

VU Research Portal

Using linked open geo boundaries for adaptive delineation of functional urban areas

Khalili, Ali; van den Besselaar, Peter; de Graaf, Klaas Andries

published in

The Semantic Web: ESWC 2018 Satellite Events: Revised Selected Papers
2018

DOI (link to publisher)

[10.1007/978-3-319-98192-5_51](https://doi.org/10.1007/978-3-319-98192-5_51)

document version

Publisher's PDF, also known as Version of record

document license

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Khalili, A., van den Besselaar, P., & de Graaf, K. A. (2018). Using linked open geo boundaries for adaptive delineation of functional urban areas. In A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan, & M. Alam (Eds.), *The Semantic Web: ESWC 2018 Satellite Events: Revised Selected Papers: ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers* (pp. 327-341). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 11155 LNCS). Springer/Verlag.
https://doi.org/10.1007/978-3-319-98192-5_51

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal@vu.nl



Using Linked Open Geo Boundaries for Adaptive Delineation of Functional Urban Areas

Ali Khalili¹(✉), Peter van den Besselaar², and Klaas Andries de Graaf¹

¹ Department of Computer Science, Vrije Universiteit Amsterdam,
Amsterdam, Netherlands
{a.khalili,ka.de.graaf}@vu.nl

² Department of Organization Sciences, Vrije Universiteit Amsterdam,
Amsterdam, Netherlands
p.a.a.vanden.besselaar@vu.nl

Abstract. The concentration of people, companies, research organizations and other activities in urban areas is a key process in the development of economies and societies. In order to investigate how these urban systems function, the OECD (Organization for Economic Co-operation and Development) in collaboration with EC (European Commission) and Eurostat have introduced the concept of Functional Urban Areas (FUAs). FUAs consider a preliminary set of socio-economic and environmental factors and provide a basis for an agreed definition for measuring development of metropolitan areas. However, because FUAs are predefined they do not meet the need for designing policies and research questions involving different types of urban areas that are defined by weighting some factors more than others or by using additional factors. Therefore, providing an adaptive approach for dynamic and multi-faceted delineation of FUAs, rather than merely relying on a rigid schema with a fixed list of FUAs per country, allows to more flexibly reflect the socio-economic geography of where people live and work. This adaptive definition of FUAs demands integration of data from multiple up-to-date linked data sources. In this paper, we describe an approach and implementation for a Linked Open Geo-Data space, which combines openly available spatial and non-spatial resources on the Web to classify urban areas with the aim to more flexibly monitor and research urban development.

1 Introduction

The concentration of people, companies, research organizations and other activities in urban areas is a key process in the development of economies and societies. How urban systems function is crucial to future economic prosperity and better quality of life for more than three billion people [1]. In order to investigate how these urban systems function, the OECD (Organization for Economic Co-operation and Development) in collaboration with EC (European Commission) and Eurostat have developed a new approach to classifying urban areas with the aim to better monitor urban development within and across countries.

The new notion of urban areas called *Functional Urban Areas* (FUAs) considers several factors beyond the formal city boundaries such as population, area, GDP, environment (CO₂ emissions and air pollution), labour market (employment and unemployment growth), innovation (patent intensity), urban form and territorial organization to develop a harmonized definition of urban areas in 28 OECD countries. Even though this new way to measure metropolitan areas provides a basis for an agreed definition of functional urban areas, to support the design of better policies for different types of urban areas, one needs to weight some factors more than others or use additional factors, which are not predefined in the current OECD methodology. Governments and policy makers need a way to dynamically redefine different types of urban areas [2]. Therefore, providing an adaptive approach for dynamic and multi-faceted delineation of FUAs, rather than merely relying on a rigid schema with a fixed list of FUAs per country, allows to more flexibly reflect the socio-economic geography of where people live and work. This adaptive definition of FUAs demands integration of data from multiple up-to-date linked data sources. Spatial data often has a temporal component; things move, and boundaries change over time.

Tackling the challenges of data integration on a dynamic environment such as WWW has been the mission of Linked (Open) Data technologies since the introduction of Semantic Web in 2001. Providing a *Linked Open Data* space, which brings together structured and interlinked geospatial data on the Web, facilitates delineating of FUAs. Currently, access to the micro data used by OECD for calculating different indicators of FUAs is limited, and negotiation with OECD is required to retrieve detailed data for regeneration of OECD FUAs.¹ On the other hand, openly available geospatial datasets on the Web such as *OpenStreetMap*² are already interlinked with existing structured data published on the Linked Open Data (LOD) cloud and provide the opportunity to reproduce a more flexible and dynamic list of FUAs. For example, the public sector in Europe creates lots of statistical data on different levels of administrative boundaries such as *NUTS* (Nomenclature of Units for Territorial Statistics), *LAU* (Local Administrative Unit), *HASC* (Hierarchical Administrative Subdivision Codes) and *ISO 3166* country codes which could be utilized to dynamically identify FUAs adapted to the context of the corresponding policy studies.

In this paper we propose a Linked Data approach and implementation which combines openly available spatial and non-spatial resources on the Web to more flexibly classify urban areas. To achieve this goal, we followed methodology described in the LOD Lifecycle [4], consisting of several steps (cf. Fig. 1) for geospatial data collection, extraction, storage, linkage and exploitation which are discussed in the following sections of this paper. This paper makes the following contributions:

- Report an approach and implementation for dynamically defining FUAs based on linked open data.

¹ <http://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm> provides the shapefiles for FUAs together with some descriptions in PDF format.

² <http://www.openstreetmap.org>.

- Use linked open data to reconstruct the closed OECD FUA dataset.
- Report a use case on the implemented approach.

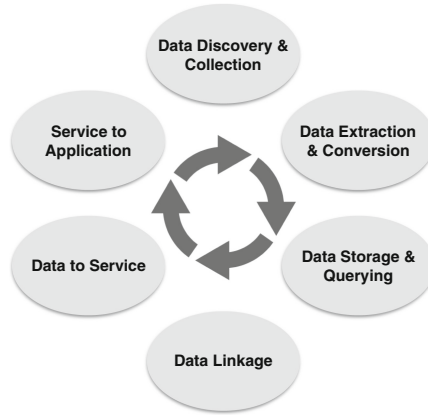


Fig. 1. An overview of the steps for adaptive delineation of FUAs.

2 Step 1 - Data Discovery and Collection

As first step (step ‘*search/browse/exploration*’ in LOD Lifecycle [4]), we performed an extensive offline/online search to find existing relevant geospatial datasets, which provide data for world-wide administrative boundaries. We were particularly looking for datasets that contain shapefiles for those boundaries. For the offline search, we used our network to find research groups working on spatial data infrastructures and through them either get access to available geospatial datasets or find other related institutes that publish geospatial data related to urban areas. For the online search, we used both general-purpose search engines (e.g. Google) as well as search engines indexing only structured content (DataHub³, LOTUS⁴ and EU Open Data Portal⁵). Progressive moves toward Open Data are creating frameworks through which geographic data assets that have often previously not been in the public domain, or been in the public domain under more restrictive licenses, can be released for free to re-use in either commercial or non-commercial applications [11]. As result of our search for open geo data, we discovered the following resources on the Web providing geospatial data for administrative boundaries:

OpenStreetMap (OSM) Data. OSM is a collaborative project to build a free editable map of the world. OSM offers up to 10 administrative boundary levels

³ <http://datahub.io>.

⁴ <http://lotus.lodlaundromat.org>.

⁵ <https://open-data.europa.eu/en/data/>.

as subdivisions of areas/territories/jurisdictions recognized by governments or other organizations for administrative purposes.⁶ And for these administrative units, all kind of socio-economic, demographic, and other data are available. These administrative boundaries range from large groups of nation states right down to small administrative districts and suburbs. There are different methods to access the properties (including the shape coordinates) of administrative boundaries in OSM. Nominatim Web API⁷ allows querying OSM for a name or address(forward search) or look up data by its geographic coordinate(reverse search). The Overpass API⁸ allows fetching selected parts of the OSM map data by search criteria such as location, type of objects, tag properties, proximity, or combinations of them. In addition to the API access, users can directly download the latest data dump of the OSM through Planet.osm mirrors⁹ in two main available formats namely PBF and compressed OSM XML.

Database of Global Administrative Areas (GADM). GADM¹⁰ provides a curated database of the administrative areas in the world. GADM provides some properties of these administrative areas such as name, variant names and “spatial features” about the location of the areas. Administrative areas in this database include up to 6 levels of details starting from level 0 which refers to countries. Level 1 to 5 cover lower level subdivisions such as provinces, departments, counties, etc. depending on the size and availability of data for the underlying country. The GADM data are publicly available for download by country or the whole world in different formats such as shapefile, ESRI geodatabase, RData, and Google Earth kmz format.

Flickr Shapefiles Dataset. Flickr Shapefiles Public Dataset¹¹ provides data from 190M geo-tagged photos on Flickr. The shapefiles are generated by plotting all the geotagged photos associated with a particular place and by generating a mostly accurate contour of that place. Flickr offers 6 levels of boundaries identified by so called Where On Earth (WOE) IDs. The levels range from country (level 1), region (level 2) county (level 3), locality (level 4) to neighborhood (level 5). The dataset is publicly available for download in GeoJSON format.

Published Shapefiles for Individual Countries. In addition to crowd-sourced and curated datasets on global administrative boundaries, local administrative offices or geo-related research centres in specific countries provide shapefiles and other properties related to the administrative units in that country. For example, *Centraal Bureau voor de Statistiek* (CBS) or *Bundesamt für Kartographie und Geodäsie* (BKG) provide shapefiles of administrative boundaries for the Netherlands and Germany respectively.

⁶ for some specific countries level 11 is also defined. Check http://wiki.openstreetmap.org/wiki/Template:Admin_level.11.

⁷ <http://nominatim.openstreetmap.org/>.

⁸ <http://overpass-api.de/api/>.

⁹ <http://wiki.openstreetmap.org/wiki/Planet.osm>.

¹⁰ <http://www.gadm.org>.

¹¹ <http://www.flickr.com/services/shapefiles/2.0/>.

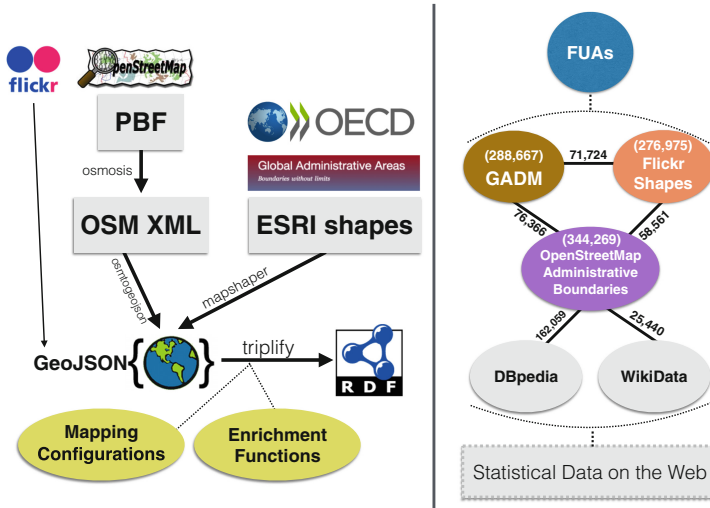


Fig. 2. (left) Steps to convert geospatial data to RDF. (right) General and domain-specific interlinked datasets for delineating FUAs.

Published Geospatial RDF Datasets. There are already several efforts to implement a spatial dimension on the Web of Data (a.k.a. Semantic Web). GeoKnow [3]¹² and LinkedGeoData projects [12]¹³ collect and publish the information extracted from the INSPIRE [8] and OpenStreetMap data sources as an RDF knowledge base interlinked with other knowledge bases in the Link Open Data initiative. To the best of our knowledge, LinkedGeoData dataset¹⁴ does not provide “relations” elements for OSM administrative units which are required to create precise polygon or multi-polygon shapes for them. GeoVocab.org is another related effort which provides an RDF spatial representation of the administrative boundaries represented in the GADM database called *GADM-RDF*¹⁵. We also found several geo datasets on particular countries, for instance Spanish open geo datasets [13] or Ecuadorian geospatial Linked Data [10], etc.

3 Step 2 - Data Extraction and Conversion

The existing diverse landscape of standards for spatial data on the Web makes the task of data extraction and conversion very cumbersome and time-consuming. The left side of Fig. 2 depicts our implemented approach to deal with spatial data extraction, processing and conversion. This corresponds to step ‘*extraction*’ in the LOD Lifecycle [4]. We used *GeoJSON* as our terminal data format for the conversion to RDF. The Flickr Shapefiles Dataset was

¹² <http://geoknow.eu>.

¹³ <http://linkedgeodata.org/>.

¹⁴ <http://downloads.linkedgeodata.org/releases/2015-11-02/>.

¹⁵ <http://gadm.geovocab.org/>.

already available in GeoJSON format. Additional processing and conversion to GeoJSON format was needed for some of the other collected data. We downloaded the OSM dataset in PBF (Protocolbuffer Binary Format) which provides a more compressed format comparing to the XML format. We then used the Osmosis¹⁶ tool to process the data and to only extract the data about administrative boundaries in OSM format. The OSMtoGeoJSON¹⁷ tool was then applied on the extracted subset which resulted in GeoJSON version of data. We used MapShaper¹⁸ to convert the OECD shapefiles and GADM dataset from ESRI format to GeoJSON format.

We utilized a set of *Mapping Configurations* and *Enrichment Functions* to convert spatial data encoded in GeoJSON to RDF format. The Mapping configurations provided a mapping between the given properties of data in original dataset and their best matching RDF properties expressed in the Linked Open Data cloud. Linked Open Vocabularies¹⁹ were used to produce suggestions from existing vocabularies on the Web. In case no existing RDF properties are available, a new proprietary RDF property is created and defined as part of our proposed vocabulary.

Enrichment functions were defined to clean up, standardize, and enrich the property values. For example, we added the ISO 3166 code of the countries by processing the given country names and converted the given Wikipedia URLs to their corresponding DBpedia URIs. We also set the right data types for the converted literal values. The convertor scripts are available as separate repositories on Github²⁰. In addition to the spatial data, we extracted tabular data about the OECD list of municipalities and FUAs²¹, as well as metadata on different OSM levels provided as HTML tables on Wikipedia²², and converted this data to RDF. The final RDF dataset consisted of 344,269 administrative boundaries from OSM, 288,668 from GADM and, 276,975 from Flickr.

4 Step 3 - Data Storage and Querying

We used Openlink Virtuoso triple store for storing the generated RDF data. The main reason for using Virtuoso was its extensive support for geometry data types and spatial indexing²³. At the time of conversion to RDF, we adapted all the GeoJSON shapes coordinates to WKT (Well-Known Text) Polygon and MultiPolygon representations. Virtuoso's stored procedures such as

¹⁶ <http://wiki.openstreetmap.org/wiki/Osmosis>.

¹⁷ <https://github.com/tyrasd/osmtogeojson>.

¹⁸ <https://github.com/mbloch/mapshaper>.

¹⁹ <http://lov.okfn.org/>.

²⁰ <https://github.com/alilk/osm-rdf>
<https://github.com/alilk/gadm-rdf>
<https://github.com/alilk/flickrshapes-rdf>.

²¹ <http://www.oecd.org/gov/regional-policy/List-municipalities.xls>.

²² http://wiki.openstreetmap.org/wiki/Template:Admin_level_10.

²³ <http://docs.openlinksw.com/virtuoso/sqlrefgeospatial.html>.

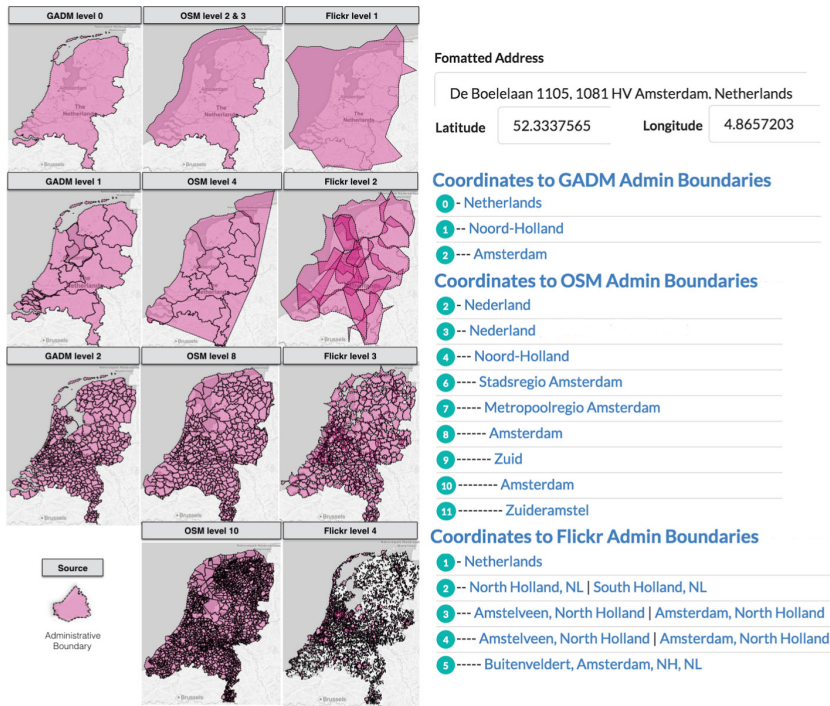


Fig. 3. Example mapping of an address to the extracted administrative boundaries.

`st_intersects`, `st_contains` and `st_within` were then used to test whether two geometries overlap in different ways. For example to find all the administrative boundaries which contain a certain point. This step corresponds to step ‘*Storage/Querying*’ in the LOD Lifecycle [4].

5 Step 4 - Data Linking

In order to exploit the power of Linked Data, we established links between the converted RDF datasets and other open datasets available on Linked Open Data cloud. This corresponds to step ‘*Interlinking/Fusing*’ in the LOD Lifecycle [4]. The right side of Fig. 2 shows the connectivity of the main datasets. The OSM dataset already contains links to general knowledge bases, e.g., DBpedia and WikiData, which serve as hubs to interlink with other open statistical datasets.

In order to compute direct links between OSM, GADM and Flickr, we followed a hybrid approach combining string similarity with the geometric overlapping of administrative boundaries. We first created a mapping between different levels of boundaries provided in OSM, GADM and Flickr by comparing the granularity of divisions in different countries. We took into account the provided OSM metadata per country for each administrative boundary level. Figure 3 shows a sample of extracted administrative boundaries for the Netherlands which

reflects the possible mappings at different levels for a specific address (top-right of Fig. 3). Secondly, we checked the overlaps of areas at the similar level, and for the matching areas we applied string matching to make sure that they refer to the same administrative boundary. Code 1.1 brings an example of CONSTRUCT queries used to create linksets between the OSM and GADM datasets. To showcase the output of our query, for Amsterdam in the Netherlands, the approach will result in the following linked entities: `oecd:NL002`, `gadm:158-9-266`, `hasc:NL-NH-AD`, `osm:relation.47811`, `flickr:727232`, `dbpedia:Amsterdam`, `wikidata:Q9899` and `geonames:2759794`.

Code 1.1. An example of making links between OSM and GADM datasets.

```

1 PREFIX GADMV: <http://geo.risis.eu/vocabulary/gadm/>
2 PREFIX OSMV: <http://geo.risis.eu/vocabulary/osm/>
3
4 CONSTRUCT {?s1 owl:sameAs ?s2 .} WHERE {
5   graph <http://geo.risis.eu/osm> {
6     ?s1 a OSMV:AdministrativeArea ;
7     OSMV:level "2"^^xsd:integer ;
8     dcterms:title ?title1 ;
9     geo:geometry ?polygon1 . }
10  graph <http://geo.risis.eu/gadm> {
11    ?s2 a GADMV:AdministrativeArea ;
12    GADMV:level "0"^^xsd:integer ;
13    dcterms:title ?title2 ;
14    geo:geometry ?polygon2 . }
15  FILTER (regex(?title1,?title2, "i") && bif:st_intersects ( ?
    polygon1,?polygon2)) }
```

6 Step 5 - Data to Service

In addition to a SPARQL endpoint²⁴ provided for Semantic Web users, we also exposed a set of predefined SPARQL query templates as RESTful Web services to facilitate use of the interlinked data by developers who are unfamiliar with the SPARQL query language. The Web services also allow for better management of data access (in case authentication and authorization are needed) whilst monitoring the data usage to optimize the queries and to provide load balancing on the services infrastructure (due to reasons of data size and performance of the respective geospatial queries, scalability of Linked Geo Data platforms is a critical issue [6]). We used Swagger²⁵ to document the APIs of the exposed Linked Geo Data services²⁶. The APIs are generally categorized as following:

- Find administrative boundaries containing a given point (e.g. Point ToOSMAdmin).

²⁴ <http://sparql.sms.risis.eu/>.

²⁵ <http://swagger.io>.

²⁶ <http://api.sms.risis.eu/#/Geo-Services>.

- Find details of a given administrative boundary (e.g. `OSMAdmin`).
- Find (multi-)polygon shapes of a given administrative boundary (e.g. `OSMAdminToPolygon`).
- Find FUAs related to a given administrative boundary (e.g. `BoundaryToOECDFUA`) or a given point (e.g. `PointToOECDFUA`). In case of adaptive FUAs, for a given indicator, the service will return its corresponding FUA.

Invoking the services will result in executing the SPARQL query templates filled in with the given input.

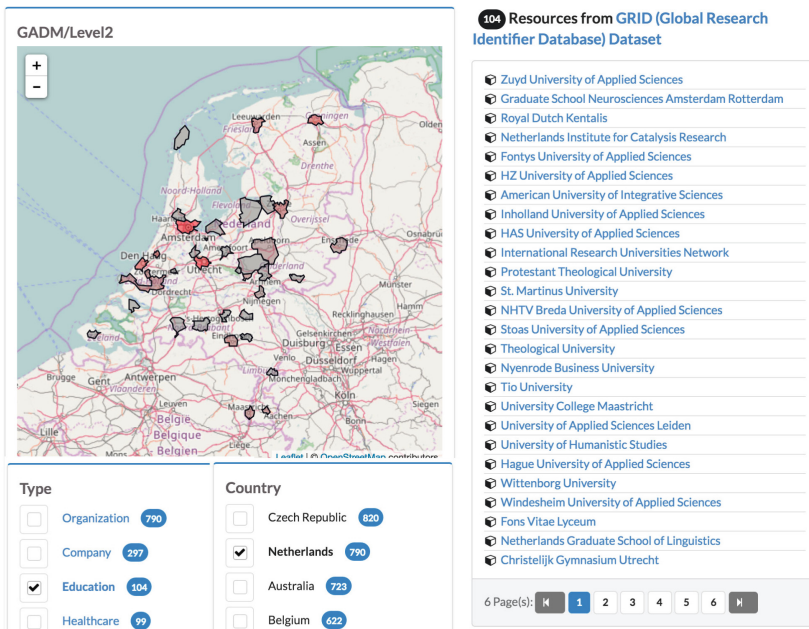


Fig. 4. Exploring <http://grid.ac> dataset using the extracted geo boundaries.

7 Step 6 - Service to Application

An important benefit of exposing data as service is the ability to combine one or more services with other existing services and applications to build novel and innovative applications. With regards to our domains of interest, we created several applications to better demonstrate the value of the provided services²⁷. Utilizing GoogleMap and Mapbox APIs to explore a dataset based on the extracted boundaries was one example of these applications. For instance, Fig. 4 shows our geo-boundaries faceted browser [5] which allows users to browse a map with areas delineated based on different attributes of a dataset.

²⁷ <http://sms.risis.eu/demos>.

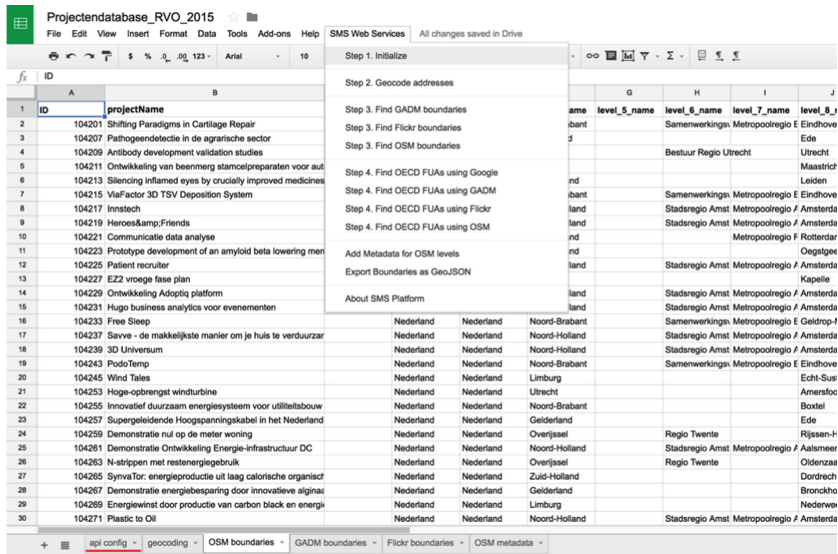


Fig. 5. The screenshot of the Google spreadsheet add-ons for geocoding addresses using SMS Web services (available at <http://sms.risis.eu/demos>).

Another practical application we built for batch processing of addresses was a Google spreadsheet add-on, depicted in Fig. 5, which chains Google Geocoding API with our `PointToAdmin` and `AdminToFUA` services (see Sect. 6). Given addresses in a spreadsheet are enriched with different levels of administrative boundaries and FUAs. The users are then able to export the extracted boundaries and process them in geodata analysis tools such as `CartoDB`²⁸.

In order to evaluate our spreadsheet add-ons, we performed a controlled usability case study with 20 participants of the `RISIS` geo summer school²⁹. Participants included researchers in the science & technology domain from different European research institutes with no knowledge of Semantic Web and Linked Data who wanted to enrich their datasets using our proposed Linked Data services. In the first part of the evaluation, we explained the idea of Linked Open Data in general and then specific Linked Geo Data services provided by our SMS platform were presented. We then asked them to install our spreadsheet add-ons and follow the steps to geocode their datasets using different sources and levels of administrative boundaries and then connect them to FUAs. In the second part, we asked them to fill in the questions recommended by *System Usability Scale* (SUS) [7] system to grade the usability of the app. SUS is a standardized, simple, ten-item Likert scale-based questionnaire³⁰ giving a global view of subjective assessments of usability. It yields a single number in the

²⁸ <https://carto.com/>.

²⁹ <http://risis.eu/event/geography-training/>.

³⁰ www.usabilitynet.org/trump/documents/Suschart.doc.

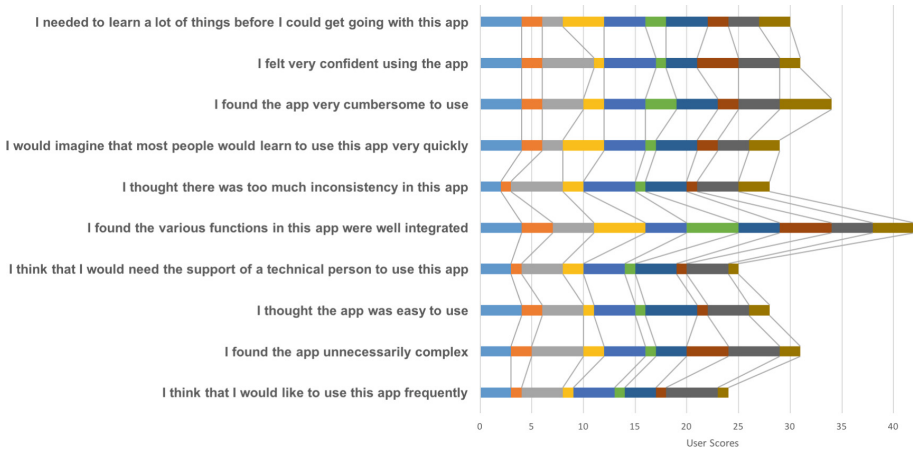


Fig. 6. Result of the SUS questionnaire evaluating our spreadsheet add-ons.

range of 0 to 100 which represents a composite measure of the overall usability of the system. The results of our questionnaire filled by 10 of the participants showed a mean usability score of **71.5** for our add-ons which indicates a good level of usability. Figure 6 shows the scores per user for the SUS questions. In addition to quantitative results, we also collected a number of user suggestions to further improve the application. For instance, to Important feedback from participants include; facilitate initial setup, add more metadata for GADM and Flickr boundaries, and clarify possible level of detail of an address field.

8 In-Use Case Study: Adaptive Delineation of FUAs

We applied the provided services and applications of linked geo-boundaries to several use cases within the context of the RISIS project. RISIS EU project³¹ aims to build a distributed infrastructure on data relevant for research and innovation dynamics and policies [9]. One of the objectives in the project is to integrate different science and technology (S&T) datasets centered on the geographical dimension and thereby propose a S&T map of Europe. To achieve this goal, geographical harmonization of different datasets in the S&T domain seems necessary. Therefore, Functional Urban Areas (FUAs) are employed as unit of harmonization. To create an S&T map of Europe we reconstructed and dynamically delineated FUAs from open datasets, starting with the following actions (linked to the steps in our approach described in previous sections):

- Find S&T related indicators which refer to different levels of administrative boundaries; and are suitable for the underlying study.
- Identify a weighted subset of open geo boundaries called adaptive FUAs to serve as unit of geo harmonization (Using steps 1–3 in Sects. 2, 3 and 4).

³¹ <http://risis.eu>.

- Geocode addresses in the targeted datasets to enrich the datasets with geo coordinates (Using steps 4 & 5 in Sects. 5 and 6).
- Identify the corresponding adaptive FUAs in different sources/levels surrounding the extracted coordinates (Using steps 6 in Sect. 7).
- Compare datasets based on the identified FUAs (Using step 6 in Sect. 7).

We conducted a concrete case study in this direction which is both exploratory, to gain insight into how a researcher in the science & technology domain (a co-author of this paper) can create an S&T map, and descriptive in nature, to illustrate what results (delineation in an S&T map based on different attributes) are achieved and how these results can be interpreted. We investigated the effect of socio-economic and structural properties of the urban areas on innovative activities, as stimulated by recent RTD³² policies in the Netherlands. This policy is oriented at the ‘top sectors’ of the economy, which were selected in a consultation of policy makers, representatives of the research system and entrepreneurs in the country. After selecting these ‘top sectors’, a large part of public research funding was devoted to this new policy. Consortia can apply for funding, and they should exist of companies and research organizations (such as universities) with a company as main applicant. Because of this context, the funded projects can be considered as a useful representation of *RTD collaboration* for innovation.

In this use case we were interested in the geographical properties of these collaboration networks. In order to investigate this we needed data about the projects, and statistical data about the characteristics of the geographical units. These data are openly available on The Dutch data portal³³. In this case, we employed the following open datasets:

- *RVO dataset*³⁴ provides a list of R&D projects that have received subsidies and financial support from the Netherlands Enterprise Agency³⁵. Projects information includes companies and research institutes which are collaborating on the project together with the geographical coordinates of the projects.
- *CBS dataset*³⁶ published by the statistics office of the Netherlands³⁷ provides different types of statistical information on dimensions such as labour, income, economy, society and regional aspects of regions in the Netherlands.

As we did not know *ex ante* what the level of geographical organization of the consortia was, we needed to define these in different granularities. This enabled us to find out at what geo-level the consortia were organized. We could then identify the characteristics of these geographical ‘containers’ of the projects. To realize that, we first calculated different sets of Urban Areas based on different statistics

³² Research and Technology Development.

³³ <https://data.overheid.nl>.

³⁴ <http://www.rvo.nl/open-data-van-rvonl>.

³⁵ Rijksdienst voor Ondernemend Nederland: RVO.nl.

³⁶ <https://www.cbs.nl/en-gb/our-services/open-data>.

³⁷ Het Centraal Bureau voor de Statistiek: CBS.nl.

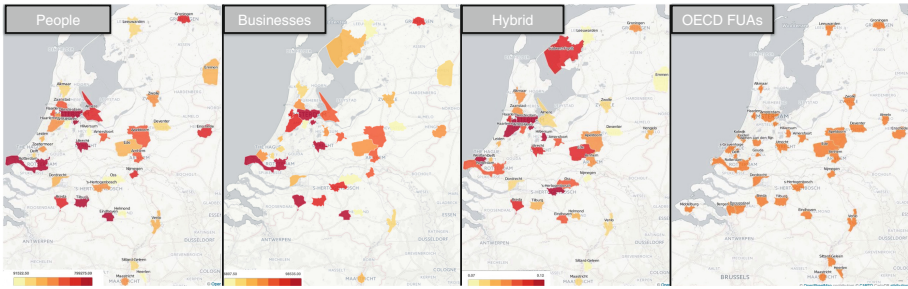


Fig. 7. An example of the adaptive delineation of FUAs for the Netherlands based on the open statistical data (populations, business establishments, hybrid and OECD).

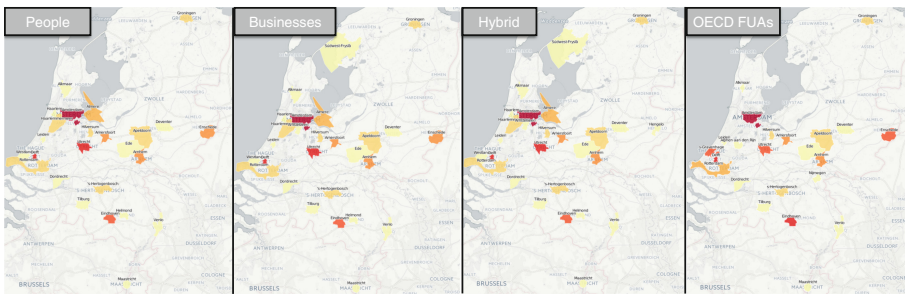


Fig. 8. Amount of RVO project subsidies mapped to the dynamically delineated FUAs defined based on the CBS open statistical data and OpenStreetMap boundaries.

provided by the CBS dataset and different levels of open administrative boundaries. Figure 7 shows the delineation of these Urban Areas through population, business establishment, and combinations of these two indicators in the municipality level. Boundaries typically differ when defined by different characteristics. When compared to the OECD FUAs³⁸ (right map in Fig. 7), the adaptive Urban Areas take into account additional regions (administrative boundaries) and enable the user to put different weights for the delineated boundaries which could be used for focused analysis of specific factors.

Our open data linking (step 4 in Sect. 5) allowed us to then map geographical coordinates of RVO projects to these FUAs (as a baseline for analysis of different S&T indicators) using SPARQL queries, to analyse the correlation of projects to the designated socio-economic factors. Figure 8 shows the result of the mapping where frequency of the projects on different factors are highlighted: the darker the color, the higher the number of awarded projects. As can be seen when comparing Figs. 6 and 7, by far not all (Functional) Urban Areas have projects. But more importantly, the different ways the Urban Areas are defined leads

³⁸ We used the shapefile from the Eurostat used in the large study named urban audit: <http://ec.europa.eu/eurostat/web/cities/data/database>.

to different outcomes. Using the OECD FUAs (right map), or the population density based FUA (left map) would miss some of the relevant areas³⁹.

As an additional use case, we also worked on exploiting background knowledge provided by DBpedia, WikiData and other open datasets to find the relation between the structural properties of the universities (e.g. number of students and size, or position on rankings) and properties of their container FUAs (e.g. various demographic and socio-economic characteristics). For example, have the locations of the higher ranked universities systematically different characteristics than lower ranked universities? Describing this use case is beyond the scope of this paper.

9 Conclusion and Future Work

With more than half the world's population now living in urban areas, defining a metropolitan area is critical to reflect the reality of where people live and work as well as the connections between surrounding cities, educational institutes, and businesses. The issue of comparability of metropolitan areas needs an in-depth, dynamic and multi-faceted analysis of administrative boundaries bringing together data about all the influencing factors. The OECD has already defined functional urban areas to address factors beyond the predefined city boundaries, and to better reflect the economic geography of where people live and work.

In this work we report an approach, implementation, and case study on the use of Semantic Web and Linked Data technologies to establish an open data space to more flexibly delineate FUAs by intergrating spatial and non-spatial data from the openly available data sources on the Web. We describe how our approach allows dynamic recreation of FUAs using linked open data and we illustrate our implementation in case studies involving researchers in the science & technology domain. In addition to better geographical coverage, our integration of Flickr, OpenStreetMap, and GADM open boundaries enables researchers and government policy makers to have different views on urban areas⁴⁰. GADM as a curated dataset focuses mainly on formal administrative boundaries while OpenStreetMap and Flickr boundaries as social crowdsourced datasets provide more details and flexibility in defining boundaries.

As future work, we envisage to (1) analyze the quality of data by applying the services to several real-world scenarios defined in the RISIS⁴¹ project; (2) create more connections to the relevant datasets on Linked Open Data cloud; (3) design intuitive user interfaces for end-users to explore FUAs while combining several indicators.

³⁹ in this case South-West Friesland is missing in population based indicators because it is less populated but still hosts a large set of businesses.

⁴⁰ We write about *urban* areas because we reconstructed FUAs, however, our approach and implementation supports delineation of any functional area.

⁴¹ <http://risis.eu>.

Acknowledgement. We would like to thank our colleagues from the Knowledge Representation & Reasoning research group at Vrije Universiteit Amsterdam for their helpful comments during the development of our approach for delineation of functional urban areas. This work was supported by a grant from the European Union's 7th Framework Programme provided for the project RISIS (GA no. 313082).

References

1. Redefining "Urban": A New Way to Measure Metropolitan Areas. OECD (2012)
2. Definition of Functional Urban Areas (FUA) for the OECD metropolitan database. OECD (2013)
3. Athanasiou, S., Hladky, D., Giannopoulos, G., García-Rojas, A., Lehmann, J.: GeoKnow: making the web an exploratory place for geospatial knowledge. *ERCIM News* **96**, 2014 (2014)
4. Auer, S., Lehmann, J., Ngonga Ngomo, A.-C., Zaveri, A.: Introduction to linked data and its lifecycle on the web. In: Rudolph, S., Gottlob, G., Horrocks, I., van Harmelen, F. (eds.) *Reasoning Web 2013*. LNCS, vol. 8067, pp. 1–90. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39784-4_1
5. Khalili, A., Loizou, A., van Harmelen, F.: Adaptive linked data-driven web components: building flexible and reusable semantic web interfaces. In: Sack, H., Blomqvist, E., d'Aquin, M., Ghidini, C., Ponzetto, S.P., Lange, C. (eds.) *ESWC 2016*. LNCS, vol. 9678, pp. 677–692. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-34129-3_41
6. Kritikos, K., Rousakis, Y., Kotzinos, D.: Linked open GeoData management in the cloud. In: *Proceedings of the 2nd International Workshop on Open Data, WOD 2013*, pp. 3:1–3:6, New York, NY, USA, 2013. ACM (2013)
7. Lewis, J.R., Sauro, J.: The factor structure of the system usability scale. In: Kurosu, M. (ed.) *HCD 2009*. LNCS, vol. 5619, pp. 94–103. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02806-9_12
8. Patroumpas, K., Georgomanolis, N., Stratiotis, T., Alexakis, M., Athanasiou, S.: Exposing INSPIRE on the semantic web. *Web Semant. Sci. Serv. Agents World Wide Web* **35**(Part 1), 53–62 (2015). *Geospatial Semantics*
9. van den Besselaar, P., Khalili, A., Koudous Idrissou, A., Schlobach, S., van Harmelen, F.: SMS: a linked open data infrastructure for science and innovation studies. In: *Peripheries, Frontiers and Beyond; proceedings of the 21st STI Conference*, pp. 106–114. University Valencia (2016). <https://www.dropbox.com/s/u5rk2bzxaupssdn/20160914%20SMS%20STI.pdf?dl=0>
10. Saquicela, V., Espinoza, M., Piedra, N., Terrazas, B.V.: Ecuadorian geospatial linked data (2014)
11. Singleton, A.D., Spielman, S.E., Brunsdon, C.: Establishing a framework for open geographic information science. *Int. J. Geogr. Inf. Sci.* **30**(8), 1507–1521 (2016)
12. Stadler, C., Lehmann, J., Höffner, K., Auer, S.: LinkedGeoData: a core for a web of spatial open data. *Semant. Web J.* **3**(4), 333–354 (2012)
13. Vilches-Blázquez, L.M., Villazón-Terrazas, B., Saquicela, V., de León, A., Corcho, O., Gómez-Pérez, A.: Geolinked data and inspire through an application case. In: *18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 446–449. ACM (2010)