda-9e35-6805d50bafae"></owl:sameAs>
    <dct:description></dct:description>
    <foaf:homepage
rdf:resource="http://demo.ckan.org/dataset/monuments
"></foaf:homepage>
    <rdfs:label>monuments</rdfs:label>
    <dct:identifier>monuments</dct:identifier>
    <dct:title>monuments</dct:title>
    <dcat:distribution>
        <dcat:Distribution>
            <dcat:accessURL
rdf:resource="http://demo.ckan.org/dataset/3aecf676-
e1cd-4eda-9e35-6805d50bafae/resource/53ed83a3-8d89-
4c81-b7a9-
eaa55553ea5b/download/table.txt"></dcat:accessURL>
        </dcat:Distribution>
    </dcat:distribution>
    <dcat:distribution>
        <dcat:Distribution>
            <dcat:accessURL
rdf:resource="http://demo.ckan.org/dataset/3aecf676-
e1cd-4eda-9e35-6805d50bafae/resource/5f3441dd-8473-
4c44-af20-
de71af3c625c/download/photo.jpeg"></dcat:accessURL>
            <dct:title>photo</dct:title>
        </dcat:Distribution>
    </dcat:distribution>
    <dct:relation>
     <rdf:Description>
      <rdfs:label>1.
LocationDescription</rdfs:label>
         <rdf:value>1. Find me, I am standing by
the monument.</rdf:value>
      </rdf:Description>
    </dct:relation>
    <dct:relation>
      <rdf:Description>
         <rdfs:label>2. Hint</rdfs:label>
         <rdf:value>2.         Describe        the
monument.</rdf:value>
      </rdf:Description>
    </dct:relation>
    <dct:relation>
      <rdf:Description>
         <rdfs:label>3. PhotoButton</rdfs:label>
         <rdf:value>3.  Take  a  photo  of  the
monument.</rdf:value>
      </rdf:Description>
    </dct:relation>
    <dct:relation>
      <rdf:Description>
         <rdfs:label>AppName</rdfs:label>
         <rdf:value>Crowd Enabler</rdf:value>
      </rdf:Description>
    </dct:relation>
  </dcat:Dataset>
</rdf:RDF>
```

Fig. 4. DCAT RDF extension in CKAN XML dataset description

VI. SYSTEM ARCHITECTURE

This section of this publication contains the architecture and functionality of Crowd Enabler application. This mobile application is dedicated for Android operating system and was developed using Eclipse and Android Development Tools (ADT)
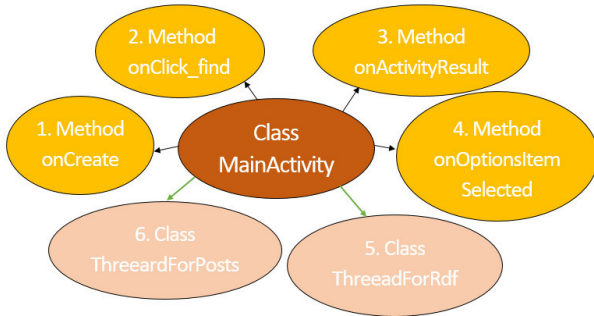
A) Programing environment

Fig. 5. Application class structure

The main methods and classes implemented in the application are as follows:

1. Method onCreate handles methods which are connected to buttons and starting ThreeadForRdf and ThreeadForPosts.

2. Method onClick_find sets GPS coordinates.

3. Method onActivityResult makes sure that our photo exists and is displayes in the app.

4. Method onOptionsItem Selected handles pressing Settings.

5. The main Method getRdfXml in Class ThreeadForRdf, gets useful informations from rdf about dataset. Using this, sets descriptions of buttons and editText.

6. Method makeRequestImage in Class ThreeardForPosts sends image to dataset. Method makeRequest sends full data (in JSON) of record to dataset.
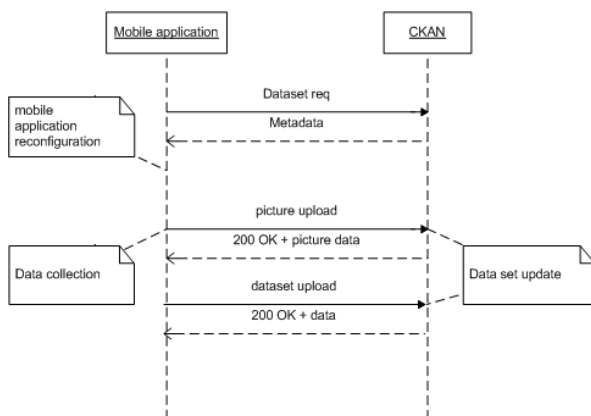
-

Fig. 6. Application data flow

Data flow between Crowd Enabler and CKAN platform is presented on Fig.6

1) Dataset request – application after launch based on defined http resource connects with CKAN Platform, downloads metadata and RDF extensions in key – value form used for Crowd Enabler configuration.

2) Crowd Enabler self-configures and shows GUI.

3) User takes photo and uploads it on the CKAN server as open data resource.

4) User performs an additional description of data record (text ,GPS coordinates), uploads and sets it to the CKAN.

B) Application GUI

End application user is a person who uses Crowd Enabler on it's mobile phone – and collects open data records. After start, application connects to CKAN and catches from RDF scheme of metadata to change names of buttons and hints defined in the application GUI.

When user wants to collect data using geolocation, must stand with his smartphone nearby an object which he wants to upload (e.g. memorial or monument).
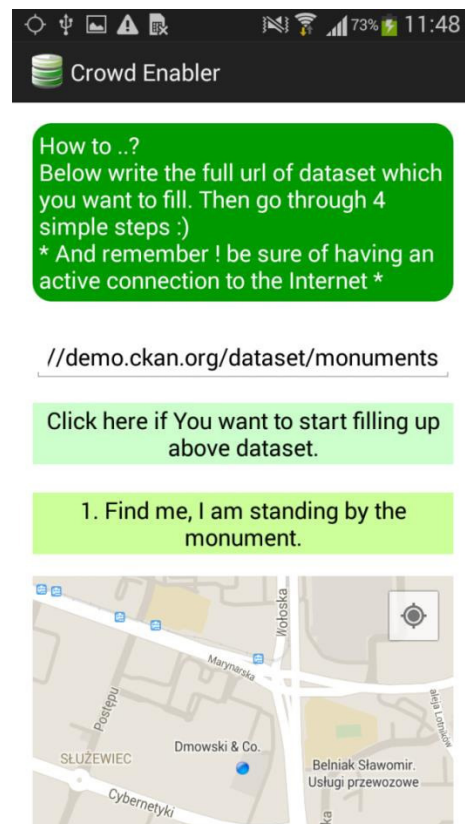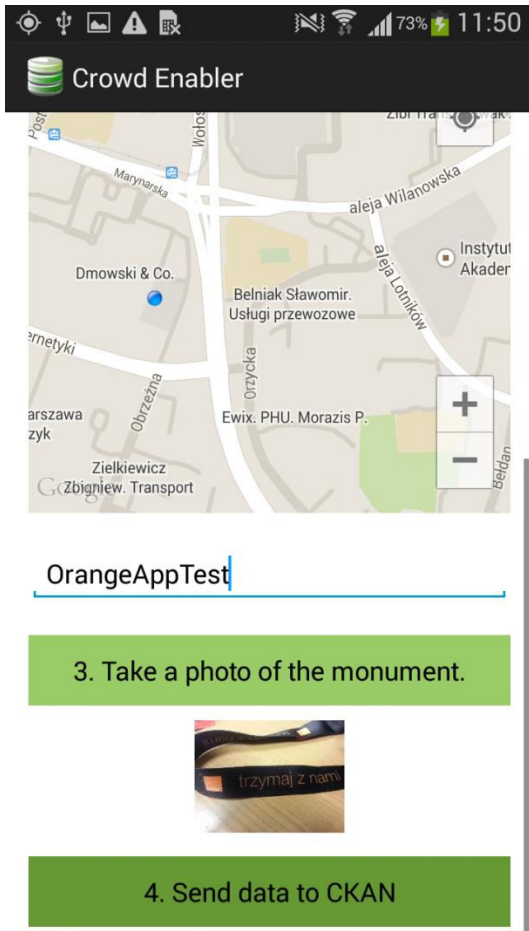

Fig. 7. Application GUI

Fig. 8. Application GUI

After taking photo and adding additional description end application user sends data to a server by pressing Send Button.

VII. MEASURENMENT

Using Crowd Enabler application several data sets were collected
1) monuments - historic buildings and monuments in Warsaw
2) forgottenPlaces - some interesting places in the city
3) Trees - the test group of trees
The Figures bellow shows the example data sets with forgottenPlaces and monuments in Warsaw
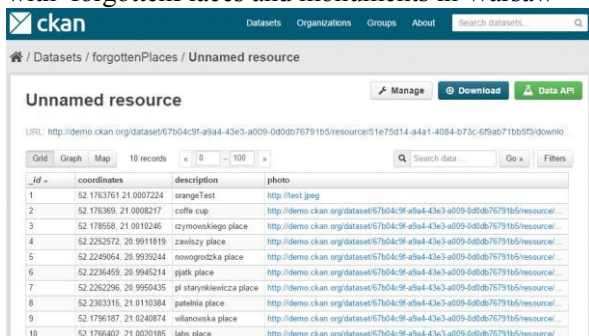

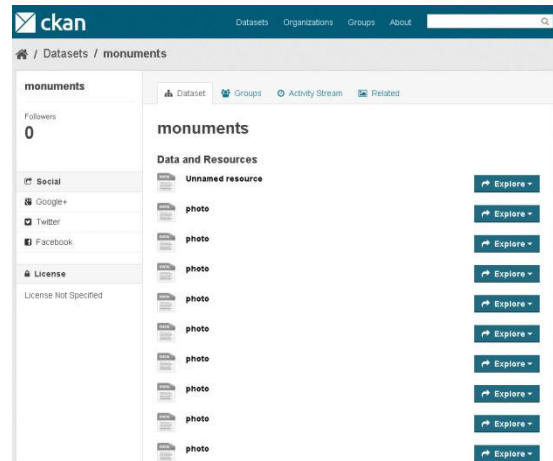Fig. 9. Data set with forgottenPlaces in Warsaw

Fig. 10. Example dataset with monuments in Warsaw

It needs to be stressed that for application test only small data sets in the amount of several records each were collected.

VIII. SUMMARY AND FUTURE WORK

In the future, authors intend to make some modifications in the Crowd Enabler application
1) Improvement of application installation process - application can be exposed in Google Play application shop and installed after connection and download without "install from unknown source" Android option usage
2) Usage of another geolocation sources instead of GPS (e.g. location based on Telco operator infrastructure, Wifi infrastructure etc.).

The main goal of this project was to build a mobile client for CKAN open data platform.

The universality of developed application based on self-configuration capabilities as well as proposed and tested nonstandard RDF meta data usage needs to be emphasized. It should be mentioned that Crown Enabler application has been created in Open Middleware Model using Open Application Programming Interfaces allowing in ease way to create innovative applications and services such as: [17], [19], [20], [21].

The potential of usage of the presented application is very wide. End application users can store large sets of data, that can be described by location, photos and extended text description.

Prototype of Crowd Enabler application was made as part of the Open Middleware 2.0 Community by Orange Labs program [22]

REFERENCES

[1] Auer, S. R.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. (2007). "DBpedia: A Nucleus for a Web of Open Data". "The Semantic Web". Lecture Notes in Computer Science 4825. p. 722.

[2] Portal http://en.wikipedia.org/wiki/Open_data [27.04.2015]

[3] Portal http://opendatacommons.org/licenses/ [27.04.2015]

[4] Portal http://yanosik.pl/ [29.04.2015]

[5] Portal https://warszawa19115.pl/ [29.04.2015]

[6] Open Data Kit (ODK) portal https://opendatakit.org/about/deployments/ [23.10.2014]

[7] CKAN Portal http://ckan.org/ [24.10.2014]

[8] https://github.com/davidebukali/Ukansearch

[9] https://github.com/47deg/labs-opendata-adopta-playa-android

[10] The Open Knowledge Foundation Portal https://okfn.org/ [24.10.2014]

[11] Portal http://datacatalogs.org/catalog?q=platform [13.06.2014]

[12] Portal https://www.data.gov/ [23.10.2014]

[13] Portal data.gov.uk [23.10.2014]

[14] Portal http://publicdata.eu [23.10.2014]

[15] Klyne, G., Carroll, J. J.: Resource description framework (RDF): concepts and abstract syntax. Technical report W3C, 2 (2004)

[16] http://www.w3.org/DesignIssues/LinkedData.html

[17] Sören Auer Volha Bryl Sebastian Tramp (Eds.) Linked Open Data – Creating Knowledge Out of Interlinked Data Results of the LOD2 Project Lecture Notes in Computer Science Springer 2014

[18] Trusiewicz, P.; Witan, M.; Kuzia, M., "Mobile Payment System - Telco 2.0 application dedicated for payments," Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on , vol., no., pp.859,864, 8-11 Sept. 2013

[19] Wawrzyniak, P.; Korbel, P.; Borowska-Terka, A., "Student information delivery platform using telecommunications open middleware APIs," Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on , vol., no., pp.871,874, 8-11 Sept. 2013

[20] Korbel, P.; Skulimowski, P.; Wasilewski, P.; Wawrzyniak, P., "Mobile applications aiding the visually impaired in travelling with public transport," Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on , vol., no., pp. 825,828, 8-11 Sept. 2013

[21] P. Wawrzyniak, Ł. Wronkowski, D. Kuniszewski, A. Cackowski, P. Czapliński i K. Szymański "Send It Safe – A Novel Application for Secure Key Exchange Using Telecommunications Open Middleware APIs," w Frontiers in Network Applications, Network Systems and Web Services (SoFAST-WS'14) w Federated Conference on Computer Science and Information Systems FedCSIS 2014, Warszawa 2014

[22] Open Middleware 2.0 Community portal – http://www.openmiddleware.pl [20.05.2013]