Privacy-aware geolocation interfaces for volunteered geography: a case study

Maria Luisa Damiani
University of Milan, Italy
damiani@di.unimi.it

Colette Cuijpers
Tilburg University, NL
C.M.K.C.Cuijpers@uvt.nl

ABSTRACT

The standard W3C Geolocation API can significantly facilitate geospatial data collection as it provides a simple set of operations for requesting geolocation services across indoor and outdoor spaces through the Web. Importantly, this API is privacy-aware in that it provides a basic privacy mechanism for requesting the user's consent to location acquisition. In this paper we address the question on whether this privacy mechanism is sufficient to conduct a project for the collection of geospatial content, in compliance with privacy laws. The question is of practical relevance as the use of geolocation standards in line with privacy regulations would make the development of volunteered geography projects easier. In this paper we present an interdisciplinary analysis spanning across technology and law, and driven by an application case. We show the limitations of this API and discuss a possible extension in line with privacy norms. Although we confine ourselves to consider European regulations, we believe that this study can be of more general concern.

Categories and Subject Descriptors

H.2.8 [**Database management**]: Database applications—*Spatial databases and GIS*; K.4.1 [**Computers and society**]: Public Policy Issues—*Privacy*

General Terms

Legal Aspects,Human Factors,Standardization

Keywords

Privacy, location-based services, geolocation standards

1. INTRODUCTION

Volunteered geography (VGI) embraces the wide spectrum of applications which aim at collecting geospatial content through the direct involvement of citizens contributing to the construction or enrichment of maps or of some other geography-grounded model [11]. Citizens can participate in different ways to geospatial data collection. For example, they can provide fine-grained knowledge of the territory and annotate shared maps, as in wikimapia[1]. Or can act as mobile sensors, for example to report the location of illegal dump sites to some organization which can verify the information and report it to the appropriate local authorities. Also, citizens can collaborate to the sampling of individuals' behavior and to the creation of virtuous cycles through closed loop solutions. For example, car drivers can choose to disclose their footprint to enable traffic monitoring in real time and in this way reduce traffic congestion.

In this paper we are concerned with privacy issues posed by these scenarios. Privacy, along with trust and credibility, is widely recognized as one of the key challenges in volunteered geography [11, 7]. Privacy is especially relevant when users leave traces of their movement. While sporadic locations of a mobile device may not be particularly sensitive, the historical trail of past locations, i.e. the user's trajectory, can reveal much about a user's behavior.

In most countries, location data cannot be collected without providing users with privacy guarantees in compliance with law. As an example, consider a participatory project aiming at studying the social habits in a city, for instance where citizens spend most of their time, which are the social services they use and so forth. Volunteers living in the city could be equipped with a smartphone and provided with a simple application which periodically transmits their location to an application server. Analytical methods can then be applied on the collected data to discover mobility patterns as in [17]. Even though this initiative is carried out for legitimate and valued purposes, i.e. improving the quality of social life, it remains the problem of how to practically deal with the collection of personal location data in compliance with data protection law. The problem stems from the fact that location represents personal data and as such can only be processed in compliance with data protection legislation which contains obligations for data controllers (e.g. application providers) and rights and guarantees for data subjects (i.e. users). For example, the processing of location data under EU law may only be allowed with prior consent of the data subject. In what follows, we refer to this contractual view of privacy as *notice-and-consent model*.

Notice-and-consent privacy model. It is the dominant privacy model in on-line applications [14]. A way to put in place such a model in our previous example is by recruiting a limited number of participants in advance and

[1]http://wikimapia.org

ACM SIGSPATIAL GEOCROWD'12 Nov. 6-9, 2012. Redondo Beach, CA, USA

requesting them to formally grant their consent to the collection and processing of location information, after providing them with appropriate information. However, ensuring that consent is actually informed is by far not trivial when the community is open and individuals can dynamically adhere to the data collection initiative, typically through the Web. For example Nissembaum coined the term *informed consent paradox* to mean that the more information is provided to users, the less their understanding [14].

In order to provide stronger protection of personal data (i.e. not only location), different communities advocate the need of novel privacy rules and concepts, such as the concept of personal data store [8], and, in social sciences, the notion of privacy-in-context [14]. Moreover, computer scientists have developed a broad set of privacy enhancing techniques to enable the users of on-line applications to exercise some form of control on the way the location is collected by third parties as well as to prevent attacks from untrustworthy parties, such as [13, 5]. Yet, to the best of our knowledge, none of these techniques are deployed in real applications.

W3C geolocation API. In this paper, we focus on the privacy issues posed by a specific class of on-line services which seem relevant for VGI, namely the geolocation services and in particular those provided through the W3C geolocation API. The W3C geolocation API (simply API hereinafter) is qualified as *W3C recommended standard* since May 2012 [15]. This API provides the abstract specification of a set of operations, which embedded in Web pages, enable to estimate the location of the website visitors. Notably the specification is privacy-aware in that it states that users shall provide their explicit consent to location acquisition, thus in line with the notice-and-consent model. This API is embedded in HTML5, the platform *that is revolutionizing the way the Web evolves, works and is used* [1]. Importantly this API is supported by all major Web browsers.

Research question and contribution. Our research is driven by the following question: *can we rely on the privacy mechanisms offered by this API to conduct a participatory project for the collection of personal location data, in compliance with data protection law?* The question has important practical implications as the use of geolocation standards compliant with privacy regulations would greatly facilitate the development of participatory applications.

In this paper, we argue that the privacy offered by the API is not enough to comply with the European data protection law. Therefore additional measures need to be undertaken by the developers and this inevitably impacts on the complexity of the application. This is especially true when the Web applications are provided by small organizations or amateurs, which, as such, cannot easily count on the support of a legal staff. To support our argument, in what follows we present an application in which location data from a community of students are collected in a participatory way on Web. Based on this experience, we present an analysis of some legal aspects and discuss a possible extension of the API to provide more transparency to users.

The remainder of the paper is organized as follows: Section 2 presents a brief overview of location privacy enhancing techniques in on-line applications. Background knowledge on geolocation services and data protection legislation in Europe is provided in Section 3. Section 4 presents the application case and the related privacy analysis. Section 5 oulines the proposal of an extended API. Final considerations are reported in Section 6.

2. RELATED WORK

In this section we overview technological approaches to the protection of location privacy in on-line applications. Actually, we are not aware of significant location privacy solutions specifically designed to support participatory data collection. However, there are techniques developed in other settings which can be usefully transposed in this domain.

A first stream of work regards anonymous network communication, e.g. onion routing techniques [10]. Using this technique, the users of our participatory data collection project can communicate their location without disclosing the true IP address. Anonymous communication is a viable approach if users can be deprived of their identity. In reality, in our scenario, identity may be required to increase the level of trustworthiness of the collected data. Moreover, even though users can participate anonymously, it is the case that users can be re-identified based on their location, such as home. In such a case, supplementary techniques are needed such as location k-anonymity. e.g. [13] and mix zones, e.g. [2]. Moreover in certain circumstances, as we will see later on, certain service providers do not accept anonymous requests.

Another stream of related work regards the protection of location information in location-based services (see [12] for a survey). For example, cloaking is a popular privacy strategy which consists of degrading the accuracy of the location information. However, whenever the cloaking method is applied to every location, the quality of the collected data and thus the effectiveness of the participatory effort can be easily compromised. More appropriate for the problem at hand are the cloaking methods which apply to a subset of locations, to protect, for example, those locations that are considered as sensitive. Semantic location cloaking methods [5] address this requirement without excessively compromising the utility of the collected data.

Policy-based solutions is another important research line. In this case, machine enforceable authorization rules are specified to regulate the access to shared location information [9]. For example, we can envisage applications in which the location data is disclosed to different parties, depending on the situation at hand. The use of privacy policies within VGI applications is however substantially unexplored. Finally we mention an important contribution related to the analysis of the privacy issues in the W3C geolocation API [6].

3. BACKGROUND KNOWLEDGE

Before presenting the key features of the W3C geolocation API, we overview geolocation techniques commonly available on popular devices, e.g. smartphones. Next, we briefly describe those aspects of the European data protection regulation that are relevant to understand the limitations of the API.

3.1 Geolocation techniques and services

3.1.1 Local vs. third party positioning

An important distinction for privacy is between local and third party positioning services. For example, GPS positioning provides the user with strong privacy guarantees because the location is estimated by the receiver installed on the device. However, GPS positioning presents heavy limitations. It is power-consuming, moreover the user must be located in a convenient location for the location to be estimated, i.e. outdoor and in light-of-sight with satellites. As people spend most of their time indoor, the fraction of time in which the GPS signal can be heard is likely marginal. That motivates the popularity of positioning services which leverage the telecommunication and public/private Wi-Fi infrastructure to provide seamlessly geolocation services both indoor and outdoor. These services are provided by third parties, i.e. the location providers (LP).

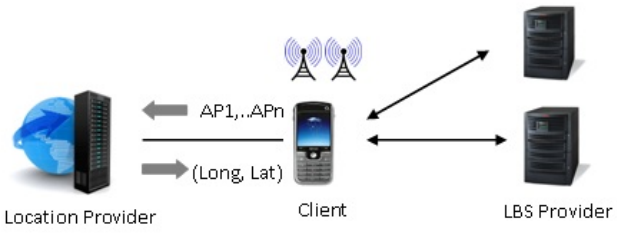

Figure 1: LBS Architecture: the client transmits to the LP contextual information (e.g. set of Wi-Fi access points, AP_1,..AP_n) to obtain the location

Figure 1 illustrates the typical architecture of a location-aware application in which the location is computed by a LP. The client requests the location from the LP and then forwards such location along with the service request to the LBS (location-based service) provider. Typically, LPs compute the location based on the contextual information sent by the client, e.g. Wi-Fi access points in proximity (e.g. $AP_1.AP_n$), by properly matching these patterns against a large and proprietary database of geo-referenced access points and cell towers.

In an urban setting, with high density of Wi-Fi networks, the location returned by the LP can have an accuracy of 10-20 meters. Note that users can interact with diverse LBS providers offering different services, therefore LPs are in a location to compile extensive records of users' location and movement. If LPs are untrustworthy, privacy is at stake. We will come back to this point later on.

3.1.2 Requesting geolocation services

Geolocation requests are encoded using machine readable operations. Operations can be defined at different levels of abstraction. For example, at operating system level, the Android system provides three classes of operations, for using GPS-based positioning, for requesting geolocation services from the LP and for handling passive positioning, respectively[2]. As the location can be determined using different techniques, it is up to the LBS provider, in charge of the application development, to choose the most appropriate method, based on the application requirements and the user's context. This requires a certain programming skill.

The W3C geolocation API provides a much simpler way to request the geolocation services. First, the API is agnostic of the positioning technology, therefore the geolocation service can be requested without specifying how the location is to be estimated. Secondly, operations can be embedded in Web pages, using a simple scripting language like Java Script. Therefore, as the API is endorsed by all major Web browsers, visitors can be localized across different operating systems. Note however that this simplicity of use is paid in term of flexibility, in that the LBS provider is not in the position of exercising any control on the way the location is computed, in particular on whether the location is computed locally or through a LP while this is important for privacy. Indeed, what happens in practice is that the Web browser translates such operation into a geolocation service request for the LP[3]. Accordingly, any time the user is geolocated the location is communicated to the LP.

3.1.3 Technical features of the API

For the sake of completeness, we briefly overview the key aspects of this API related to the geolocation functionalities, in particular the information provided by the LP and the operations for handling the geolocation request. The privacy-related aspects will be presented later on in the paper.

- **Position information returned upon the request of geolocation service**. Some of these data are provided on developer's discretion (i.e. are optional):
  - Timestamp of the location
  - Position as latitude and longitude
  - Altitude (optional)
  - Accuracy of the location (Long, Lat) in meters
  - Accuracy of the altitude (optional)
  - Heading (the direction is expressed as angle with respect to North) (optional)
  - Speed in meters per second (optional)

- **Geolocation operations.** The interface provides two main operations, to get the single location and to continuously update the location, respectively. Note that the latter operation allows users' tracking. The abstract specification of these operations is reported in Figure 2, while a succint description is provided here below:
  -*getCurrentPosition()*: this operation returns the location of the client at the time in which it is requested
  -*watchPosition()*: this operation is to request the repeated updating of the location until the operation is explicitly stopped (i.e. using the clearWatch operation).

3.2 Location data protection in Europe

In view of the privacy issues raised in this paper, it is important to have some basic understanding of the general Data Protection Directive and the so-called ePrivacy Directive, the main pillars of the EU legal framework regarding processing of personal and location data.

3.2.1 Data Protection Directive (95/46/EC)

The general Data Protection Directive (hereafter: DPD) consists of a layered system of three levels. The first level

[2]http://developer.android.com/guide/topics/location/strategies.html

[3]http://www.mozilla.org/en-US/firefox/geolocation/

```
interface Geolocation {
  void getCurrentPosition(PositionCallback successCallback,
                          optional PositionErrorCallback errorCallback,
                          optional PositionOptions options);

  long watchPosition(PositionCallback successCallback,
                     optional PositionErrorCallback errorCallback,
                     optional PositionOptions options);

  void clearWatch(long watchId);
};

callback PositionCallback = void (Position position);

callback PositionErrorCallback = void (PositionError positionError);
```

Figure 2: Abstract specification of the geolocation operations

is the general level that applies to all processing of personal data. The second level is applicable when sensitive data are being processed. The third level is applicable when personal data are being transferred to third countries. The layered system is cumulative, meaning that if sensitive data are being transferred to third countries, all three levels apply. The issues we address in this paper mainly concern the first level, therefore the second and third level will not be discussed. Besides the three levels within the DPD, the EU legal framework on data protection consists of two more levels of protection, including Directive 2002/58/EC (ePrivacy Directive).

The DPD applies to the processing of personal data which is defined as "any information relating to an identified or identifiable natural person (data subject)", while processing covers "any operation or set of operations which is performed upon personal data". Both concepts are interpreted in a very broad manner. As a main rule, the DPD stipulates that personal data "may only be processed fair and lawful". What fair and lawful entails, can be derived from the other provisions in the Directive. Main requirements: having a specific purpose and legitimate basis for the processing of personal data. Processing is only allowed in accordance with the specified purpose and may not go beyond this purpose. Regarding the quality of the data it is determined that data must be relevant, accurate, not excessive and up to date. Sound security measures also need to be taken in order to protect data from being corrupted or destroyed. Furthermore, the data controller has the obligation to inform data subjects (and in some cases the Data Protection Authority) regarding data processing. Data subjects have the right to access, rectification, erasure, blocking and the right to object. To ensure enforcement, Member States are obliged to put in place effective sanctioning mechanisms in case of infringement of the data protection rules.

The obligations in the DPD are addressed to the data controller, defined as the entity that "determines the purposes and means of the processing of personal data". In view of the actual processing, a controller can engage a processor, an entity "which processes personal data on behalf of the controller". However, the controller is responsible for and must put in place processing contracts with their data processors. As we will see in the legal considerations in Section 4.2.1, it can be rather difficult to identify the controller and processor(s) when the W3C geolocation API is used to provide a LBS. This is however very important, as it is the controller who must comply with all rights and obligations laid down in the legal framework.

3.2.2 The ePrivacy Directive (2002/58/EC)

The provisions of Directive 2002/58/EC (as amended by 2009/136/EC) particularize and complement the DPD in the field of electronic communications services. Importantly this Directive lays down rules regarding the processing of location data, in particular it states that: "location data may only be processed when made anonymous or with prior consent and only for the duration necessary for the provision of a value added service", being: "any service which requires the processing of traffic data or location data other than traffic data beyond what is necessary for the transmission of a communication or the billing thereof". The main difference between the two Directives concerns the fact that the regime to process location data in providing value added services is stricter than the general regime regarding the processing of personal data. Prior consent and anonymisation are the only valid grounds for processing location data. In practice, the only valid ground in most cases will be prior consent as the Art. 29 WP has stipulated that "true anonymisation is increasingly hard to realise and (...) the combined location data might still lead to identification" [16]. Therefore, the core legal consideration addressed in Section 4.2.1 concerns the difficulties of the requirement that location data may only be processed with the prior consent of the data subject, i.e. user.

4. APPLICATION CASE

The application case consists of a Web application involving the participants of a recent seminar on location privacy[4] in an experiment of location data collection. The purpose of this project is manifold: i) to experience an active learning approach to location privacy based on user's participation; ii) to highlight location privacy issues related to the use of W3C geolocation API; iii) to collect data on the accuracy of the geolocation service offered by popular LPs throughout Europe. This application has been developed in the framework of the EU project Modap (www.modap.org).

4.1 Location data collection through the Web

The Web application is conceptually simple: it estimates the location of each user visiting the Web site from any device, mobile or not, and stores this information together with the user's IP and few additional information in a repository on the server. Users can then inspect the content of the repository.

The location is estimated as follows: first the application invokes the geolocation operation provided by the API (i.e. the getCurrentPosition operation). If the location is correctly returned, it is recorded in the repository along with additional information. Conversely, if the location is not available, because for example the user denies his/her consent to location collection, then the application computes a coarse location based on the IP address and using commercial datasets reporting the association between IPs and geographical locations. In this case, the location is determined locally. Note that the application does not include extra code for privacy support beyond what is provided by the API. The application, called WhereAreYouNow (Wayn) is available on http://wayn.modap.org. The participants of the seminar - more than 50 students (PhD and M.Sc stu-

[4]Summer School on Privacy-aware Social Mining, mss2012.modap.org

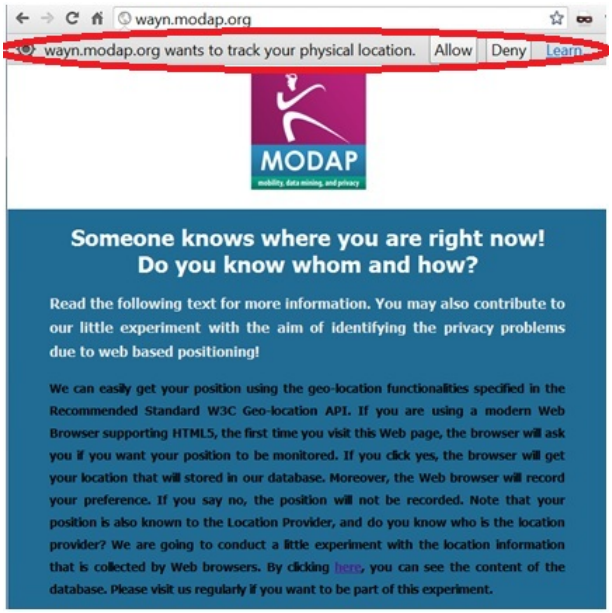

Figure 3: As the Web page is accessed the user is prompted with the request of consent (in red)

dents) from 13 countries mainly in Europe - were invited to visit the Web site in advance and leave their footprints.

4.1.1 Accessing the Web site

The interaction with the Web site is as follows: the visitor of the Web site, using for example Google Chrome, is presented with the page in Figure 3. The text explains the purpose of the data collection project and asks visitors to contribute to it. The first time the user visits the Web site, the Web browser asks the user to grant or deny his/her consent to the disclosure of location. Note that the interaction is completely transparent to the application (i.e. LBS) provider. The user can then reply yes or no or even not reply. If the user replies yes then the Web browser records the preference, i.e. consent will not be requested again, unless the user deletes this preference, then forwards the geolocation request to the LP. Conversely, if the user denies consent or does not reply in a given amount of time (e.g. 10 seconds), our application records the IP and further information, e.g. time.

In addition, users are provided with a feedback on their location. Feedback is important to increase user's awareness and enhance user's experience. Feedback is provided by displaying on a map the logged locations along with the additional data stored in the repository. The map in Figure 4 shows what the users can see, i.e. the content of the repository. This map can be accessed through the home page of the Web site. The map reports a pin for each record, i.e. visit. Pins are displayed in different colors based on the accuracy of the corresponding location. The most accurate are the locations identified by blue pins. Generally these locations are estimated based on either GPS or through Wi-Fi-based positioning. The sky colored pins are those typically detected by the LP based on the IP address. This happens for example when the user connects to the Web site using a cabled connection and no other geo-enabling device is on

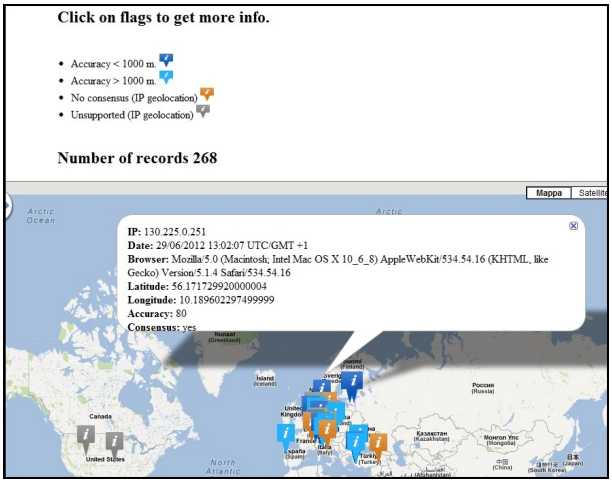

Figure 4: The map displays the content of the repository

(Wi-Fi or GPS). Conversely, if the user has denied consent or has not replied in the given amount of time the application displays the coarse location estimated based on the IP address and the pin is orange. Finally, gray pins are used in those situations in which the LP cannot compute the location, for example because of an error, or because the geolocation request is not accepted.

Each pin is shown along with the information recorded in the database, including IP address, time, coordinates, location accuracy and the Web browser footprint (i.e., Web browser configuration). One additional information that is currently not displayed to users, is the unique code, acting as pseudonym, stored as cookie on the mobile device. The pseudonym is used to correlate locations from the same user i.e. build the user's trace. If the user removes the cookie, a new pseudonym is created and thus also a new corresponding cookie. Note that this practice is left outside the legal analysis in this paper as currently a lot of debate exists within the EU regarding how to practically deal with the revised legal regime on cookies. However, the legality of this practice constitutes an interesting question.

4.1.2 Lessons learned

Before discussing the specific privacy issues which emerge from this experiment, it is worth mentioning interesting outcomes, of more general interest, of this application:

- *Creating privacy awareness in education.* This experiment allowed us to illustrate how location can be used to break the apparent anonymity of users and reveal personal details. For example, it may be relatively easy to discover where a person lives based on his/her trace and temporal information. For example, if an individual has accessed the Web site early in the morning from a residential area (the typology of area can be identified for example through StreetView), it is likely that the location is in proximity of user's home. Moreover, if the same person has left a footprint at the seminar location (we remind that locations of the same users can be correlated), the identification is fairly straightforward.

- *Access constraints to geolocation services.* A few participants tried to visit the Web site anonymously using the Tor onion routing system[5]. In that case, the geolocation request does not return a location, but an error. The gray pins displayed on the map in Figure 4 report points which are completely meaningless as the location is computed based on the IP of one of the nodes of the onion routing infrastructure. In essence the LPs are generally aware of the user's IP address.

- *Insights into the accuracy of geolocation services.* From this experiment, it turns out that the accuracy achievable from a LP is a few tens of meters uniformly across the different countries, such as Estonia, Turkey and Switzerland. Note however that official statistics from LPs are not available. Ideally, by promoting a larger scale participatory project, it would be possible to gain deeper understanding on the capabilities of third party geolocation services and come up with useful statistics. We leave this point for future work.

4.2 Privacy issues

We turn to consider the privacy issues which emerge from this experiment. We recall that no additional support for privacy is provided beyond the basic mechanism offered by the API. Therefore users have only limited control over the disclosure of their location, as they can only accept or deny consent. Moreover, the experiment makes clear that users have insufficient information to express informed consent. We report some informal considerations here below. Then we analyze the question from a legal point of view.

- Users are typically not aware of the third party that behind the scenes computes the location, i.e. the LP. Even though certain Web browsers somehow provide this information, the actual implications are not clear to most of the users.

- As a consequence, the denial of consent is not correctly understood. Actually denying consent means that the location is not computed by the LP. This, however, does not prevent the application from computing the location in some other way, for example based on IP.

- Users are not aware of the purpose of the Web application, unless it is explicitly provided by the application itself. This simply follows the fact that the API operations only display the domain name of the Web application requesting the location, while there is no way to add additional information on the privacy policy.

- Users are not aware of the location actually transmitted, unless using application-dependent functionalities, as we have seen before. Users are surprised of the high accuracy that can be achieved, whenever GPS is not used.

- Users are not aware on whether the location is just computed once or repeatedly, i.e. users are tracked. This because the current implementation of the standard does not make clear which geolocation operation is precisely used (i.e. getCurrentPosition or watchPosition, see Section 3.1.3). Therefore it may happen that the user believes to give his/her consent to the disclosure of the single location while in reality the user is tracked for the time the Web page is accessed.

- Users are not aware of the fact that by law they could request the removal of information recorded in the repository.

[5]https://www.torproject.org/)

4.2.1 Legal considerations

In the introduction the question is raised whether we *can rely on the privacy mechanisms offered by the W3C geolocation API to conduct a participatory project for the collection of personal location data, in compliance with data protection law.* In this section we assume the applicability of the ePrivacy Directive without entering into the discussion whether or not the ePrivacy Directive is applicable. For a deeper understanding of this very complex discussion we refer to [3]. Moreover, we discuss some of the legal considerations that are directly relevant in view of amending the API, while acknowledging a much wider array of legal implications in need of more in depth research.

We focus in particular on the difficulties regarding the legal requirement of prior consent when use is being made of the standard W3C geolocation API. Prior consent, as a legal ground for processing location data, is only valid when consent is specific, freely given and based on information which is clear, comprehensive, understandable for a broad, non-technical audience and permanently and easily accessible. In relation to geolocation services on smart mobile devices the Art. 29 WP has clarified the concept of consent: it cannot be obtained freely through mandatory acceptance of general terms and conditions, nor through opt-out possibilities; the default should be that location services are 'OFF'; when switched on with consent, the user must continuously be warned that geolocation is switched on; data subjects need to renew their consent after changes have occurred in the service and even without changes reaffirm their consent at least once a year; besides the possibility to withdraw consent at any time, there must be a simple means, free of charge, to temporarily refuse the processing of location [16].

In view of the information to be provided in order to obtain valid prior consent the ePrivacy Directive requires information regarding the type of location data, the purposes and duration of the processing and whether the data will be transmitted to a third party for the purpose of providing the value added service. From the ePrivacy Directive it also follows that the processing must be restricted to persons acting under the authority of the provider of the public communications network or publicly available communications service or of the third party providing the value added service. Furthermore, processing must be restricted to what is necessary for the purposes of providing the value added service. From the foregoing three main legal considerations emerge regarding: *valid consent; restriction to persons acting under authority; and, restriction to what is necessary for the purpose.*

Valid consent. When use is being made of the W3C geolocation API Figure 3 shows that the user is informed that wayn.modap.org wants to track the physical location of the user. The user has the option to click the yes button, the no button, to click learn more or not to click anything at all. A user might expect that the *learn more* button will lead to information provided for by Wayn, but it leads to

the privacy policy of the Web browser. Even though here some information can be found on things like withdrawal, the validity of the consent can still be questioned. For example is providing a *learn more* button that refers to a Web page of the Web browser "prior information provided for by the provider of the value added service"? And what about other characteristics such as the fact that consent is only requested once, and other legal requirements such as the need to make ongoing tracking visible for the user at all times. As the standard W3C geolocation API does not leave any room for Wayn to customize the location request, the only option for Wayn in order to comply with data processing legislation is to provide all information necessary in order to obtain valid prior consent on the Web site that appears with the location request. As the request appears in a separate dialogue box on top of the page, it is questionable if the user is completely aware of the relation between the Web content and the request (aggravated by the fact that the learn button leads to the Web browser). From the perspective of Wayn it might not be preferred to provide long lists of information on its Web page, or worse, Wayn might not have this information available as will become clear below. Moreover, the information will probably be rather confusing (remember the information paradox) as Wayn will need to explain how other parties, such as the Web browser and the Location Provider are involved in the processing of personal and location data, while in fact Wayn has no control over this process whatsoever. This leads to the second legal consideration.

Restriction to persons acting under authority of. Even though we consider Wayn to be the controller as it is Wayn wanting to learn its users positions, from a legal perspective this qualification is questionable. It is not Wayn who determines the means of processing, but the Web browser/Location Provider. Wayn has no control over the type of technologies and data that are being used to pinpoint the users' location. Also, not Wayn but the user determines the Web browser. So, as Wayn cannot specify the type of positioning technology to use, including the specification whether the position should be determined locally or by a third party, it might not even be possible for Wayn to provide users with all the information necessary to obtain valid consent. This means we encounter two legal problems. First, there is no possibility for Wayn to exercise any control over the way in which the user's consent is obtained. And second, Wayn might not even be capable to provide proper information regarding the data processing process, which is required in view of valid prior consent. So Wayn must rely on this information to be properly provided by the Web browser, which party is chosen by the user.

This leads to a whole array of interesting legal questions regarding the processing of personal and location data by the third parties involved within the process of providing LBS and the lack of clarity for the average user regarding the involvement of these parties. A user of a Web site expects to be dealing with the owner of the Web site when he is asked if his location can be pinpointed, while in fact it is his Web browser that acquires the location information and then passes this onto the Location Provider, a party probably not even known to the user. The LP then redirects the service request of the user, along with his location information, back to the Web site Wayn. So the Web browser as well as the LP are aware of IP address and (coarse, depending on whether or not consent has been given) location of the user.

If we look at the Firefox browser we can see that it presents a seemingly clear privacy policy (http://www.mozilla.org/en-US/legal/privacy/firefox.html): the Web browser collects the Wi-Fi info and forwards the request to the LP. So the Web browser is the intermediary between the user and Wayn, but involves the LP in this process. Moreover, as such the Web browser and the LP are aware of the information and thus in a position to process the personal and location data as well, even beyond the purpose of the provider of the LBS for whom the request is being "translated". Whether or not and how this is being done is unclear to us, but also to the user whose privacy is at stake. But even when these parties only process the data as far as is necessary in view of the location request of Wayn, legal problems remain. The qualification of the Web browser and the LP under the DPD and the ePrivacy Directive is problematic as they do not fall under the strict authority of Wayn. Moreover, no formal contracts exist establishing a legal controller - processor relationship between Wayn and the Web browser on the one hand and Wayn and the Location Provider on the other hand. What the implications are of this rather complex situation is in need of more in depth research, from a computer science as well as from a legal perspective.

Restriction to what is necessary for the purpose. In relation to the requirement that processing is allowed only if necessary for the purpose, a comment can be made relating to the characteristic of Wayn that it calculates and stores a coarse location position on the basis of an IP address even when consent is denied. This does not comply with data protection legislation. IP addresses are personal data and thus the processing of this data in combination with (coarse) location data is only allowed on the basis of a legal ground. Without consent the location may not be calculated.

5. EXTENDING THE API

In the light of this analysis, we propose a possible extension of the API in the direction of transparency. For the sake of clarity, we illustrate these additional features through the screenshots of a prototype. In what follows, we refer to this set of functionalities as extended API. It is worth mentioning that the prototype has been built by simply re-defining the API operations and encapsulating the native operations in a JavaScript object. This means that the Web browser has not been extended with additional functionalities (i.e. plug-in) while the graphical interface handling the interaction with the user is not fully integrated. These limitations are not really relevant for the purpose of demonstrating the functionalities of the system. The extended API provides the following features:

- The geolocation operations are extended with one additional parameter, the *url* of the privacy policy of the application. Even though the idea is simple, we believe that this is a substantial improvement in the direction of aligning data collection to data protection norms. The privacy policy is an XML document which reports key information, one of the obligations deriving from data protection regulation.
- Users are made aware of the fact that their location

is either acquired once or conversely users are tracked. In the latter case the user can explicitly stop location tracking. As an additional functionality, users can get a feedback on the quality of the location information being acquired.

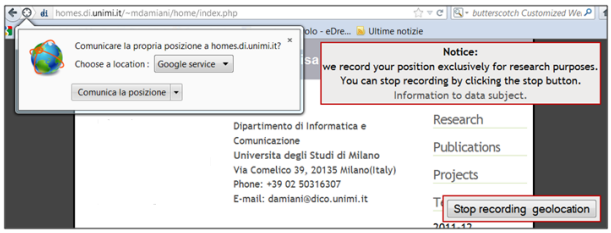

Figure 5: Extended API: displaying the notice

The prototype has been developed for the geo-enabled Web site of one of the authors of this paper, as the Web site can be quickly modified. Figure 5 shows the revisited interface which is presented to the user when the Web browser, in this case Firefox, requests the user's consent. Besides prompting the request of consent (in Italian), the geolocation operation displays two additional information items: (a) A brief notice informing the user of the general purpose of data collection. In this case, the notice specifies that location information is only collected for research purposes. This form contains a link to a more detailed notice. (b) The information that the location would be repeatedly collected (button Stop Recording geolocation). Depending of which geolocation operation is requested the user can be presented with different messages. Figure 6 reports the information which is shown to the user in response to his/her consent. This information includes accuracy, the address and the location on a map.

Figure 6: Extended API: feedback on the location

6. CONCLUSION

This API has the potential to support the development of large scale geospatial data collection projects through the Web. In this paper, we have chosen to adopt an interdisciplinary approach because the deployment of the API requires not only a technological background but also some basic understanding of the data protection regulations. Yet, a number of issues at different levels are still open and will be addressed in future work, including: extending the legal analysis beyond the European law; defining law-compliant privacy-aware data collection interfaces; and devising privacy-enhancing technologies to ensure stronger protection from untrustworthy LP, e.g. [4].

7. ACKNOWLEDGEMENTS

The work reported in this paper has been partially supported by the EU FET-OPEN project MODAP. We are sincerely grateful to Yucel Saygin, project coordinator, who supported the development of the Wayn case study and Vincenzo Pupillo who developed the software.

8. REFERENCES

[1] G. Anthes. HTML5 Leads a Web Revolution. *ACM Comm.*, Vol. 55 No. 7:16–17, 2012.

[2] A. R. Beresford and F.Stajano. Location Privacy in Pervasive Computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.

[3] A. C. Cuijpers and M. Pekarek. The Regulation of Location Based Services: Challenges to the European Union Data Protection Regime. *Journal of Location Based Services*, (5)3-4:223–24, 2011.

[4] M.L. Damiani and M. Galbiati. Handling user-defined private contexts for location privacy in LBS. In *Proc. ACM GIS '12*, 2012.

[5] M.L. Damiani, C. Silvestri, and E. Bertino. Fine-grained cloaking of sensitive positions in location-sharing applications. *IEEE Pervasive Computing*, 10(4):64–72, October 2011.

[6] N. Doty, D. Mulligan, and E. Wilde. Issues of the W3C Geolocation API. Technical report, UC Berkeley, School of Information, 2010.

[7] S. Elwood. Volunteered geographic information: future research directions motivated by critical, participatory, and feminist gis. *GeoJournal*, 72:173–183, 2008.

[8] World Economic Forum. Personal data: The emergence of a new asset class, 2011. *http : //www3.weforum.org/docs/WEF_ITTC_-PersonalDataNewAsset_Report_2011.pdf*.

[9] GEOPRIV. http://www.ietf.org/html.-charters/geopriv-charter.html.

[10] D. Goldschlag, M. Reed, and P. Syverson. Onion routing for anonymous and private internet connections. *ACM Comm.*, 42(2):39–41, 1999.

[11] M. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69:211–221, 2007.

[12] C. S. Jensen, H. Lu, and M.L. Yiu. Location Privacy Techniques in Client-Server Architectures. In *Privacy in Location-Based Applications: Research Issues and Emerging Trends*. Springer-Verlag, 2009.

[13] M. F. Mokbel and W. G. Chow, C-Yand Aref. The new Casper: query processing for location services without compromising privacy. In *Proc. VLDB*, pages 763–774, 2006.

[14] H. Nissembaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 2009.

[15] W3C. Geolocation api specification. http://dev.w3.org/geo/api/spec-source.html, 2012.

[16] Article 29 Data Protection Working Party WP185. Opinion 13/2011 on geolocation services on smart mobile devices. http://ec.europa.eu/justice/policies/-privacy/docs/wpdocs/2011/wp185_en.pdf.

[17] Y. Zheng, L. Zhang, Xie X, and W-Y Ma. Mining interesting locations and travel sequences from gps trajectories. In *WWW*, 2009.