Using Foursquare place data for estimating building block use

Spyridon Spyratos
Digital Earth and Reference Data Unit, Institute for Environment and Sustainability (IES), Joint Research Centre (JRC), European Commission, Italy; Department of Planning and Regional Development, University of Thessaly, Greece

Demetris Stathakis
Department of Planning and Regional Development, University of Thessaly, Greece

Michael Lutz
Digital Earth and Reference Data Unit, Institute for Environment and Sustainability (IES), Joint Research Centre (JRC), European Commission, Italy

Chrisa Tsinaraki
Digital Earth and Reference Data Unit, Institute for Environment and Sustainability (IES), Joint Research Centre (JRC), European Commission, Italy

Abstract
Information about the land use of built-up areas is required for the comprehensive planning and management of cities. However, due to the high cost of the land use surveys, land use data is outdated or not available for many cities. Therefore, we propose the reuse of up-to-date and low-cost place data from social media applications for land use mapping purposes. As main case study, we used Foursquare place data for estimating nonresidential building block use in the city of Amsterdam. Based on the Foursquare place categories, we estimated the use of 9827 building blocks, and we compared the classification results with a reference building block use dataset. Our evaluation metric is the kappa coefficient, which determines if the classification results are significantly better than a random guess result. Using the optimal set of parameter values, we achieved the highest kappa coefficient values for the land use categories “hotels, restaurants and cafes” (0.76) and “retail” (0.65). The lowest kappa coefficients were found for the land use categories “industries” and “storage and unclear”. We have also applied the methodology in another case study area, the city of Varese in Italy, where we had similar accuracy results. We therefore conclude that Foursquare place data can be trusted only for the estimation of particular land use categories.

Corresponding author:
Spyridon Spyratos, Spatial Analysis, GIS and Thematic Mapping Laboratory, Department of Planning and Regional Development, University of Thessaly, Pedion Areos, 38334 Volos, Greece.
Email: spiros2@gmail.com
Keywords
Urban building block use, land use, Foursquare, volunteered geographic information, place data

Introduction
Management and planning of urban areas requires up-to-date land use (LU) data. LU patterns affect the demand for transport (Litman and Steele, 2014) and energy (Pérez-Lombard et al., 2008), the land prices (Cheshire and Sheppard, 1995), the crime rates and types (Stucky and Ottensmann, 2009), the factors that affect public health (Frank et al., 2006), and the urban activity patterns (Jacobs-Crisioni et al., 2014). LU data is required by various professionals and organizations, such as planning agencies, city councils, departments of transportation, utility companies, real estate companies, developers, tax authorities, law enforcement agencies, and health insurance companies (Jensen and Cowen, 1999). The availability of LU data can contribute toward a more holistic understanding of the built environment and ultimately facilitate evidence-based decision making.

Still, for many urban areas, the required LU data is not available. Evidently, extensive in situ LU surveys require a significant volume of human and financial resources. Less costly and widely utilized methods for LU and land cover (LC) identification include the use of satellite and aerial images. Such data sources can be utilized to identify the physical characteristics of an area, namely the LC, but they cannot be used for estimating the uses of built up areas. At European level, three projects collect LU and LC data, which are the European Urban Atlas project (European Union, 2011), the Land Use/Cover Area frame Survey (LUCAS) (Eurostat, 2013), and the COoRdination of INformation on the Environment (CORINE) project (EPA, 2014). Nevertheless, none of the above European projects provides LU data that can be reused in order to identify the LU of built up areas at a detailed level.

Considering the high cost and the absence of detailed LU data in urban areas, in this study, we propose the re-use of openly available place data contributed by citizens to social media applications for the production of low cost LU datasets. In the literature, geographic data contributed intentionally by citizens to Internet applications are often described as volunteered geographic information (VGI) (Goodchild, 2007) or crowd-sourced geographic information. In this study, we refer to this type of data as citizen-contributed data. Examples of popular social media applications that collect citizen-contributed data are the Facebook social networking application and the Foursquare social recommendation application. The users of these applications are contributing information about places, such as, the name of a place, its location, and its category (restaurant, museum, etc.).

The rationale of this study is to investigate whether citizen-contributed place data can be used for estimating LU in built-up areas, as a low cost alternative to the traditional in situ LU surveys. To this end, we used Foursquare place data of unknown quality for estimating the LU of built-up areas in the city of Amsterdam. Places that are contained fully or partially within building blocks, determine the type of the socio-economic activities that are hosted within these building blocks, namely their building block use (BBU). Thus, based on the type of the places contained in an urban building block, we assigned LU categories to that building block. The LU allocation was performed at the spatial scale of urban building block level and not at the building level. The reason for this is that many buildings in Amsterdam are narrow and, as a consequence, many Foursquare places are erroneously described by Foursquare users to be located in the nearby buildings. We evaluated the
accuracy of the estimated BBU dataset, by comparing it with a reference BBU dataset that has been generated using the nonresidential use dataset of the city of Amsterdam. The nonresidential use dataset of the city of Amsterdam includes surfaces of built-up areas with nonresidential use, which were recorded by LU surveyors. Finally, we repeated the methodology in another case study area, the city of Varese in Italy, where we had similar accuracy results.

Previous studies explored methods for identifying building use and LU using geographic data generated intentionally or unintentionally by citizens. Huang et al. (2013) and Fan et al. (2014) identified building use by considering the geometry of building footprint data, which has been contributed by citizens to the OpenStreetMap project. Jokar Arsanjani et al. (2013) identified LU and LC patterns using OpenStreetMap feature data. Fritz et al. (2012) developed a crowdsourcing tool called Geo-Wiki, where citizens are directly contributing LC data. Frias-Martinez et al. (2012) identified large-scale LU clusters of four categories based on geo-located tweets. Noulas et al. (2013) used the type and the popularity of Foursquare places for identifying the type of the dominant activity that was taking place in wide geographical areas of Madrid and Barcelona. Finally, several studies have used mobile phone activity data for identifying LU clusters (Pei et al., 2014; Soto and Frias-Martinez, 2011; Toole et al., 2012).

To the best of our knowledge, there is no study that has used place data for the LU estimation at building or building block level.

The remainder of the article is structured as follows. Amsterdam study area and data section describes the study area and the data. Methodology section describes the methodology followed for estimating BBU using Foursquare place data as well as for assessing the accuracy of these estimations. Results section presents the results of the accuracy assessment. In Reproducibility of the proposed methodology section, we present the result of the application of the proposed methodology, in another study area, the city of Varese in Italy and we propose a method for selecting appropriate parameter values. In Discussion section, we discuss the methodology and its limitations, and finally, our conclusions and our future work directions are outlined in Conclusions section.

Amsterdam study area and data

The availability of official LU data is essential for our study, since this data will be used as reference for the evaluation of the BBU dataset that has been estimated using citizen-contributed place data. In contrast to the official LU data, citizen-contributed place data are available for all the major European cities. The city of Amsterdam, in the Netherlands, offers detailed LU data, and is thus considered as a suitable source for the generation of a reference BBU dataset. Therefore, Amsterdam was selected as the main study area, and in particular, the area bounded by the A10 motorway (Figure 1). The A10 motorway creates a physical boundary that encloses a 72.12 km² area, which is well representing the diversity of the urban environment.

Official datasets

In this study, we have used two official datasets as reference for the evaluation of the accuracy of the estimated BBU dataset. These datasets are the official buildings dataset and the nonresidential use dataset of Amsterdam. The buildings dataset is the Basisregistraties Adressen en Gebouwen, it was retrieved in February 2015 from the geo-data portal of the VU University Amsterdam (VU geoplaza, 2015), and it had been last
updated in August 2013. The nonresidential use (niet-woonfuncties Functiekaart) dataset was retrieved in February 2015 from the geo-data portal of the city of Amsterdam (City of Amsterdam, 2015). The nonresidential use dataset includes surfaces of built-up areas with nonresidential use. These surfaces have gradually been surveyed during the period March 2010–November 2014.

The original building dataset includes 143,804 building units within the area bounded by the A10 motorway. From the building dataset, we merged the attached buildings to derive the building block dataset. In this study, a building block is defined as a unit of attached buildings that are separated by others by a street or other open space. Moreover, we removed building blocks with less than 100 m² footprint area, since, according to the building dataset, they mostly represent garages and sheds of residential premises. The cleaned building block dataset formally represented by the set \( b \) in equation (1), has 9,827 members.

\[
b = (b_1, b_2, \ldots, b_{m-1}, b_m), \ m = 9,827
\]  

The nonresidential use dataset of Amsterdam includes 23,313 surfaces within the area bounded by the A10 motorway and is represented by the set \( s \), which is defined in equation (2). These surfaces represent the area of the footprint of one or more buildings that a facility, for example, a supermarket, covers.

\[
s = \{s_1, s_2, \ldots, s_{n-1}, s_n\}, \ n = 23,313
\]  

Every surface of the nonresidential use dataset \( s_i(b^i, a^i), i = 1 \ldots n \), that is a member of the \( s \) set, has the attributes \( b^i \) and \( a^i \), where:

- \( b^i \) is the building block that geometrically contains \( s_i \). The \( b^i \) attribute is a member of the \( b \) set which is defined in equation (1).
- \( a^i \) is the official Amsterdam LU category of the \( s_i \) surface. The \( a^i \) attribute is a member of the \( a \) set which is defined in equation (3).
The official Amsterdam nonresidential LU classification is represented by the $a$ set, has eight members and is described in equation (3).

$$a = \{a_1, a_2, \ldots, a_7, a_8\}$$ (3)

Every $a_z$, $z = 1 \ldots 8$, represents one of the following LU categories. Details about the LU types that these eight LU categories and their subcategories describe, can be found in City of Amsterdam (2011).

- $a_1 = \text{Bedrijven}/\text{Industries}$,
- $a_2 = \text{Kantoren}/\text{Offices}$,
- $a_3 = \text{Detailhandel}/\text{Retail}$,
- $a_4 = \text{Horeca}/\text{Hotels, restaurants and cafes}$,
- $a_5 = \text{Maatschappelijk}/\text{Societal}$,
- $a_6 = \text{Vrije tijd}/\text{Leisure}$,
- $a_7 = \text{Parkeren and Openbaar vervoer}/\text{Parking and public transport}$,
- $a_8 = \text{Opslag and onduidelijk}/\text{Storage and unclear}$.

In order to evaluate the accuracy of the estimated BBU dataset, whose production is described in the next section, we used a reference BBU dataset. Buildings typically host facilities of diverse functions, for example, retail shops on the ground floor and offices on the upper floors. An urban building block $b_k$, $k = 1 \ldots m$, can have more than one surfaces $s_i$ of the same LU category or of different LU categories. We therefore calculated the reference BBU dataset $R_{z,k}$ using equation (4), which computes for each building block $b_k$ whether there exist or not $s_i$ surfaces for each of the eight $a_z$ LU categories. For example, as shown in Figure 2, the building block $b_1$ has both the “retail” and “offices” use.

$$R_{z,k} = \begin{cases} 
1 & \text{if } \exists s_i(b^s, a^s) \in s, \text{ where } a^s = a_z, \ b^s = b_k \\
0 & \text{if } \nexists s_i(b^s, a^s) \in s, \text{ where } a^s = a_z, \ b^s = b_k 
\end{cases}$$ (4)

![Figure 2. Building blocks with LU “retail” (on the left side of the figure) and LU “offices” (on the right side of the figure). A building block can have multiple LU categories as long as it contains at least one surface of those LU categories.](image)
Citizen-contributed data, Foursquare place dataset

Recent social and technological developments, such as the increased educational attainment and the diffusion of smartphone devices, enable citizens to collect and share geo-referenced data on the Internet. These citizen-contributed data are collected in the context of various types of human activities such as socially oriented activities, educational activities and scientific ones, for example, citizen science activities. In the literature, several terms are often being used interchangeably to describe this wide category of data. The most commonly used terms are volunteered geographic information (Goodchild, 2007), crowdsourced geographic information or user-generated geographical content.

Our aim in this study is to reuse data that has been collected by citizens in the context of socially oriented activities for LU mapping purposes. Place data is a data type that can be reused for that purpose, since the type of a place, for example, restaurant, determines the LU of the surface that this place covers. Places, from the humanistic perspective, are defined as the “enclosed and humanized space” (Tuan, 2001). In line with the above definition, places can be described as spaces enriched with human experiences and meaning (Couclelis, 1992).

In this study, we define places as surfaces where socio-economic activities occur (e.g., bars, offices, and neighbourhoods). The types of the socio-economic activities that occur in places that are hosted in building blocks are used for BBU identification.

Two social media applications that collect citizen-contributed place data were taken into consideration for the selection of the place data source. For the study area, there were available 24,486 Facebook places and 37,482 Foursquare places which were harvested using the Facebook Graph Application Programming Interface (API) (Facebook developers, 2014) and the Foursquare venues API (Foursquare, 2015b) respectively. Due to the higher volume of Foursquare place data, the Foursquare application was selected to be used in this study. The combined use of both data sources was not implemented, since it would have introduced methodological and semantic problems. This is due to two reasons:

(a) Facebook uses a different place type classification from Foursquare. Consequently, the combination of the two sources would introduce semantic problems.
(b) Facebook users determine the place locations using the Global Navigation Satellite System (GNSS) receivers of their smartphones, while the Foursquare users determine the place locations on top of web maps. Both these approaches may introduce biases, but the combination of both types of biases would affect the consistency of the methodology.

Foursquare is a social media application that allows users to discover and evaluate places. Until late July 2014, the users of Foursquare were able to declare their presence in a place, an activity widely known as “check-in”. After July 2014, this check-in feature has been moved to the Swarm application (Foursquare, 2014). Information about a place such as its geographic location, its name and its type is contributed to the Foursquare application by its users. Foursquare users are also correcting the descriptions of existing places if they are inaccurate. The geographic location of the Foursquare places is approximately determined by users who are adding a point on top of a web map. Since the places are not abstract points of the Euclidian geometry (Frank, 1996), the determination of the place point location is based on the Foursquare users’ subjective judgment. The spatial distributions of the Foursquare places and the Foursquare checked in users, in the city of Amsterdam, are shown in Figure 3. The majority of both places and checked in users are located in the city centre of Amsterdam.
Methodology

In this section, we present the methodology followed for estimating BBU using Foursquare place data and for assessing the accuracy of these estimations. The methodology consists of three steps. As shown in Figure 4, the first step is the Foursquare place data preparation, which is explained in Foursquare place data preparation section. In the second step, we calculate the estimated BBU dataset taking into consideration combinations of the parameters $d$ and $c$, as is explained in BBU estimation section. The last step, which is presented in BBU accuracy assessment section, includes the definition of the method used for the accuracy assessment of the estimated BBU dataset.

**Figure 3.** The study area is divided in pixels of 300 m by 300 m size, and each pixel is colored based: a) on the total number of places that are located in it (left side); and b) the total number of Foursquare users that have checked in places that are located in it (right side).

**Figure 4.** The methodology followed for estimating BBU and assessing the accuracy of these estimations.
Foursquare place data preparation

First, in the Foursquare data preparation phase, we aligned the Foursquare place classification to the Amsterdam nonresidential LU classification. For this alignment, only the Foursquare categories that describe nonresidential functions that exist within buildings were used. The Foursquare application uses a predefined classification for the description of the category of a place (Foursquare, 2015a). As of January 2015, there were 712 Foursquare place categories. From these categories, 633 were manually aligned to their corresponding eight Amsterdam LU categories, taking also into consideration their detailed subcategories. The remaining Foursquare categories were not taken into consideration since they were referring to irrelevant categories such as outdoor activities and residential uses. The manual alignment of the Foursquare and Amsterdam classifications was based on personal interpretation and it was made difficult due to semantic and spatial granularity differences between the Foursquare place classification and the Amsterdam LU classification. For example, the Foursquare place category “College Cafeteria” was aligned to the Amsterdam LU category “Societal,” which describes Colleges and Universities among others, and not to the “hotels, restaurants and cafes” LU category, which describes cafes.

The Foursquare place classification serves different purposes from the Amsterdam nonresidential LU classification. As shown in Table 1, those classifications have different structure. The former is used to classify places that are used for socially oriented purposes, and as a result activities that are on the main interest of Foursquare users are described in more detail. For example, the Foursquare place classification is very detailed in place categories that belong to the “hotels, restaurants and cafes (a4)” LU category. The Amsterdam nonresidential classification is used for LU mapping purposes and, as a result, it reflects the interests of urban planners and managers.

After aligning the Foursquare place classification to the Amsterdam one, we cleaned the Foursquare place dataset. As of February 2015, there were 37,482 Foursquare places within the area bounded by the motorway A10. From them, 30,036 were referring to one of the 633 Foursquare place categories, which were aligned to the Amsterdam LU and were used for the estimation of the BBU. The numbers of Foursquare places that belong to each Amsterdam LU category are presented in Figure 5. Clearly, places that refer to recreational and commercial LU categories are well represented in the Foursquare place dataset. On the contrary, places that belong to the “storage” or “industries” LU categories are underrepresented. This is due to the voluntary nature of crowdsourcing data collection. Foursquare users add descriptions of places that they are willing either to evaluate or to state their presence in them.

| Amsterdam LU category | a1  | a2  | a3  | a4  | a5  | a6  | a7  | a8  |
|-----------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| Number of LU subcategories | 7   | 7   | 24  | 13  | 36  | 51  | 9   | 2   |
| Percentage of LU subcategories (%) | 5   | 5   | 16  | 9   | 24  | 34  | 6   | 1   |
| Number of place categories | 10  | 20  | 138 | 261 | 80  | 112 | 10  | 2   |
| Percentage of place categories (%) | 2   | 3   | 22  | 41  | 13  | 18  | 2   | 0   |
The cleaned place dataset, which is represented by the set $p$, has $x = 30,036$ members and is formally described in equation (5).

$$p = \{p_1, p_2, \ldots, p_x\}, \quad x = 30,036$$

Every place $p_j(f, a^p, c^p, d^p, b^p), j = 1 \ldots x$, is a member of the $p$ set and it has the attributes $f, a^p, c^p, d^p, b^p$.

- The attribute $f$ represents the Foursquare category of a place $p_j$.
- $a^p$ is the Amsterdam LU category, which corresponds to the Foursquare place category $f$ of a place $p_j$. The $a^p$ attribute is a member of the $a$ set, which is defined in equation (3).
- $c^p$ is the total number of Foursquare users that have declared a visit, or as is widely known, have checked in a place $p_j$.
- $d^p$ is the Euclidian shortest distance between the point, where a place $p_j$ is located, and the footprint of its nearest building block. In case a place is located within a building block, then $d^p = 0$.
- $b^p$ is the building block that either geometrically contains a place $p_j$ or is the nearest building block to it.

**BBU estimation**

The estimation of BBU was performed at the spatial scale of urban building block level and not at the building level. The reason for this is that in Amsterdam, and particularly in its historic centre, many buildings are narrow, due to the fact that in the past they were taxed based on their frontage (Farmer, 1993). The determination of the position of the places that are within buildings with small frontage requires high precision. Since the location of Foursquare place data is not very precise (Figure 6), many places are falsely described to be located in the nearby attached building. Due to that reason, an earlier attempt to allocate LU to buildings failed in terms of the accuracy of the estimations. In this study,
this limitation is overcome by increasing the spatial scale of analysis from the building level to building block level.

In order to estimate the BBU for the study area, two parameters were taken into account for LU allocation in the building blocks: the parameters \( d \) and \( c \), which are presented in this section. Many Foursquare places are erroneously described to be located close to building blocks but not inside their footprint, for instance the retail shops in Figure 6. These places need to be assigned to their closest building blocks, given that the building blocks lie within a specific maximum distance. To allocate these places in the building blocks that they belong, we have introduced the parameter \( d \). This parameter describes the maximum value that the distance \( p_j(d_p) \) from a place \( p_j \) to its closest building block \( b_k \) may take, in order to include \( p_j \) in \( b_k \). The \( d \) parameter may take any integer value in the range \((0-50)\) and its highest value, \( d = 50 \text{m} \), was determined empirically from the accuracy assessment of the estimated BBU datasets. The rationale was to include all the possible \( d \) parameter values that are needed for the identification of the highest kappa coefficient (see Results section).

In order to assess how the positional accuracy of Foursquare places varies based on the application of different \( d \) parameter values, we geocoded the place addresses, and we assessed whether their address locations fall within the same building blocks as their geographic locations. From the 30,036 Foursquare places used in this study, 9,845 had complete address information, and their addresses were geocoded using the OpenStreetMap’s Nominatim service (OpenStreetMap, 2015). The results of the accuracy assessment, in Figure 7, show that as the \( d \) parameter value is increasing, the total positional accuracy of places that lie within distance \( d \) from their closest building block decreases, while the number of total places taken into account increases. Clearly, the selection of the \( d \) parameter value is a trade-off between place data quantity and place data positional accuracy.

To further assess the accuracy of Foursquare places we made use of the Linus’ law (Raymond, 2001). Linus’ law states that, the higher the number of users or contributors of a product, the higher the probability that an error will be identified or fixed by someone is. Haklay et al. (2010) demonstrate that Linus’s law is valid for OpenStreetMap data, since...
there is a positive correlation between the number of contributors and the data quality. Linus’ law applied on the Foursquare place dataset, would state that the higher the number of Foursquare users that have declared a visit in a place is, the higher is the probability that this place is accurately described. To test this assumption, we took into consideration parameter $c$, which represents the number of Foursquare users that have checked in a place $p_j(c^p)$. The $c$ parameter may take any integer value in the range $(0, 50)$, and its maximum parameter value was roughly determined as for parameter $d$. By geocoding the address of Foursquare places, as described above for parameter $d$, we assessed the positional accuracy of Foursquare places by taking into account different values of parameter $c$. As Figure 8 shows, the positional accuracy of Foursquare places increases, with an exception when $c = 1$, as the $c$ parameter value increases. This verifies that Linus’ law applies also to Foursquare place data.

Finally, taking into consideration the above two described parameters, we computed the estimated BBU dataset $E_{z,k,d,c}$. Using equation (6) we calculated for each building block $b_k$ whether places of each LU category $a_z$ are hosted in it, taking into account all the possible combinations of the parameter values $c$ and $d$. For example, as shown in Figure 9, for $c^d = 0$ and within a distance $d^p = d = 2m$ from the building block $b_2$ there are three places: place $p_3$, which belongs to the $a_5$ LU category, and the places $p_4$ and $p_5$ that belong to the $a_6$ LU category. As a result, in this building block the LU categories $a_5$ and $a_6$ are assigned.

$$
E_{z,k,d,c} = \begin{cases} 
1 & \text{if } \exists \ p_j(d^p, b^p, d^p \leq d, c^d \geq c) \in p, \\
& \text{where } d^p = a_z, \ b^p = b_k, \ d = [0 \ldots 50] \text{ and } c = [0 \ldots 50] \\
0 & \text{if } \not\exists \ p_j(d^p, b^p, d^p \leq d, c^d \geq c) \in p, \\
& \text{where } d^p = a_z, \ b^p = b_k, \ d = [0 \ldots 50] \text{ and } c = [0 \ldots 50]
\end{cases}
$$

Figure 7. Percentage of places for which both the address and their geographic location refer to the same building block versus number of Foursquare places, using the $d$ parameter values in the range $(0 \ldots 50$ m). For example, for the parameter value $d = 6$ m there are 21,000 places from which, based on the 9,845 place sample, 70% are positionally accurate.
The accuracy assessment was performed in order to evaluate the correctness of LU category assignment on the building blocks. We compared the BBU estimated datasets, which were calculated using equation (6), with the reference BBU dataset which was calculated using equation (4). This comparison was performed 2,601 times in order to take into consideration any possible combination of the $d$ and $c$ parameter values. Cohen’s kappa coefficient (Cohen, 1960), which is described in equation (7), was selected to be used for the determination of the optimal BBU classification for each LU category. The reason is that this statistical measure normalizes for the expected chance of agreement (Carletta, 1996), and it determines whether the classification results are significantly better than a random result (Congalton, 1991).

**Figure 8.** Percentage of places for which both their address and geographic location refer to the same building block versus number of places, using $c$ parameter values in the range (0 – 50).

**Figure 9.** BBU estimation using Foursquare place data and taking into consideration the parameters $c$ and $d$. In the building block $b_2$, using parameter values $c^p \geq c = 0$ and $d^p \leq d = 2m$, the $a_5$ and $a_6$ LU categories are assigned.

**BBU accuracy assessment**

The accuracy assessment was performed in order to evaluate the correctness of LU category assignment on the building blocks. We compared the BBU estimated datasets, which were calculated using equation (6), with the reference BBU dataset which was calculated using equation (4). This comparison was performed 2,601 times in order to take into consideration any possible combination of the $d$ and $c$ parameter values. Cohen’s kappa coefficient (Cohen, 1960), which is described in equation (7), was selected to be used for the determination of the optimal BBU classification for each LU category. The reason is that this statistical measure normalizes for the expected chance of agreement (Carletta, 1996), and it determines whether the classification results are significantly better than a random result (Congalton, 1991).
When there is total agreement between the estimated and the reference BBU dataset, the kappa coefficient value is one. On the contrary, when there is no agreement other than that which would be expected by chance, the kappa coefficient value is zero. Negative kappa coefficient values can also occur and they indicate agreement less than that achieved by chance (Viera and Garrett, 2005).

\[ k = \frac{p_o - p_e}{1 - p_e} \]  

(7)

The \( p_o \) value in equation (7), represents the proportion of times that the estimated and the reference BBU datasets agree for a given LU category. In detail, the \( p_o \) value is estimated using the confusion matrix presented in Table 2 as described in equation (8). The \( p_e \) value in equation (7) represents the proportion of times that the estimated and the reference BBU datasets are expected to agree by chance only. In detail, the \( p_e \) value is estimated using equation (9).

\[ p_o = \frac{TP + TN}{TP + FP + FN + TN} \]  

(8)

\[ p_e = \left( \frac{TP + FP}{TP + FP + FN + TN} \right) \times \left( \frac{TP + FN}{TP + FP + FN + TN} \right) + \left( \frac{FN + TN}{TP + FP + FN + TN} \right) \times \left( \frac{FP + TN}{TP + FP + FN + TN} \right) \]  

(9)

In addition to Cohen’s kappa coefficient, we estimated the precision, the sensitivity and the specificity for each LU category, which are described in equations (10), (11), and (12). The precision refers to the probability that a building block classified, for example, to have the “retail” LU in the estimated BBU dataset, actually has that LU category in the reference BBU dataset, as shown in Figure 10. The sensitivity refers to the probability that a building block that has the “retail” LU in the reference BBU dataset is correctly identified as having the “retail” LU in the estimated BBU dataset. Finally, the specificity refers to the probability that a building block is correctly estimated not to have the “retail” LU. The accuracy assessment results are presented in the next section.

\[ \text{precision} = \frac{TP}{TP + FP} \]  

(10)

\[ \text{sensitivity} = \frac{TP}{TP + FN} \]  

(11)

\[ \text{specificity} = \frac{TN}{FP + TN} \]  

(12)

| Table 2. Confusion matrix for the “retail” LU category. TP represents the number of building blocks that have “retail” LU in both the estimated and the reference BBU datasets. |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| | Reference BBU dataset | | | |
| | Retail | Nonretail | Retail | Nonretail |
| Estimated BBU dataset | True positives (TP) | False positives (FP) | False negatives (FN) | True negatives (TN) |
| Retail | | | | |
| Nonretail | | | | |
Results

We present the best accuracy results for each LU category of the estimated BBU dataset in Table 3. The best accuracy results refer to the optimal BBU classifications for each LU category that were produced using the set of the parameter values $d$ and $c$, for which the highest kappa coefficient was achieved when compared to the reference BBU dataset. These optimal sets of parameter values $c$ and $d$ for each LU category are presented in the last two columns of Table 3. As shown in Table 3, the highest Cohen’s kappa coefficient value, 0.76, was estimated for the “hotels, restaurants and cafes” LU category. About 75% of the building blocks with “hotels, restaurants and cafes” use were correctly identified using

![Figure 10. Building blocks estimated to have retail use (in black) versus building blocks that have retail use in the reference dataset (in black) in the “Bos en Lommer” neighborhood, Amsterdam.](image)

| Parameters         | Kappa | Precision (%) | Sensitivity (%) | Specificity (%) | $d$ | $c$ |
|--------------------|-------|---------------|-----------------|-----------------|-----|-----|
| Industries         | 0.06  | 45.70         | 5.72            | 98.51           | 43  | 0   |
| Offices            | 0.50  | 56.23         | 61.25           | 90.64           | 18  | 0   |
| Retail             | 0.65  | 72.48         | 67.14           | 95.48           | 29  | 5   |
| Hotels, restaurants and cafes | 0.76  | 82.41         | 75.12           | 97.63           | 22  | 26  |
| Societal           | 0.46  | 58.01         | 52.29           | 92.18           | 21  | 1   |
| Leisure            | 0.42  | 48.05         | 49.58           | 92.61           | 20  | 3   |
| Parking and public transport | 0.17  | 30.38         | 14.63           | 98.84           | 17  | 0   |
| Storage and unclear| 0.01  | 33.33         | 0.46            | 99.91           | 30  | 0   |

Table 3. Best accuracy results for each LU category of the Amsterdam case study. In the last two columns we present the $d$ and $c$ parameter values for which the highest Cohen’s kappa coefficient values were achieved.
Foursquare place data. The lowest kappa coefficient was achieved for the LU categories “industries,” “Parking and public transport,” and “storage and unclear,” and is for all of them below 0.2. For the other categories, the kappa coefficient ranges from 0.42 to 0.65. The specificity for all the LU categories is high, above 90%, and shows that building blocks without nonresidential LU were correctly estimated as not having such use.

The parameter values, $d$ and $c$ affect the accuracy of the estimated BBU dataset. As shown in Figure 11, the kappa coefficient for each LU category is increasing as the maximum distance between places and their closest building blocks increases up to the optimal $d$ parameter values, which range from 17 to 43 depending on the LU category. This is because some LU categories, like for example “industries,” are mostly located in sparsely built-up areas, while others such as “offices” in densely built-up areas. Regarding parameter $c$, as shown in Figure 12, for some categories with high precision, the kappa coefficient is increasing as the number of minimum checked in users increases up to the optimal $c$ value. This optimal value ranges from 0 to 26, since it depends on the characteristics of the places of each individual LU category. For example, for the category “hotels, restaurants and cafes,” which is the category with the highest number of places, the optimal $c$ parameter value is 26.

As shown in Figures 13 and 14, the precision and the sensitivity of the eight LU categories vary according to the value of parameter $d$. The sensitivity, for each of the eight LU categories, is increasing as the $d$ distance is increasing up to the optimal $d$ value. This is because more Foursquare places are assigned to building blocks, and thus more building blocks are estimated as having nonresidential LU. Consequently, the probability that a building block with one or more nonresidential LU in the reference dataset is being correctly identified in the estimated dataset is increasing. Inversely, when the $d$ distance value is increasing, the precision of the detection of the 8 nonresidential LU are slightly decreasing. This is because, as shown in Figure 7, when we take into consideration places

![Figure 11](image)

**Figure 11.** Cohen’s kappa coefficient of the estimated BBU dataset calculated using the parameter values $d = (0–50 \text{ m})$ and $c = 0$. 
that are far from building blocks, the probability that these places are positionally accurate is decreasing.

**Spatial variation of the accuracy results**

In order to assess how the accuracy of the proposed methodology varies by space, we repeated the methodology in two subareas of the case study. These subareas are the Amsterdam-Centrum, which is shown in Figure 1, and the A10-periphery, which is the area that remains when we exclude the Amsterdam-Centrum from the area that
is enclosed by the A10 motorway. As shown in Table 4, the kappa coefficient, the precision and the sensitivity in the Amsterdam-Centrum is higher than in the A10 periphery. The reason for that is that Amsterdam-Centrum is the urban centre of Amsterdam, where as shown in Figure 3, most of the Foursquare places and most of the activity of Foursquare users are located. The specificity in the A10 periphery is much higher that it is in the Amsterdam-Centrum. This is because in the A10 periphery there are many exclusively residential building blocks, which are correctly estimated as not having any of the nonresidential LU categories. The $c$ parameter values remain stable for the well-predicted LU categories, such as the “retail” and the “hotels, restaurants and cafes.” This proves that the $c$ parameter values are consistent across the same case study area. Regarding the $d$ parameter, its values vary significantly in the two subareas of the case study.

**Table 4.** Best accuracy results for the two subareas of the Amsterdam case study.

|                | Amsterdam-Centrum | A10 Periphery |
|----------------|-------------------|---------------|
|                | Precision (%)     | Sensitivity (%) | Specificity (%) | d | c | Precision (%) | Sensitivity (%) | Specificity (%) | d | c |
| Industries     | 0.08              | 56            | 8              | 98           | 35 | 0  | 0.06              | 43            | 5              | 99 | 43 | 0  |
| Offices        | 0.55              | 71            | 80             | 75           | 35 | 0  | 0.45              | 50            | 55             | 92 | 18 | 0  |
| Retail         | 0.67              | 80            | 81             | 86           | 47 | 5  | 0.60              | 67            | 63             | 96 | 27 | 4  |
| Hotels,        | 0.81              | 89            | 89             | 92           | 17 | 26 | 0.70              | 80            | 67             | 98 | 25 | 28 |
| restaurants    |                   |               |                |              |    |    |                   |               |                |    |    |    |
| Societal       | 0.46              | 65            | 56             | 88           | 21 | 3  | 0.46              | 59            | 50             | 94 | 23 | 1  |
| Leisure        | 0.43              | 65            | 62             | 81           | 9  | 3  | 0.35              | 41            | 41             | 94 | 20 | 3  |
| Parking & public transport | 0.15 | 40 | 15 | 96 | 17 | 0 | 0.16 | 31 | 12 | 99 | 17 | 4 |
| Storage & unclear | 0.00 | NA | 0 | 100 | 0 | 2 | 0.01 | 50 | 1 | 100 | 30 | 2 |
In this section, we analyze the density of places that belong to the two LU categories, for which we had robust estimations, the ‘‘retail’’ and the ‘‘hotels, restaurants and cafes.’’ As shown in Figure 15, a linear regression analysis was performed in order to assess whether the number of Foursquare places $p_j(b^p, a^p)$ for each LU category in each building block is correlated to the number of surfaces $s_i(b^s, a^s)$ of each LU category as is recorded the Amsterdam nonresidential LU dataset. The slope of the fitted line is 0.86 for the ‘‘hotels, restaurants and cafes’’ LU category and 1.2 for the ‘‘retail’’ LU category. The coefficient of determination, denoted by $r^2$, is used for the evaluation of the correspondence between the Foursquare places and the LU surfaces.

**Density of “retail” and “hotels, restaurants and cafes” LU**

In this section, we analyze the density of places that belong to the two LU categories, for which we had robust estimations, the “retail” and the “hotels, restaurants and cafes.” As shown in Figure 15, a linear regression analysis was performed in order to assess whether the number of Foursquare places $p_j(b^p, a^p)$ for each LU category in each building block is correlated to the number of surfaces $s_i(b^s, a^s)$ of each LU category as is recorded the Amsterdam nonresidential LU dataset. The slope of the fitted line is 0.86 for the “hotels, restaurants and cafes” LU category and 1.2 for the “retail” LU category. The coefficient of determination, denoted by $r^2$, is used for the evaluation of the correspondence between the Foursquare places and the LU surfaces.
Reproducibility of the proposed methodology

In this section, we assess whether and how the proposed methodology can be applied to other urban environments. For that purpose, we repeated the proposed methodology in the city of Varese, in Italy (see Assessment of the methodology in the city of Varese, Italy section), and we describe a method for selecting appropriate $d$ and $c$ parameter values in the Selection of appropriate parameter values section.

Assessment of the methodology in the city of Varese, Italy

To test the reproducibility of the proposed methodology, we repeated the study in the city of Varese, in Italy. Compared to the city of Amsterdam, the city of Varese has a different urban morphology and almost 10 times less population. The estimated BBU datasets of Varese were produced using the same methodology as the ones used for the Amsterdam case study. The accuracy assessment of the BBU estimations was performed on 150 randomly selected building blocks of the study area, which are shown in Figure 16. The reference LU data for these 150 building blocks was collected through a ground survey. For the collection of the ground survey data, we recorded both the uses that were visible at the exterior of the buildings and the uses described in the mailboxes and doorbells of the buildings. Both the ground survey and the Foursquare place data collection took place in the same time period, July 2015.

The results of the accuracy assessment are presented in Table 5. Compared to the Amsterdam case study, the kappa coefficient values for the “hotels, restaurants and cafes” LU category is 0.76 in Amsterdam and 0.73 in Varese. For the “retail” LU category, the kappa coefficient values are 0.65 in Amsterdam and only 0.48 in Varese. The Kappa coefficient values are lower in the Varese case study for the aforementioned LU categories, due to their lower sensitivity. The low sensitivity demonstrates that there is an absence of Foursquare place data in Varese. This absence of place data has as result low $c$ parameter values, since the lower the $c$ value is, the more Foursquare place data are included in the analysis. The parameter $d$ varies significantly across the LU categories.
Selection of appropriate parameter values

As revealed from the accuracy assessment results, the selection of the appropriate parameter values $d$ and $c$ is crucial for optimal accuracy results. The selection of these values depends on the purpose of the application and is a trade-off between data quality and data quantity. In this study, the kappa coefficient was used for the determination of the optimal parameter values. The optimal sets of parameter values for each LU are not consistent across the Amsterdam and Varese case studies, since they are affected by the local characteristics of the case study area (e.g., urban morphology) and the characteristics of the Foursquare dataset in that area (e.g., number of Foursquare users). For $d$ parameter values of 20 m or more, the number of Foursquare places, the kappa coefficient value, the precision, and the sensitivity of the estimation in the Amsterdam case study, remain almost stable (Figures 7, 11, 13 and 14). As a result, the precise $d$ parameter value as long as it is higher than 20 m, does not have any considerable effect on the accuracy of the estimations, and thus as a general guidance we suggest a fixed $d$ value of 25 m for all the LU categories. On the contrary, for optimal accuracy results, the $c$ parameter value needs to be chosen separately for each LU category. For the selection of the $c$ value, a calibration is suggested. Such a calibration can be performed by the application of the methodology in a sample of 5% of randomly selected building blocks of the study area, for which there is ground truth data. The optimal parameter values that have been found in the accuracy assessment phase of the sample can then be applied to the whole case study area. To test whether that calibration method is valid, we generated 20 estimated BBU datasets using $c$ parameter values as is specified in the 20 different calibrations and a fixed $d$ parameter value $d = 25$. Figure 17 shows that the majority of the 20 estimated BBU datasets that were produced using the calibration method had, when compared to the reference BBU dataset, kappa coefficient values slightly lower than the BBU dataset that was estimated using the optimal parameter values.

Figure 17. The kappa coefficient value of the BBU datasets that was estimated using parameter values $d = 25$ and $c$ as defined by calibrations on 20 randomly selected samples of the Amsterdam case study.
Discussion

There is an important difference between the estimated BBU dataset and the reference BBU dataset. Since the estimated BBU dataset is constructed using place data, it reflects how space is experienced and subjectively perceived by the Foursquare users. The reference BBU dataset reflects how space is observed by the LU surveyors. According to Tuan (1979: 389) ‘‘the space that we perceive and construct, the space that provides cues for our behaviour, varies with the individual and cultural group.’’ For example, the Starbucks stores are perceived by the majority of their users as coffee shops, but some individuals experience them as co-working spaces. The variety of how people perceive and experience space is potentially expressed in the place datasets but not in traditional LU surveys that do not involve the space users through questionnaires or interviews.

Apart from the above-described conceptual difference of the reference and estimated BBU datasets, their comparison led in a series of interesting scientific results. As an outcome of the accuracy assessments, using the proposed methodology we identified with high confidence the urban building blocks of the case study areas with at least one ‘‘retail’’ or ‘‘hotels, restaurants and cafes’’ use. On the contrary, our methodology failed to reliably identify building blocks with ‘‘industries,’’ ‘‘storage and unclear’’ or ‘‘parking and public transport’’ use. The reason for the low identification rate of some LU categories (Table 3) is that many Foursquare places that belong to these categories are missing from the Foursquare place dataset. The completeness of the Foursquare place dataset varies depending on the LU category (Figure 5), since the Foursquare users, who are mostly young smartphone owners (Zickuhr, 2013), decide independently about which places, of what type and from which area, they will add in the Foursquare dataset. Therefore, recreational and commercial places, which are the places of the Foursquare users’ main interest, are better represented in the Foursquare place dataset compared to other types of places.

In addition to the above-described limitation of the Foursquare place dataset, as shown in Figure 7, the positional accuracy of the Foursquare places’ location is low. Many places are falsely described to be located in other building blocks or close to building blocks they belong but not inside their footprint (Figure 6). As it is also valid in general for crowd-sourced data, the positional and thematic accuracy of the Foursquare place data is not assured and in-situ controlled through an established quality assurance and quality control mechanism. The resolution of errors in the description and the location of places rely on the willingness of Foursquare users, who might also not possess the required scientific and technical skills for describing the category and the location of a place accurately, or who might experience space in different ways. Apart from the characteristics of the Foursquare place dataset, the accuracy results of the Amsterdam case study are negatively affected by limitations introduced from the use of the nonresidential use dataset as a reference. These limitations include: (a) the fact that the LU survey was limited to the exterior of the buildings; (b) the fact that the nonresidential use dataset is outdated in some areas; and (c) the fact that the ‘‘storage and unclear’’ category includes, under a single nondifferentiable category, surfaces that have storage, unclear, or no use.

Conclusions

In this study, we evaluated the use of Foursquare places for estimating the use of urban building blocks. The main case study has been conducted in the city of Amsterdam, and an additional one has been conducted in the city of Varese, Italy. Based on Foursquare place data, and particularly on the type of these places (like, for example, offices), we assigned to each building block LU categories that describe the types of the activities that are hosted in
it. For the estimation of the LU categories of each building block, two parameters were used. These are the distance between a Foursquare place and its nearest building block, and the number of Foursquare users that have checked in a Foursquare place. The distance was used in order to geometrically allocate places to their nearest building blocks, and the number of checked in users in order to assess whether a Foursquare place is accurately described. The estimated BBU dataset, which represents how Foursquare users experience space, has been compared to the reference BBU dataset which represents how LU surveyors observe space.

Our evaluation metric is the kappa coefficient, which determines if the classification results are significantly better than a random result (Congalton, 1991). Among the accuracy assessment outcomes, for both the Amsterdam and Varese case studies, the highest kappa coefficient values were achieved for the individual LU categories “hotels, restaurants and cafes” and “retail.” This is because places of the above two categories are well represented in the Foursquare place dataset, since these places are of the Foursquare users main interest. On the contrary, the methodology failed to identify building blocks with “industries,” “storage and unclear,” and “parking and public transport” use. This is because for the above three LU categories many places are missing from the Foursquare dataset.

The proposed methodology can be used for the generation of up-to-date, low cost, and globally harmonized datasets about urban building blocks with “hotels, restaurants and cafes” or “retail” LU. In the near future, technological and social developments are expected to increase the quality of citizen-contributed data. The increased educational attainment, the diffusion of smartphone devices and the emergence of indoor position systems are expected to further increase the use of location-based social media applications and the precision of citizen-contributed data. In our future work, we will examine the use of other datasets, such as business directories, with an aim to improve the identification rate of LU categories that are underrepresented in the Foursquare place dataset.

Acknowledgments

We would like to thank Massimo Craglia, Sven Schade, Elena Roglia, Chris Jacobs-Crisioni, Henk Scholten the editor and two anonymous reviewers for their valuable comments and suggestions. Additionally we thank Klaas-Bindert de Haan for supporting us regarding the nonresidential use dataset of Amsterdam.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: This research has been supported through a Ph.D. research grant of the Joint Research Centre of the European Commission.

References

Carletta J (1996) Assessing agreement on classification tasks: The kappa statistic. Computational Linguistics 22(2): 249–254.
Cheshire P and Sheppard S (1995) On the price of land and the value of amenities. *Economica* 62: 247–267.

City of Amsterdam (2011) Functiekaart-classificatie 2011-6. Available at: https://drive.google.com/file/d/0BxRbFXteeEQjSHViaThaWUXSNGe/view.

City of Amsterdam (2015) Non-residential functions (Function Map). Available at: http://maps.amsterdam.nl/open_geodata/ (accessed 4 February 2015).

Cohen J (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1): 37–46.

Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment* 37: 35–46.

Couclelis H (1992) Location, place, region, and space. In: Abler RE, Marcus MG and Olson JM (eds) *Geography’s Inner Worlds Pervasive Themes in Contemporary American Geography*. New Brunswick, NJ: Rutgers University Press, pp. 215–233.

EPA (2014) *Corine Land Cover Mapping*. Environmental Protection Agency. Available at: http://www.epa.ie/soilandbiodiversity/soils/land/corine (accessed 17 December 2014).

European Union (2011) Mapping guide for a European urban Atlas. Available at: http://www.eea.europa.eu/data-and-maps/data/urban-atlas/tab-methodology (accessed 11 January 2015).

Eurostat (2013) LUCAS 2012 (Land Use/Cover Area Frame Survey) Technical Reference Document: C-1 Instructions for Surveyors. Luxembourg: Eurostat.

Facebook developers (2014) The Graph API. Available at: https://developers.facebook.com/docs/graph-api (accessed 12 August 2014).

Fan H, Zipf A and Qing F (2014) Estimation of building types on OpenStreetMap based on urban morphology analysis. In: Huerta J, Schade S and Granell C (eds) *Connecting a Digital Europe Through Location and Place*. Springer International Publishing, pp. 19–35.

Farmer B (1993) Needs and means. In: Farmer B and Louw HJ (eds) *Companion to Contemporary Architectural Thought*. London: Routledge, pp. 21–28.

Foursquare (2014) A brand new Foursquare, with a brand new logo and look, is almost ready for you. Available at http://blog.foursquare.com/post/92636287198/a-brand-new-foursquare-with-a-brand-new-logo-and-look (accessed 18 August 2014).

Foursquare (2015a) Foursquare for Developers – Foursquare Category Hierarchy. Available at: https://developer.foursquare.com/categorytree (accessed 18 January 2015).

Foursquare (2015b) Foursquare for Developers – Venues Service. Available at: https://developer.foursquare.com/overview/venues.html (accessed 12 February 2015).

Frank A (1996) The prevalence of objects with sharp boundaries in GIS. In: Burrough P and Frank A (eds) *Geographic Objects with Indeterminate Boundaries*. London: Taylor & Francis, pp. 29–40.

Frank LD, Sallis JF, Conway TL, et al. (2006) Many pathways from land use to health and air quality. *Journal of the American Planning Association* 72(1): 75–87.

Frias-Martinez V, Soto V, Hohwald H, et al. (2012) Characterizing urban landscapes using geolocated tweets. In: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing. IEEEEE, pp.239–248.

Fritz S, McCallum I, Schill C, et al. (2012) Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software* 31: 110–123.

Goodchild MF (2007) Citizens as sensors: The world of volunteered geography. *GeoJournal* 69(4): 211–221.

Haklay M, Basiouka S, Antoniou V, et al. (2010) How many volunteers does it take to map an area well? The validity of Linus’ law to volunteered geographic information. *Cartographic Journal* 47(4): 315–322.

Huang H, Kieler B and Sester M (2013) Urban building usage labeling by geometric and context analyses of the footprint data. In: *Proceeding of 26th International Cartographic Conference (ICC)*. Dresden, Germany: International Cartographic Association.

Jacobs-Crisioni C, Rietveld P, Koomen E, et al. (2014) Evaluating the impact of land use density and mix on spatiotemporal urban activity patterns: an exploratory study using mobile phone data. *Environment and Planning A* 46(11): 2769–2785.
Spyridon Spyratos is a PhD student at the University of Thessaly, Greece. He holds an MSc degree in Geographical Information Science from University College London, UK (2011) and a BSc degree in Urban and Regional Planning from the University of Thessaly (2010). He worked at the Joint Research Center of the European Commission as a trainee (2011–2012) and as a PhD researcher (2012–2015). His research focuses on the re-use of geo-referenced data from social media applications for land use mapping and site evaluation purposes.

Demetris Stathakis is an associate professor at the Department of Urban and Regional Planning Engineers (PRD) of the University of Thessaly (Volos, EL). He previously (2004–2008) served as a postdoctoral researcher at the Joint Research Center of the European Commission (Ispra, IT) on GIS and remote sensing. He graduated and obtained his doctorate degree from PRD (1994, 2003). Since 1995 he holds an MSc in Geodetic and Geographical Information Systems of the University of London.
(UCL, UK). His current research interests include monitoring, analyzing and forecasting urban sprawl.

**Michael Lutz** holds an MSc degree in landscape ecology and a PhD in geoinformatics from the University of Münster, Germany. He has 15 years of experience in the areas of semantic modelling of data and processes to support the discovery, access, use and integration of geographic information. Michael is the technical coordinator of the implementation and maintenance of the Infrastructure for Spatial Information in Europe (INSPIRE). In the framework of the EU’s ISA² programme on interoperability solutions and common frameworks for European public administrations, businesses and citizens, he is also investigating how INSPIRE principles, tools and guidelines can be used in other sectors.

**Chrisa Tsinaraki** holds a diploma (1995), MEng (2000) and PhD degrees (2008) in computer engineering from the Electronics & Computer Engineering Department of the Technical University of Crete (Chania, Greece). Since 2013 she is a postdoc researcher in the Joint Research Centre of the European Commission. In the past she has worked as a researcher at the Technical University of Crete and the University of Trento and as a visiting lecturer at the Technical University of Crete. Her research interests include open data, interoperability support, big data, digital libraries, multimedia semantics, multilingualism, event modelling and visualization, and citizen science.