

CLODA: A Crowdsourced Linked Open Data Architecture

Georgios Larkou, Julia Metochi, Georgios Chatzimilioudis and Demetrios Zeinalipour-Yazti

Department of Computer Science

University of Cyprus

P.O. Box 20537, 1678 Nicosia, Cyprus

{glarkou, jmetoc01, gchatzim, dzeina}@cs.ucy.ac.cy

Abstract—In this paper we present our Crowdsourced Linked Open Data Architecture (CLODA), a first attempt to combine crowdsourcing, localization and location-based services to generate, collect, validate and relate real-world, geo-spatial and multi-dimensional information using smartphones and other mobile devices. CLODA focuses on the construction of URI addressable, interlinked and semi-structured data following the Linked-Open Data (LOD) paradigm. The validity of the constructed data is then contributed by a participating crowd. We present our prototype implementation on top of Google Maps and a blend of in-house technologies, particularly our indoor positioning framework, coined Airplace, our trajectory similarity framework, coined SmartTrace, our neighborhood detection framework, coined Proximity and our smartphone testing platform coined SmartLab.

I. INTRODUCTION

Linked data describes a method of publishing structured data so that it can be interlinked and become more useful [1], [23]. Although the Web traditionally has been interlinked, the absence of semantics in published data made it difficult to derive meaningful conclusions from this data. As a result, many efforts nowadays, including CKAN, DBpedia, GeoNames, etc., create semantically-rich data sources from the Web. Such a process is severely hampered by the lack of tools for generating vast amounts of metadata for providing valid semantics at a reasonable scale (e.g., every bar-coded item on this planet).

On the other hand, Crowdsourcing emerges as one of the most suitable means for the seamless realization and adoption of the LOD paradigm at a larger scale, providing validation and completeness as a primitive operation. Crowdsourcing refers to a new distributed social collaborative problem-solving model in which a crowd of undefined size is engaged to solve a complex problem through an open call [3]. Smartphones and mobile devices can unfold the full potential of crowdsourcing and significantly accelerate its role in fulfilling the LOD vision. Mobile devices, nowadays, feature unprecedented multi-sensing capabilities (e.g., geo-location, light, movement, audio and visual sensors, among others), which provide efficient means for opportunistic data collection for big-data LOD scenarios. Even more importantly, they are always connected and are used daily for all kinds of tasks.

Another important aspect related to the fulfillment of the

LOD vision with smartphones, is where the data collection is taken place. According to recent statistics people spend 80-90% of their time in indoors environments, including shopping malls, libraries, airports or university campuses, while 70% of cellular calls and 80% of data connections originate from indoors [4]. This has triggered an increasing interest in indoor Location-based Services (LBS) and location-aware applications, e.g., in-building guidance and navigation, inventory, elderly support for Ambient and Assisted Living (AAL), etc. To enable such applications and facilitate their wider acceptance alternative solutions are required for the provision of accurate and reliable location estimates. Satellite-based positioning, e.g., GPS, is unavailable or significantly degraded inside buildings due to the blockage or attenuation of signal strength.

In this paper we present an architecture that exploits the power of the crowd to generate, collect and validate real-world, multi-dimensional data, using smartphones and mobile devices in indoor and outdoor spaces. Our architecture enables the creation of new open datasets and support data owners in publicizing their data. Such datasets published under the Linked Open Data principles are expected to enable new applications and services (e.g., indoor inventory and data management for shops, warehouses, public buildings, malls, airports, indoor games etc.)

Our motivation was the fact that even if there are many attempts for developing inventory management applications (e.g., Inventory Tracker, Barcode & Inventory Pro and Inventory Android applications), all of them concentrate on the individual user and not the benefit of the crowd. As a result, all data are privately stored and there is no way of sharing them. Our goal is to provide an architecture and a prototype application for validated linked open data acquisition through crowdsourcing.

CLODA consists of several in-house modules we have developed over the years. These modules provide: i) Indoor geo-location [6]; ii) Proximity interactions [7]; iii) The ability to answer social queries based on movement [8]; and iv) Remote smartphone application testing [5].

For the remainder of this section we will focus on the description of two envisioned usage scenarios of our architecture.

Scenario 1: Professors, researchers, staff, students at the University often need to borrow books for their research and teaching activities. Usually, the book needed already exists on campus and is owned by the university, by a colleague or by a laboratory. It would be difficult and inefficient to create and maintain a campus-wide online book inventory system. With the proposed CLODA the owner of a book or equipment would just need to scan the barcode of the product. The item would automatically be geo-tagged, metadata would be fetched from the Semantic Web (e.g., Google API) and the information would be linked and published according to the LOD paradigm. When another student or staff requires access to the item, the position of the item is augmented on a mapping framework (e.g., Google Maps) and the user can easily navigate to the item and ask permission to borrow it. (e.g., see Figure 2).

Scenario 2: An emergency incident takes place in a public place (e.g., airport, theater or a football stadium). A patient with cardiovascular problems needs immediate attention. Even if doctors are among the crowd they are unable to help the patient as they lack the essential equipment. With the proposed architecture and the derived prototype at hand, a user is able to locate the nearest automated external defibrillator (AED) and a first aid kit. The emergency equipment has been previously geo-tagged by a governmental agent and its existence was confirmed by a number of visitors.

The main contributions of the work presented in this paper:

- We present the architecture behind CLODA, a framework providing crowdsourced construction of linked open data.
- We discuss how a variety of innovative technologies can be integrated in a unified application providing the ability to geo-tag items using conventional smartphones.

The CLODA prototype mobile application will pursue the following challenging scientific and technological objectives: i) Develop a framework that combines different data streams (online, sensor readings, user input) to create structured information that follows the LOD principles, ii) Develop an easy-to-use application that will help users register indoor views and items, augmented with indoor location and environment information in addition to the online metadata available for bar-coded items.

The rest of the paper is organized as follows: Section II looks at the related work, Section III introduces our CLOD architecture thoroughly covering each of its sub-modules. Section IV presents our prototype application developed for Android while Section V looks into future issues and concludes the paper.

II. BACKGROUND

A. Indoor Localization Technologies

People spend a significant amount of their daily life indoors¹, thus they interact with items located indoors. Research

on indoor localization has made significant progress in terms of accuracy and required infrastructure during the last decade, but a dominant applicable solution has not emerged yet. In our opinion, the problem is the lack of places and scenarios where indoor localization might be useful, i.e., is limited to airports, malls, etc. In our architecture and its applicable scenarios, indoor localization emerged as the ideal means for generating geo-tagged crowdsourced LOD even in spaces where indoor localization did not seem to be interesting.

The localization literature is very broad and diverse as it exploits several technologies. Assisted GPS (A-GPS) is obviously ubiquitously available but has an expensive energy tag and is also negatively affected from the environment (e.g., cloudy days, forests, downtown areas, etc.) Besides GPS, the localization community [11] has proposed numerous proprietary solutions including: *Infrared, Bluetooth, visual or acoustic analysis, RFID, Inertial Measurement Units, Ultra-Wide-Band, Wireless Sensor Networks, Wireless LANs, etc.*; including their combinations into hybrid systems [12]. Most of these technologies can deliver a high level of positioning accuracy, however they usually require the deployment and calibration of expensive equipment, such as custom transmitters and antennas, which are dedicated to positioning.

B. Crowdsourcing Technologies

Crowdsourcing refers to a new distributed social collaborative problem-solving model in which a crowd is engaged to solve a complex problem. Crowdsourcing can be discriminated into two categories: i) web-based crowdsourcing and ii) mobile crowdsourcing [3]. In our architecture, we utilize the latter as it provides the advantage of location-based crowdsourcing thus unleashing the full potential of location-aware crowdsourcing.

Crowdsourcing applications on smartphones can be classified into extensions of web-based applications or as new applications. The former class expands to users that do not have access to a conventional workstation and adds the dimension of real-time location-based information to the service. Instances of such applications are Gigwalk², Jana³ and the work of Ledlie et al. [13]. The latter class includes applications for crowdsourced traffic monitoring (e.g., Waze⁴) and road traffic delay estimation (*VTrack* [16]); constructing fine-grained noise maps by letting users upload data captured by their smartphone microphone (*Ear-Phone* [14], *NoiseTube* [15]); identifying holes in streets by allowing users to share vibration and location data captured by their smartphone (*PotHole* [18]); location-based games with a purpose to collect geospatial data (*CityExplorer* [17]); leveraging mobile phones for collaborative traffic signal schedule advisory (*SignalGuru* [19]); offer new employment opportunities to low-income workers in developing countries [20] [21] [22]; and real-time fine-grained indoor localization services exploiting the Radio Signal Strength (RSS) of WiFi access points (*Airplace* [6]) .

²Gigwalk Inc., May 2012, <http://www.gigwalk.com/>

³Jana, May 2012, <http://www.jana.com/>

⁴Waze Ltd., April 2012, <http://www.waze.com/>

¹Feb. 3, 2013: Stage Analytics, <http://www.strategyanalytics.com>

C. Linked-Open Data

Linked Open Data (LOD) is the method of publishing structured data (e.g., RDF, XML) that can be interlinked among different websites enabling data integration and querying [23]. CKAN, a registry of open data and content packages provided by the Open Knowledge Foundation contains most of the publicly available open datasets provided by a variety of authors. For instance Freebase, provided by Google, creates an entity graph of people, places and things while YAGO is a knowledge base derived from Wikipedia and GeoNames among others.

In addition to the general Linked-Open Databases, a large number of other more content specific LOD exist. Instances of such LODs include linked databases related to music (e.g., BBC Music), others related to the research community and publications (e.g., DBLP), others related to open source projects (e.g., SECOLD), open libraries (e.g., the Open Library), live linked open sensor databases [24], crowdsourcing parking availability information [25] and crowdsourcing tasks [26] within linked data management.

D. Big-Data and data mining

Big data refers to data sets whose size and structure strains the ability of commonly used relational DBMSs to capture, manage, and process the data within a tolerable elapsed time. Big data sizes commonly range from a few dozen terabytes to many petabytes in a single database and their underlying data model might be anything from structured (relational or tabular) to semi-structured (XML or JSON) or even unstructured (Web text and log files). Big data architectures are highly parallel and distributed in order to cope with the inherent I/O and CPU limitations. Such systems typically perform on mid-scale private clouds, offering higher privacy, to large-scale public clouds, both exposing operational and analytic functionality stand-alone or as-a-Service.

The evolution of Big Data and the ideal marriage of Big Data and the Semantic Web is able to provide the metadata needed for the widely adoption of the latter. Mega-modelling [27] was proposed as a new model management system for the acquisition, composition, integration, management, querying and mining of data. GLADE [28], a scalable distributed system for large scale data analytics, takes analytical functions expressed through the User-Defined Aggregate (UDA) interface and executes them efficiently on the input data. In addition, Big-Data goals, as presented in [29], highlights the significance of developing new analytic tools able to extract metadata (required for LODs) in a fast and cost-efficient manner from enormous datasets.

III. THE CLOD ARCHITECTURE

CLODA is the first attempt to combine crowdsourcing, localization and location-based services to generate, collect, validate and relate real-world, geo-spatial and multi-dimensional information. CLODA has been inspired by both DBpedia [9], a crowd-sourced community effort to extract structured

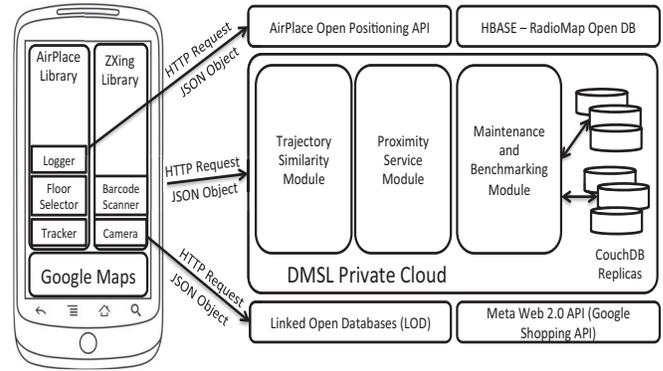


Fig. 1. The client-side and server-side building blocks of our presented CLOD Architecture.

information from Wikipedia and the GeoNames⁵ geographical database, another crowd-sourced community effort that integrates geographical data such as names of places in various languages, elevation, population and others.

In this section, we will present CLODA along with its underlying principles and technologies. Our motivation was the fact that even if there are a lot attempts for developing inventory management applications (e.g., Inventory Tracker, Barcode & Inventory Pro and Inventory Android applications), all of them concentrate on the individual user and not the benefit of the crowd. As a result, all data are privately stored and there is no way of sharing them. Our goal is to provide an architecture and a prototype application for validated linked open data acquisition using crowdsourcing.

A. Overview

The users of CLODA are able to contribute geo-tagged, validated metadata information about items using their mobile phones. The information is stored in a distributed farm of databases and is publicly accessible. The owner of the items is the crowd, willing to contribute to a publicly available inventory for an ethical benefit. For instance, in our first scenario, the owner of the items is the crowd (e.g., lectures, university staff, the students) on the other hand in our second scenario the owner might be a governmental authority or a public place (e.g., airport, museum, stadium) owner/coordinator willing to contribute to the public repository. CLODA is organized in a three-tier architecture that consists of the Android Application prototype, some highly-available Web-servers and the combination of the supporting client-side module and the respective server-side module, as illustrated in Figure 1.

In the following subsections, we will present more thoroughly the array of in-house modules used to provide: i) Crowdsourced indoor positioning independent of specialized infrastructure (i.e., Airplace [6]); ii) The ability to answer social queries based on movement (i.e., SmartTrace [8]); iii) Geo-spatial neighbor relationships (i.e., Proximity [7]); and iv) Remote smartphone applications testing (i.e., SmartLab [5]).

⁵Feb. 3, 2013; GeoNames, <http://www.geonames.org>

B. Indoor Localization Module

We utilize the AirPlace [6] library, which exploits *Received Signal Strength (RSS)* values extracted through passive scanning of the beacon packets transmitted by neighboring APs as part of the standard network functionality. To address the challenging signal propagation conditions indoors, due to multi-path, reflections and diffractions, RSS fingerprints (i.e., vectors of RSS measurements recorded from APs in the vicinity of the user) are collected a priori at predefined reference locations. Each RSS fingerprint is associated with the respective reference location and is stored in the so called *Radiomap* that covers the whole area of interest. Essentially, the Radiomap is a mapping from the multi-dimensional RSS space to the physical coordinates. Airplace overcomes the drawbacks of other indoor positioning systems as explained in paragraph II-A. A significance advantage of CLODA prototype in comparison to any existing work is the ability to augment those items over a mapping framework (e.g., Google Maps).

C. Trajectory Similarity Module

The trajectory similarity modules provides users moving similarly to other CLODA users. The CLODA prototype is responsible to gather those users and request from them to validate the existence of any geo-tagged items on their route. The given module is implemented on top of SmartTrace [8], which enables trace similarity search among smartphone users and optimizes queries with respect to response time and energy consumption. More importantly, *SmartTrace* is privacy-aware since it does not share the user trajectories to the authority, rather it only returns matching scores.

SmartTrace, relies on an in-situ data storage model, where geo-location data is recorded locally on smartphones for both performance and privacy reasons. To formalize our description, let $\{A_1, A_2, \dots, A_m\}$ denote a collection of spatiotemporal trajectories. A spatio-temporal trajectory A_i ($i \leq m$) is defined as a sequence of l multidimensional tuples $\{a_1, \dots, a_l\}$ where each tuple is characterized by two spatial dimensions and one temporal dimension (i.e., $a_j(x_j, y_j, t_j), \forall j \in 1, \dots, l$). Each trajectory A_i resides in its entirety in-situ, which is cheaper and more efficient for smartphone environments. Given a query Q , itself expressed as a spatio-temporal trajectory, we compare each A_i to the points of Q within some temporal and spatial window. SmartTrace, circumvents expensive and massive similarity executions by running an inexpensive linear-time computation on the smartphones in a pre-processing step. It then uses an iterative top-K processing algorithm in order to iteratively identify the K most similar trajectories to Q , without ever pulling the target trajectories together.

D. Proximity Service Module

CLODA utilizes the proximity service module in order to provide enhanced social collaboration between neighboring users. Additionally, the proximity service module is also used in order to support verification and validation of the crowdsourced linked data provided by the users. As previously described while users are geo-tagging items in an indoor

environment the proximity service module can identify their k nearest neighbors and provided them to CLODA. Similarly to the previous module, CLODA prototype application will then asked them to validate the existence of the items previously geo-tagged by another user.

The given CLODA module is founded on the Proximity Framework [7], which was proposed to answer such *Continuous All k Nearest Neighbor (CAkNN)* queries efficiently based on the crowdsourcing of user locations. The proximity framework is: i) *Stateless* to cope with transient user populations and high mobility patterns; ii) *Parameter-free* to be invariant to parameters that are network-specific (such as cell size, capacity, etc.) and user distribution specific; iii) *Memory-resident*, since the dynamic nature of mobile user makes disk resident processing prohibitive; iv) *Specially designed* for highly mobile and skewed distribution environments performing equally well in downtown, suburban, or rural areas; v) *Fast and scalable*, in order to allow massive deployment; and vi) *Infrastructure-ready* since it does not require any additional infrastructure or specialized hardware.

A suite of proximity-based social interaction applications, called Crowdcast, has already been build on top of our powerful Proximity framework. Proximity efficiently connects you to your closest neighbors at all times, regardless of where you are and how far they are. Those neighbors can be shown in a list or on a map. On top of functionality a whole suit of applications have been developed: (i) Helpcast, to send out SOS beacons or disseminate natural disaster warnings; (ii) Msgcast, to post local microblogging messages; (iii) Eyecast, to extend the view on the urban environment using the cameras of ones neighbors; (iv) Miccast, to post local vocal messages and warnings; (v) Taskcast, to post local tasks in your neighborhood as part of local crowdsourcing or organizing a charity event, etc. The server has the overall picture of the users whereabouts and can compute the k nearest neighbors for each user.

E. Maintenance and Benchmarking Module

Re-programming smartphones and instrumenting them for application testing and data gathering is currently a tedious, time-consuming process that poses significant logistical challenges. To this end, we utilized SmartLab⁶, a first-of-a-kind open Infrastructure-as-a-Service (IaaS) cloud that enables fine-grained control over both real and virtual smartphones via an intuitive web-based interface. SmartLab's current infrastructure is ideal for scenarios that require fine-grained and low-level control over real smartphones, e.g., OS, Networking, DB & storage, security, peer-to-peer protocols, but also for scenarios that require the engagement of physical sensors and geo-location scenarios. SmartLab has been utilized during the development process of the Android prototype application.

SmartLab has facilitated the development, testing and demonstration of CLODA considerably as explained next. Firstly, it allowed us to move around in a building localizing ourselves while exposing the smartphone screen on a

⁶System and Video available: <http://smartlab.cs.ucy.ac.cy/>.

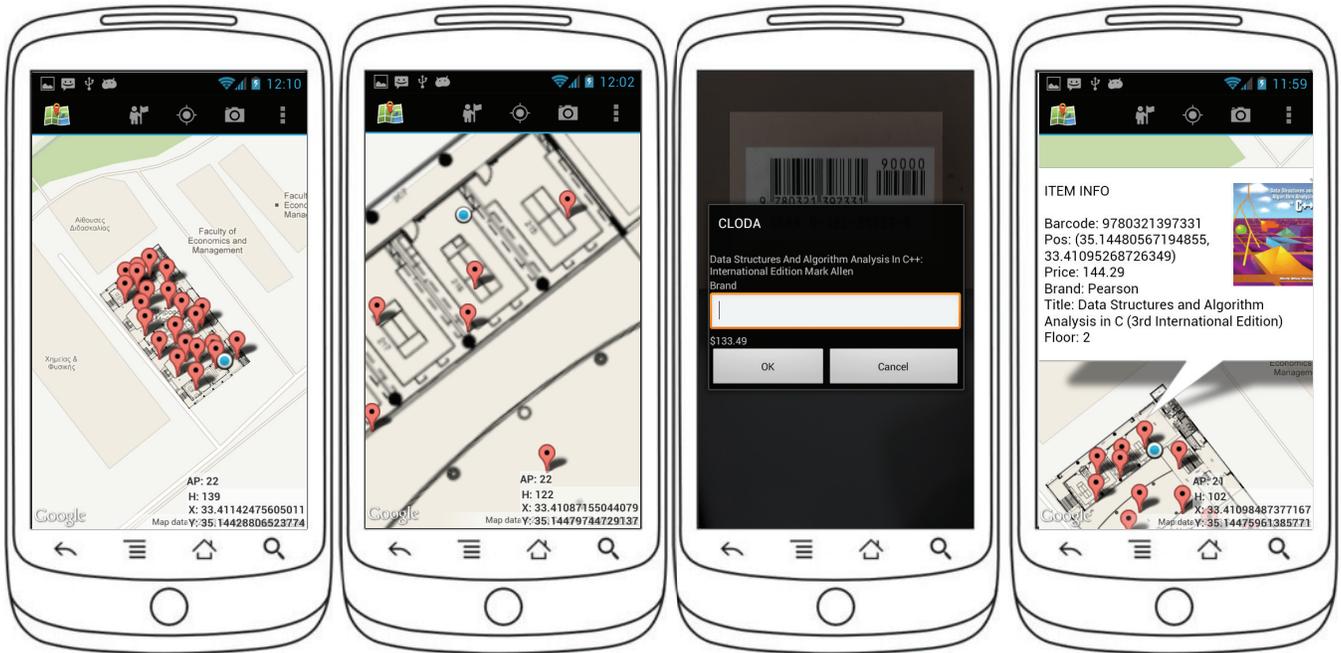


Fig. 2. CLODA Android Application Prototype Screenshots.

remote web browser through SmartLab. The particular setting has proved considerably useful as the bulk of existing AndroidScreenCapture software are both USB-based (that hinders mobility) but also inefficient (i.e., no compression or other optimizations). Secondly, SmartLab allowed us to collect and compare *Received Signal Strength (RSS)* indicators from different WiFi chip-sets, which is important for RSS measurements and would not be possible with emulators. Finally, SmartLab allowed us to test the generated APK on a variety of devices.

IV. OUR PROTOTYPE

The main goals of our prototype Android Application are to develop: i) a framework that combines different data streams (online, sensor readings, user input) to create structured information that follows the Linked Data principles and ii) an easy-to-use application that will help users register indoor views and items augmented with indoor location and environment information and online metadata available for bar-coded items.

A. Overview

The prototype provides the ability for the user to seek and “geo-tag” items in indoor and outdoor environments over a traditional geographic mapping systems (e.g., Google Maps). CLODA prototype allows the users to upload and adjust floor maps for buildings as layers over Google Maps while utilizing the AirPlace positioning library to navigate through the corridors and contribute items/inventories to the public LOD. Additionally it provides the ability to automatically retrieve basic information about the item by extracting the barcode of the item using a third party library, ZXing Library, and

consuming the meta-data of the items using Google Shopping API.

B. Implementation

In this section we overview the three-tier prototype of CLODA, i.e., the Android Application prototype, the highly-available Web-servers and the distributed databases.

The prototype mobile application was developed on top of the Android mobile OS and allows the user to geo-tag new items, edit, search or delete existing inventories while it provides the ability to the user to localize himself indoor. The interface of the prototype Android application is presented in Figure 2. The leftmost image presents an overview of the items located at the Computer Science building, University of Cyprus and the user’s location is represented by a blue dot. All information is presented as additional layers on top of the actual Google Maps. The second and the third images present the process of scanning an item and inserting this item to the LOD using the information provided from the Google Shopping API⁷ with the help of the ZXing open-source, multi-format 1D/2D barcode image processing library⁸. Additionally, the third image presents how the user is able to enhance the information provided by Google Shopping API before inserting it to the LOD.

As mentioned in the previous sections, the location of the user is provided by the AirPlace indoor positioning library. Before navigation in an indoor environment the prototype application is downloading the floormap, which is placed as a

⁷Google Shopping API - <https://developers.google.com/shopping-search/>

⁸ZXing - <http://code.google.com/p/zxing/>

layer on top of the maps using the Google Maps API and a radiomap corresponding to the nearest to the user buildings. If the floormap and the radiomap are not available the application request from the user to provide them as both are required for positioning.

The latter two tiers of our architecture constitute the back-end of CLODA prototype architecture. A variable number of CentOS powered virtual machines is created in order to provide a link between the prototype Android application and the CouchDB database. The reason behind the introduction of a middle-ware layer between the Android application and the databases was the security requirement introduced during the design phase that required all users to authenticate their request to the database. In order to not reveal any sensitive information in raw format (e.g., database credentials) through the wireless network we decided to introduce a middleware API responsible for proxying authenticated users' requests to the database. The communication between the middleware layer and the prototype Android application is achieved in a Restful manner (i.e., HTTP requests are issued by the application and JSON objects are returned by the middleware).

Finally, the third-tier of our architecture consists of multiple CouchDB replicas. The database replicas are responsible to store the information regarding each item geo-tagged by the user and serve any incoming requests for previously geo-tagged items.

V. CONCLUSIONS

This paper presents an innovative architecture for crowd-sourced linked open data, coined CLODA. The proposed architecture utilizes the advantages of each sub-module, e.g., Indoor Localization, Trajectory Similarity, Crowdsourcing with mobile devices and Proximity Services, in order to provide a sustainable framework for crowdsourced linked open data collection. The major advantages of CLODA are: i) The ability to operate indoors, on top of existing geographic mapping technologies (e.g., Google Maps), ii) Consuming publicly available APIs (e.g., Google Shopping API) and iii) Utilizing the power of the crowd in order to construct verified geo-location aware LOD.

As a future work, we will investigate the advantages of the introduction of incentives for the users in order to enhance the collaboration between them and encourage individuals contribution to the publicly available LOD while providing guarantees for the validity of the data gathered.

ACKNOWLEDGMENT

This work was supported by the last author's startup grant, funded by the University of Cyprus, EU's FP7 Planetdata NoE (2007-2013), the EU's FP7 "Mobility, Data Mining, and Privacy" project and the COST Action IC903 (MOVE) "Knowledge Discovery for Moving Objects".

REFERENCES

[1] T. Berners-Lee, J. Hendler and O. Lassila, "The Semantic Web", In Scientific American, p. 29-37, May 2001.

[2] N. Shadbolt, W. Hall, T. Berners-Lee, "The Semantic Web Revisited", In IEEE Intelligent Systems, 2007.

[3] G. Chatzimiloudis, A. Konstantinides, C. Laoudias, D. Zeinalipour-Yazti, "Crowdsourcing with Smartphones", In IEEE Internet Computing, Sept/Oct, 2012.

[4] Source: Strategy Analytics <http://www.strategyanalytics.com>.

[5] A. Konstantinides, C. Costa, G. Larkou and D. Zeinalipour-Yazti, "Demo: A Programming Cloud of Smartphones", In MobiSys'12.

[6] C. Laoudias, G. Constantinou, M. Constantinides, S. Nicolaou, D. Zeinalipour-Yazti, and C. G. Panayiotou, "The Airplace indoor positioning platform for android smartphones", In IEEE MDM'12 (Best Demo Award).

[7] G. Chatzimiloudis, D. and Zeinalipour-Yazti, W.-C. Lee, M. D. Dika-iakos, "Continuous All k-Nearest Neighbor Querying in Smartphone Networks", In IEEE MDM'12.

[8] D. Zeinalipour-Yazti, C. Laoudias, C. Costa, M. Vlachos, M. I. Andreou, D. Gunopulos, "Crowdsourced Trajectory Similarity with Smartphones", In IEEE TKDE'12.

[9] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, S. Hellmann, "A Crystallization Point for the Web of Data", In WWW'09.

[10] Source: The GeoNames geographical database - <http://www.thegeonames.org>.

[11] Y. Gu, A. Lo, I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks", In IEEE Communications Surveys & Tutorials 09.

[12] C.-L. Li, C. Laoudias, G. Larkou, G. Chatzimiloudis, G. D. Zeinalipour-Yazti, C.G. Panayiotou, "Hybrid Indoor Positioning on Multi-Sensor Android Smartphones", In IPIN'12.

[13] J. Ledlie, B. Otero, E. Minkov, I. Kiss, J. Polifroni, "Crowd translator: on building localized speech recognizers through micropayments", In ACM SIGOPS Operating Systems Review, 2010.

[14] R. K. Rana, C.-T. Chou, S. Kanhere, N. Bulusu, W. Hu, "Ear-phone: an end-to-end participatory urban noise mapping system", In IPSN'10.

[15] M. Stevens, E. D. Hondt, "Crowdsourcing of Pollution Data using Smartphones", In UbiComp'10.

[16] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, J. Eriksson, "VTrack: accurate, energy-aware road traffic delay estimation using mobile phones", In ACM SenSys'09.

[17] S. Matyas, C. Matyas, C. Schlieder, P. Kiefer, H. Mitarai, M. Kamata, "Designing location-based mobile games with a purpose: collecting geospatial data with CityExplorer", In International Conference on Advances in Computer Entertainment Technology 2008

[18] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, H. Balakrishnan, "The pothole patrol: using a mobile sensor network for road surface monitoring", In MobiSys'08.

[19] E. Koukoumidis, L.S. Peh, M.R. Martonosi, "SignalGuru: leveraging mobile phones for collaborative traffic signal schedule advisory", In MobiSys'11.

[20] A. Gupta, W. Thies, E. Cutrell, R. Balakrishnan, "mClerk: enabling mobile crowdsourcing in developing regions", In CHI'12.

[21] T. Yan, M. Marzilli, R. Holmes, D. Ganesan, M. Corner, "mCrowd: a platform for mobile crowdsourcing", In SenSys'09.

[22] N. Eagle, "txteagle: Mobile Crowdsourcing", In IDGD'09.

[23] C. Bizer, T. Heath, T. Berners-Lee, "Linked Data - The Story So Far", Int. J. Semantic Web Inf. Syst. 5(3): 1-22, 2009.

[24] D. Le-Phuoc, J. X. Parreira, M. Hausenblas, Y. Han, and M. Hauswirth, "Live linked open sensor database", In I-SEMANTICS'10.

[25] J. Kopecky, J. Domingue, "ParkJam: Crowdsourcing Parking Availability Information with Linked Data", In ESWC'12.

[26] E. Simperl, B. Norton, D. Vrandečić, "Crowdsourcing Tasks within Linked Data Management", In ESWC'12.

[27] S. Ceri, E. Valle, D. Pedreschi, R. Trasarti, "Mega-modeling for Big Data Analytics", In Conceptual Modeling, 2012.

[28] Y. Cheng, C. Qin, F. Rusu, "GLADE: big data analytics made easy", In SIGMOD'12.

[29] S. Chaudhuri, "What next?: a half-dozen data management research goals for big data and the cloud", In PODS'12.