Human Computation (2016) 3:1:143-159

ISSN: 2330-8001, DOI: 10.15346/hc.v3i1.8

GeoKey - open infrastructure for community mapping and science

OLIVER ROICK, University College London

MORDECHAI (MUKI) HAKLAY, University College London

CLAIRE D. ELLUL, University College London

ABSTRACT

The development of the geospatial web (GeoWeb) over the past decade opened up opportunities for collaborative mapping and large scale data collection at unprecedented scales. Projects such as OpenStreetMap, which engage hundreds of thousands of volunteers in different aspects of mapping physical and human-made objects, to eBird, which records millions of bird observations from across the globe. While these collaborative mapping efforts are impressive in their scale and reach, there is another type of mapping which is localised, frequently carried out over a limited period of time, and aims at solving a specific issue that the people who are living in the locality are facing. These needs are addressed in participatory mapping, which nowadays includes citizen science elements in data collection and management. The paper describes the background and design of a novel infrastructure for participatory mapping and science – GeoKey. The paper provides a differentiation between collaborative and participatory mapping, describes the state of the art and several use cases of community mapping, and the architecture of GeoKey, focussing both on the approaches to data capture and subsequent potential to share the data in an open manner where possible. It also describes the design elements that support learning and creativity in these projects.

1. INTRODUCTION

Recent years have seen a dramatic shift in the ways geographic information is produced, stored, managed and disseminated. Advances in Web technologies and growing availability of affordable GPS-equipped devices enable people from all walks of life to create, manage and share geographic data over the Web (Haklay et al. 2008). This new type of user-generated content drives successful collaborative mapping efforts, such as OpenStreetMap (Haklay & Weber 2008) and WikiMapia (Goodchild 2007). Major companies, like Google (through their Map Maker programme) or TomTom's TeleAtlas, employ collaboratively mapped information to improve their services by updating their base-maps or applying the data to generate real-time traffic information. These projects, which can involve hundreds of thousands of participants focus on creating and improving base-maps to be used in web-mapping applications or on mobile devices as context maps for a variety of geospatial applications. Others, such as Crowdmap from Ushahidi allows the mapping of information that is provided by participants using simple mobile devices and Short Messaging Service (SMS) – especially in situations where it is urgent to transfer information to understand the situation on the ground (Starbird 2011).

Importantly, from the geographical information technologies perspective, these approaches are straightforward applications that use collaborative mapping to survey physical features. Their aim is to describe and represent physical objects, or hold information that can be easily linked to a geographical place (e.g. an incident report in Crowdmap). While these applications are very successful in their reach and functionality, there is a class of mapping that is not well served by them, namely participatory mapping.

Collaborative mapping (where the main aim is to record location information accurately and efficiently, as in OpenStreetMap) and participatory or community mapping are different. In the latter the aim is to produce maps that are meaningful to the producers, respect their ownership over the information, allow them to select representations that express what is important to them and potentially empower them to take control over the decisions that will influence their life (Sieber 2006; Verplanke et al. 2016). These maps aim to stimulate collective discussions, spatial learning and information exchange that empower participants to affect decision making and policies. Participatory mapping therefore focuses on local communities consisting of people who share common objectives and priorities as well as common local knowledge on a specific area. Participatory mapping technologies build on a rich history of participatory mapping that started in the 1970s (Cambers 2006), which in the mid-1990s evolved into the area of Public Participation Geographic Information Systems (PPGIS) or Participatory GIS (PGIS) (see Sieber 2006 for a detailed discussion). Participatory Mapping is mostly focused on a bounded geographic area and frequently limited in time as the maps are used within a specific community goal. While people do collect facts, the emphasis in participatory mapping projects is on discussing, annotating and expressing opinions about a certain object or group of objects, at times with less regard to the exact location of the object but rather on the general location. For example, noting that an area is frequented by youth and therefore noisy is not about bounding a polygon around few streets, but indicating an approximate location. Collaborative mapping, on the other hand, is in general open for everyone to contribute information; mapping does not serve a certain geographic extent or a specific objective; the actual use case is not determined up-front. Ease-of-use in data entry, analysis and sense making is critical, as the maps are intended to be used by a very large group of people, some of them with low technical and map-reading abilities.

Another area of importance for participatory mapping is citizen science. Citizens have been supporting scientific research since it started in the seventeenth century by contributing observations, analysing data and disseminate research results. Recently, the collection of geographic scientific data has been simplified by the advent of affordable GPS receivers and GPS-equipped smartphones (Bonney et al. 2014). In addition to simple observations – such as bird counts – mapping for citizen science also includes subjective perceptions that are combined with sensor measurements, for instance on environmental issues such as noise or air pollution. For strictly scientific projects the quality of the data is of utmost importance as it widely influences the quality of the analysis result. However, when integrated within a participatory mapping context, there is a need to carefully balance cost, resources and practical aspects of data collection to allow high quality scientific data collection and analysis which is combined with perceptions and discussions as in the general case. Therefore, collected information has to be benchmarked against some calibrated reference data, but can also serve local needs and goals. For example, during a noise mapping campaign around London City Airport participants were asked to articulate the noise level through subjective annotations (such as acceptable, annoying, or exhausting) that were then correlated to measurements of a noise meter. This form of citizen science, or, as the Public Laboratory of Open Technology and Science prefer to call it 'civic science', includes the accurate and precise data collection through a range of instruments, with the important inclusion of meanings, annotations and discussions. An example for such integration is provided in Local Ground (van Wart 2010) – a project created at UC Berkeley School of Information – that utilises printed paper maps, which are then scanned and processed using their Local Ground platform to update and enrich information portrayed on base-maps. Against this background, this paper focuses on an open source platform that was developed to support the implementation of participatory mapping – GeoKey (www.geokey.org.uk) – both in terms of data capture and in terms of subsequent open data sharing. GeoKey is utilizing existing open-source code bases as much as possible and further propose an open API that allows users of third-party applications or social media sites to easily integrate with it by implementing connector modules that operate as gateways between our system and third-party applications. GeoKey was developed with partial support of the EU FP7 project Citizen Cyberlab, with a goal to support learning and creativity within a participatory mapping project. We review such use of the application towards the end of the paper.

GeoKey builds on the lessons that we learned from our community mapping system, which was developed between 2007 and 2014 (Ellul et al. 2011a, 2011b). In this paper we provide the rationale and the design outline of the GeoKey system. We start by describing the current state of the art in participatory mapping systems, and the potential use cases for participatory mapping from previous work that was carried out at UCL. Based on these, we explain the objectives and design rationale for GeoKey and its general architecture. We complete the report with a demonstration of an application that is using GeoKey capabilities and chart its future directions.

2. EXISTING PARTICIPATORY MAPPING PLATFORMS AND USE CASES

Besides the aforementioned collaborative mapping systems (e.g., OpenStreetMap), a number efforts emerged that aimed at implementing concepts of participatory mapping by providing means to collect and present individual environmental perceptions. One of the most comprehensive example is Geolive (Tudge et al. 2012), which was developed in the University of British Columbia by Jon Corbett and colleagues as a closed system that was run by the university, and an application of the Crowdmap system that was mentioned above. While the Geographic Information Systems (GIS) market leader has a programme call Esri Community Maps program (Esri 2015), it is important to notice that this use of the term is in the US context, where it means a settlement. Therefore, the program is aimed at allowing those with municipal responsibilities to manage their geospatial data in a single location using Esri's tools, and not to explicitly engage the public (though it is possible to do so).

Geolive and Crowdmap support participatory mapping by providing the means to post text, photos or videos to a map by assigning the piece of media to a place or location. Instead of focussing on mapping of geographic features to create high quality topographic maps, these applications address collection of participant's perceptions in the environment. Geolive was developed with special attention to the principles of participatory mapping, and does that in a controlled environment, whereas Crowdmap lean more towards the crowdsourcing model with less restrictions on the submissions and integration of information. Other projects are more focussed on a specific topic or domain. For example, the Green Map System (Perkins 2007) utilises collaborative online mapping to gather natural, cultural and social places that manifest a way of sustainable living, which are led through a participatory localised process, and are controlled by the contributors.

The services that are proposed here focus on either collecting descriptions of environmental perceptions through various types of media – such as texts, videos or photographs – or surveying information encoded in a set of predefined attributes. However, a combination of both approaches, for instance by integrating the data within one system or providing means to dynamically combine the data through public APIs, can enrich the collected data with media and in turn increase the amount information and quality for the user working with the data.
In many systems, once the data is created and stored, it is kept within the service and is not accessible through third-party applications that can help to exploit, analyse or visualise the information other than through the capabilities the respective service provides. This is mostly because of missing platform-independent APIs or the fact that the data can only be accessed using vendor-specific proprietary software, as with ESRI Community Maps Program (Esri 2015), or the way it is currently handled in Mapping for Change platform of Community Maps (Ellul et al. 2011b) or by Geolive. Providing a means to access the data through third-party applications, again through public APIs, increases the possibilities to exploit, visualise and analyse the data and extract information in manifold ways.

Furthermore, in most cases data contributed to a platform is in general available for anyone to view and access. However, in some case this may not be desired since the uploaded data may be a threat to someone's privacy, or to community control over the data, or can lead to serious risks to participants as adversaries can find details of who collected a piece of information that they find threatening. Therefore, a fine-grained management of user access down to the feature level is needed in participatory mapping platforms. Geolive already adopts the premise of ownership control as it allows for specifying which users can access the stored data.

In current systems, data is often directly integrated or fused with existing datasets. Options to correcting and editing data are often not available. This is a problem, since the data is contributed by general participant who can be expected to unintentionally add errors into the data, or might not represent the views of the community (if the specific approach that is used in the case is toward a consensus map). Crucial quality control and mechanisms that allow for correction of errors are important to ensure a certain level of data quality, and general agreement on the content of the system. However, users have to be encouraged to constantly contribute new as well as review and edit existing data. Adapting features from online social networks can increase user commitment to a collaborative project. For instance, newsfeeds on recent mapping activities can initiate further actions, as users are able to discover errors or missing features. Game-like elements, such as leader-boards that allow to compare one's performance with that of friends or honour badges for a certain number of edited features can keep users motivated to carry on mapping. Taking these limitations as a starting point, and in order to make the requirements for GeoKey clearer, we now look at three use–cases of the system, using projects that have been running at UCL in recent years.

2.1 Assisting Forest Communities to Fight Illegal Poaching

Indigenous people in the Congo basin highly depend on natural resources for their livelihoods. They are, on the other hand, rarely involved in management decisions relating to their area. In order to allow these people to express their wishes as well as identify violations of regulations, rugged smartphones can be used to record observations of illegal logging and over-hunting by commercial poachers. This information can be applied to put pressure on authorities and to make local law enforcement more effective (Stevens et al. 2013).

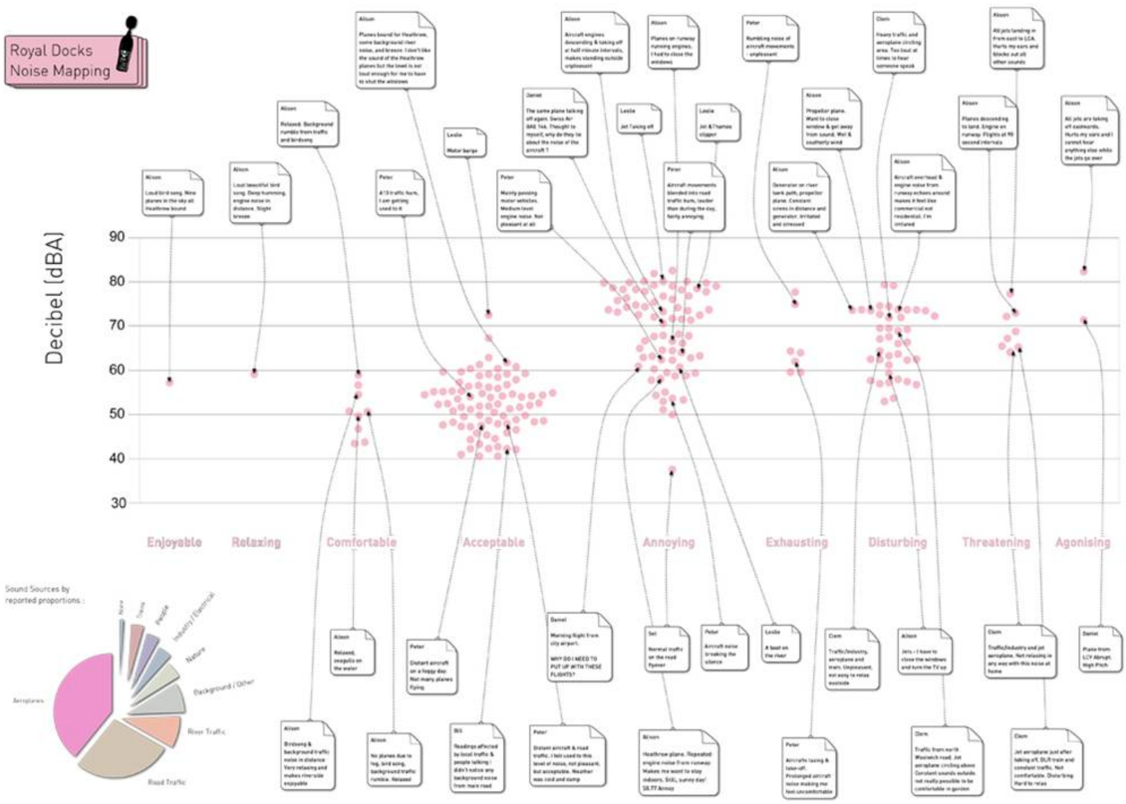
The data captured is highly sensitive and therefore should not be accessed and used by the general public as open data. A central storage and management system therefore requires the means to authorise single users or user groups, while explicitly excluding others. Moreover, the process of agreeing who has access to the information is based on Free and Prior Informed Consent (FPIC - see Lewis et al. 2008). The process requires that the community will have full control over information sharing and therefore it is required that information will not be disseminated beyond trusted parties. Moreover, some information might have an internal cultural sensitivity (e.g. should be shown only for certain people in the community while hidden from others). At the same time, other parts of the information that can be collected in the field (e.g. the location of roads or names of villages) can be shared freely.

2.2 Noise Mapping at London City Airport

In 2008, plans for changing the operations of London City Airport caused concern to residents who are living adjacent to it with regard to the level of noise pollution caused by increasing the number of flights. Using affordable noise meters residents were able to collect noise recordings over a period of two months. That data, in turn, was applied to prove correlation between high noise levels and the airport's working hours. In 2010, they requested access to the community mapping system to carry out further measurements by themselves. In the second set of measurements, residents were further encouraged to include a personal annotation of how they perceive the noise level using a set of predefined words, such as relaxing, annoying or disturbing. The captured data this needed to include not only the measurements but also to provide space for a description of contextual details, discussion (e.g. these flights are especially noisy today) and multimedia (photographs, recording) and tags. With the growth of applications such as NoiseTube (Maisonneuveet al. 2009) and WideNoise (Baker et al. 2013) which enable people to measure noise with their smartphones, there is also the need to enable the potential of adding data streams from associated websites, or redirecting the data to a specific community map. The community might also want to record specific themes (layers) of information, such as the location of vulnerable facilities (e.g. schools) or the location of people who work in shifts (who are less tolerant to daytime noise). The data need to be extracted as a map, but also to allow manipulations to produce different visualisations that combine qualitative and quantitative information, in which a cluster of points indicates the observed dB(A) measurements,linked to a description, with some of the statements by participant in baloons (see Figure 1) – these visualisations represent the community views on noise and the airport.

2.3 Mapping Community Perceptions

A final example for participatory mapping is provided from another project that was carried out in 2008, as part of the UrbanBuzz set of projects. In this case, participatory mapping was employed to gather information on how residents perceive their environment, be it positive or negative aspects. In the London suburb of Marks Gate residents were encouraged to express their opinions about aspects that they would like to act and change, such as anti-social behaviour spots and environmental issues. While the mapping helped to solve issues around the neighbourhood, residents did not want to share the information with the general public in the form of a web based map, due to concerned about external perceptions of their area. They also wanted to moderate the information and ensure that the representation expressed their consensus view and discuss it in a community meeting. While the community was interested in using digital tools to collect information, they preferred not to have an online map. Therefore, they wanted to be able to download the information under their control and visualise it locally. This example shows that easy data collection means are required as well as the option to access the data but in a way that access is limited only to contributors.



<u>Figure 1.</u> Collaborative representation of noise in Royal Docks (design by Christian Nold).

This use-case demonstrates the need for flexible data-structures and means to compile new sets of properties and their value ranges for each feature type, and the need to support creative visualisations and knowledge presentations by the participants.

3. OBJECTIVES FOR AN OPEN PARTICIPATORY MAPPING PLATFORM

Based on the characteristics and limitations of existing systems, the requirements derived from various community projects, and with an aim of developing a platform that will enhance the ability of organisations large and small to establish a robust platform for participatory mapping, we set out to develop a new platform, with a leading principle of 'open from the start' – knowingly choosing an open architecture based on an open-source stack of components. A concept note (on which part of this paper is based) was written in October 2014, containing an analysis of existing participatory-mapping platforms, an evaluation of requirements of different participatory mapping projects and a proposal for an open architecture for the new platform. This concept note was shared with experts in citizen science and participatory mapping across the world to get feedback on our plans and to adjust our proposal according to the feedback we got by the end of November 2014. The time to consult the experts and gather feedback involved not only comments on the document itself but also various in-person and online meetings, which led to new insights that were then integrated into the concept note. Based on the final version of the concept note, we implemented a first prototype of the new system. This involved an evaluation of different possible software stacks that can be used as a foundation for development, the design of a data model and its implementation in a database schema and the implementation of server-side components including an open Application Programming Interface (API). The main objectives for GeoKey are covered in the following sections.

3.1 Separation of client- and server-side components

A clear separation of server-side components and client-side applications allows for de-coupling data storage and management capabilities from accessing applications. This model has been proven to be very successful in OpenStreetMap, especially in terms of rich set of end-user applications that rely on the same API, and supports creativity among developers, as well as adapting to local needs or specific local contexts. The separation provides the ability to create multiple front ends that interact with the underlying data in multiple ways. The GeoKey platform provides means to store, access and update the data, while the business logic is implemented on the client side. However, the following considerations have to be taken into account:

1. In order to completely decouple client and server, a stateless API design needs to be implemented. That is, the server has no information on the clients' current status. The server only provides access to data represented as a group of independent resources.
2. The server is responsible for managing conflicts and maintaining consistency and integrity of the data. For instance, if two users edit the same feature simultaneously, a versioned system is required that allows for reverting and resolving conflicts.
3. Security of the data and managing how data is being used and what are the rules to access and view it should also be managed by the server. Creators of the data shall have full control on how the data can be used and exploited afterwards.
4.

Representational state transfer (REST) provides means to implement clean separation of client and server side. Capabilities to access and update data will be provided through a platform-independent API based on standard HTTP methods. That way, a variety of different applications can be implemented that assist users in contributing new data, manage existing data or provide means to visualise and analyse information.

3.2 Flexible data structure

Participatory mapping aims at helping people to map out a large variety of topics. While the mapped features will have common properties – such as geometry, version number or the date of last edit – the majority of properties that are to be stored for individual mapping projects cannot be foreseen at implementation time. A flexible tagging scheme similar to OpenStreetMap tagging, using key-value-pairs, is needed. That way, the amount and granularity of information attached to a geographic feature is completely up to the user (or, more likely in the case of participatory mapping, the developer of the front end applications). However, project administrators may want to ensure a certain level of data quality in terms of attribute completeness. Hence, the option to define a set of mandatory properties along with the required data type and the range of accepted values has to be provided by the system. Projects such as the ones mentioned in sections 2.1-2.3 demonstrated that some of the creativity in participatory mapping emerges from a group of participants deciding to use a common tag to describe their data, or create a local classification scheme that reflects local concerns.

Further, the system need to encourage rich annotations; e.g., recording observations, emotions and perceptions that help understanding characteristics of an object or a group of objects and facilitate discussion amongst contributors; for instance by enabling comments or attaching media objects (e.g. video note or audio note). Each object can be potentially linked to a discussion, which might mean a rich integration of information provided by a dedicated front-end application or by a third-party social networking system such as Twitter, Facebook, Instagram, or Ushahidi - information provided through these different routes will all be associated with a GeoKey object or objects in a way that will allow their retrieval.

3.3 Flexible user-rights management

A flexible set of options allows users and/or user groups to contribute, access and use the data of a project has been created. On one hand, administrators may select if contributed data should be entirely available to the public, restricted to certain users and user groups or entirely private. In the latter case, the data will be available for download through the administration interface only. Further, contributing data to a project or layer can be restricted in the same way. Additionally, access to data may be further constrained by the relation of user and object. If a user, who belongs to a certain group, creates an object, the same object might only be accessible and editable by other users of the same group. For example, in a project that focuses on gender-specific aspects, data created by women is then only visible to other women in the same project or, even more restricted, to a group of pre-selected participants. The initial preference has to be set by the administrator, but the decision as to whether the feature is available or not is entirely up to the user, who created the object in the database. It is recognised that the configuration of user management rights will require training and can be complex, and therefore special attention must be paid to the user interface and documentation that will support administrators, who might be workers at small Non-Governmental Organisations (NGOs).

In order to secure access to the API a robust user authentication and session management is required. A promising approach in this direction is the Oauth system, which applies secret tokens to grant an application access to an API.

3.4 Security

In addition to granting users different levels of access rights, encryption of security sensitive data can be important to some data sets. Therefore, a means to encrypt data using SSL/TLS or an alternative technology is envisaged.

Additionally, some information requires even more secure encryption than SSL/TLS, for example highly sensitive information relating to poaching incidents or recording observations of an endangered species should be only available to trusted users. In addition to encrypted data transfer, we further aim at encrypted data storage. Encrypted data includes all sensitive information such as location, the properties of the feature, information on the creator of the object and the trusted users. This will most likely require a second database as the information will be encrypted on client-side and therefore will not fit into the data model of the standard data base.

3.5 Quality control, review and moderation system

Participatory mapping encourages citizens with no formal qualifications in surveying or geographic information technologies to contribute and share data. In order to ensure a certain level of data quality, rigorous quality control is of the utmost importance in some cases (e.g. in the recording of land information). Collaboratively driven projects, such as OpenStreetMap or Wikipedia, have demonstrated that – despite the lack of formal qualification of contributors and the absence of formal quality control guidelines – the information available through such services is in general reliable. This is due to the large number of contributors who leverage the simple reviewing and editing tools these projects provide to improve the quality of initial contributions. In particular, in OpenStreetMap, users that are familiar with a certain area contribute their local knowledge to correct mapping errors and increase attribute accuracy on feature level.

For GeoKey, a set of tools that allow for reviewing and editing contributed data can be built on the platform. Users will be able to recommend a feature for review if they think accuracy has to be improved. Other registered users within a project are then informed about the feature in question and can decide if they want to fix any errors. Further, users will be able to comment on features and also on layers to enrich the data with personal opinions and perceptions at that location. Using the same system, a formal moderation process can be implemented for any contribution. In contrast, in some cases review and quality control of the data is not required. For instance, a community that makes up information (for a scavenger-hunt game, for instance) does not require quality control as the information cannot be verified in any way. Therefore, information down to the feature level is required to indicate whether the feature should be occasionally reviewed by other users.

3.6 Easy-to-use Administration

In order to enable even technical laypersons to create and administer new projects as well as to contribute data, a highly usable and easy-to use administration user interface is provided. This user interface helps people to set up and administer projects, including management of user groups and accessibility of data. Further, a query builder is envisioned, that allows for preparing database queries to create custom views of the data. The administration interface supports the design of data capture protocols, allowing data to be captured as text, numbers, multimedia files, and data streams from sensors. The administration interface is also designed to simplify the interaction with other systems, so that for each project it is possible to plug-in a suitable connectors to additional systems.

Additionally, the system can be extended to enable user authentication through different third-party services, such as OpenID, Google, Twitter or Facebook to simplify registration and authentication processes.

3.7 Supporting a wide range of analysis and visualisation

While the system should be mostly focus on back-end functionality, with data management capabilities, including user authentication and moderation, and the other functions described above, it should also facilitate the use of information by other applications and by its end users. Therefore, while it the system does not include visualisation capabilities, it stores parameters that can facilitate the visualisation and analysis across multiple platforms and devices. For example, the information about the icons to represent a class of information (e.g. monitoring site) and the colour that will be used for the representation of lines and polygons can serve a range of visualisation tools, and is therefore stored within the GeoKey database. The public API ensures that data can be extracted so it can be used in statistical analysis and visualisations using tools such as such as R. This can be facilitated by using standardised data formats.

3.8 Licensing

Licensing is a two-fold issue when implementing and running an open platform for participatory mapping. Firstly, GeoKey is released as an open-source product that can be forked and further developed by other parties and deployed and run on their own server. Therefore, GeoKey is released under an appropriate open-source software license that allows for redistribution as well as advancement and refinement of the code-base. However, in some cases the project will need to be integrated with commercial products. Therefore, the licensing framework that was selected is one that also allows for proprietary developments (Apache License version 2.0)

More critical is the choice of licensing of data contributed to the system. The creator might want to decide whether the data she contributed shall be public domain or is subject to further restrictions such as required attribution or share-alike distribution. The licenses provided by Open Data Commons might help to achieve the requirements. This will complicate the creation of the project within the system, as some data might be just out of bounds (e.g. the community deliberately don't want it to be shared outside the context of its original collection) while other pieces of information should be shared freely through systems such as OpenStreetMap. The system will therefore require a catalogue of data licensing frameworks that can be associated with different data layers and objects.

3.9 Accessing and Re-Using the Data - Automated Metadata

Metadata (data about data) is fundamental to ensuring the appropriate use and re-use of a dataset. It describes how a dataset was captured, which tools were used and for what purpose the data was originally designed. Understanding this facilitates a user's understanding of the limitations of the results of any subsequent analysis. Traditionally, metadata is captured to a standard such as the EU INSPIRE project's use of ISO-19115 and ISO-19119 (INSPIRE 2011), and requires extensive effort to create and maintain, as well as being complex to create and use (Poore and Woolf 2010, Rajabifard et al., 2009). Capturing extensive metadata is also not feasible in a Community Mapping project, where users will not have requisite expertise or time required for this task. As a proxy for this information, knowing who captured an item of data, and when, when the data was edited and by whom, and perhaps which device was used provides an opportunity to develop the concept of a 'trusted' user – i.e. a person or group of people who the community themselves knows will produce data that is 'fit for purpose'[1]. The 'purpose' for which a dataset is captured can be obtained from the project description.

3.10 Accessing and Reusing the Data - Search Functionality

Community mapping is not only about capturing data, but also about making use of the data to support decision making and advocacy within and beyond community groups. Thus, making the data accessible and re-usable is a fundamental component of any platform. GeoKey offers flexible search tools that mimic the approach taken by Google (and hence to which end users are accustomed) by allowing free text search. This is however combined with user access rights to ensure that users are only presented with information to which they have access rights. Users are also permitted to download their own data as a matter of course.

3.11 The Importance of Building a Community Identity

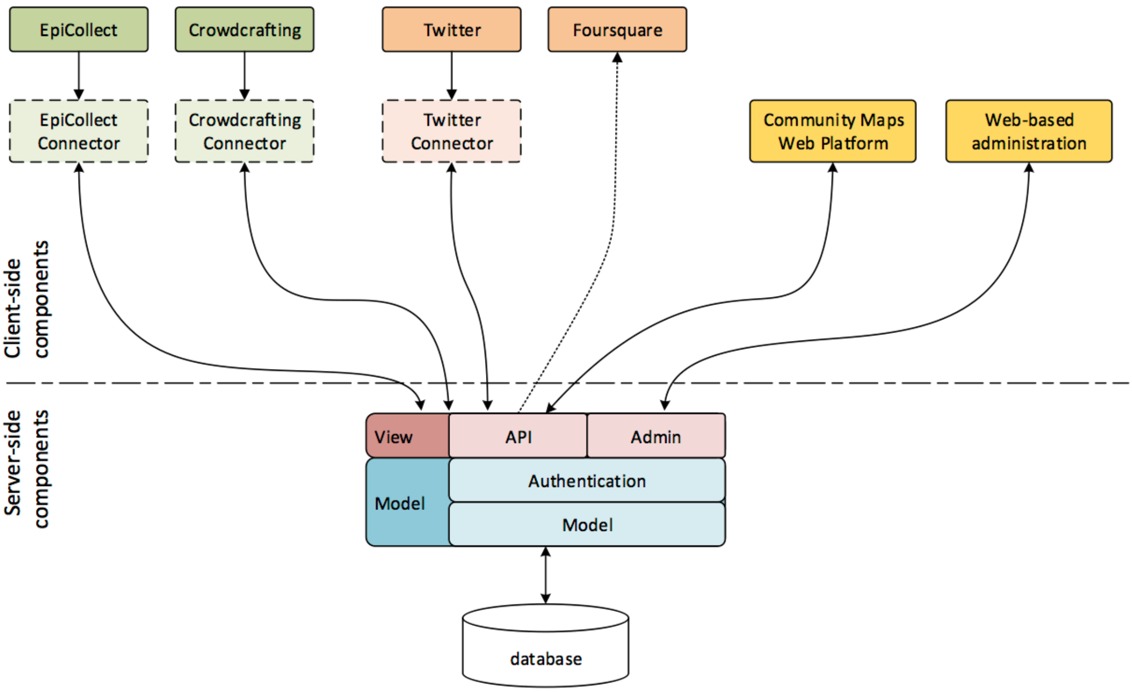
The combination of user rights, easy administration, search tools and quality control and moderation, and security permit a final aspect of GeoKey's functionality – the ability to allow front end developers to build a community, which is fundamental to the success of any project (Ellul et al 2009). The flexible data structure means that each community can define features for the map in their own way – for example, while many groups may wish to represent an entity called "local organisations", the information they would like to store about these will differ (and could include anything from name, e-mail address, contact details, opening hours, volunteering terms, funding sources, and organisation type). The GeoKey approach allows each group to define their own data structures, and group these together as projects, which is key to the bottom-

[1] The concept of 'fitness for purpose' is used when considering data quality and acknowledges that a dataset captured for one specific reason (e.g. to understand personal exposure to pollution) may not be suitable for another purpose (e.g. to generate a pollution map of a city)

up approach that drives community mapping. Similarly, custom user-rights can be configured as part of the administration set up process - some groups may require that all data is moderated, with nominated moderators, with others having automatic approval of data. The separation of front end and back end functionality allows developers to build their own custom, front-end tools and easily integrate these into existing project websites.

4. GENERAL SYSTEM ARCHITECTURE

To achieve the desired objectives, an extensible system architecture described below was developed for GeoKey. For a documentation of the current state of the underlying data model and the API, please refer to GeoKey github repository, (https://github.com/ExCiteS/geokey/).



<u>Figure 2.</u> System API architecture consisting of data storage, a core API and several connector modules.

The general system design (Figure 2) separates server-side components, including data storage and dissemination interfaces, from client applications, which access and alter the data through the interfaces. That way, we guarantee a flexible and platform-independent way for a variety of different applications to access the data. At the core of the back-end components is a relational database system with geospatial capabilities that keeps all information relevant to run the platform. That includes information on projects, the categories created by users, the contributions, comments and links to uploaded files. The database system needs to provide means for storing and querying geographic geometries and also include features of schema-less databases to accommodate for flexible store data in a flexible manner.

Business logic is implemented on top of the database. The model and authentication tier act as gateway between the database and API or web-based administration. The model describes the components of the data design and provides basic functionality for creating and managing the instances of the model components. The authentication tier authorizes the requesting user (either requesting through the API or through the web-based administration tool) and validates requests against the permissions granted through the user groups.

The web-based administration tool provides a user interface to set up and manage community mapping projects. This includes creating projects, their data structures consisting of categories and fields as well as setting up user groups and the respective permissions.

The API provides access to the system for third-party applications, such as EpiCollect which supports the creation of mobile forms for field data collection, Crowdcrafting which supports human computation and classification tasks and can be used to classify and verify information that was collected by participants, or social media, like Twitter. API functionality focusses on accessing, creating and updating and deleting data contributed to the platform (e.g., Features and Observations). It does not provide sophisticated management capabilities, such as creating and altering projects or views. This decision was made in order to keep the API as simple as possible and reduce security risks that are introduced when opening up a platform through an API.

A separate element of GeoKey implementation provides a generic API that allows for access and management of the data stored in the database and is implemented as a set of HTTP resources following the REST paradigm. HTTP methods GET, PUT and DELETE are employed to provide means to access and manage the data. The resources will be defined as a set hierarchical of URLs following a scheme. The listings of projects, layers and features only include data the user is allowed to view and access. The user has to provide authentication credentials or a session key with the request. Otherwise, only public information will be returned.

By default, information is exchanged using JSON, since it is a widely adopted and readable format to encode structured information with a small-size footprint. In addition to JSON, other formats, such as XML and CSV are supported, to allow integration with GIS.

On top of this core API, connectors are implemented that serve as interaction point between third-party applications (such as Twitter, Facebook, Instagram etc.) or specialised applications, for instance, apps on mobile phones such as EpiCollect. The connector modules employ the HTTP resources to access and store data in the system on one end and to retrieve information from the connected service on the other end. However, some projects may rely on data that is not supposed be integrated into the platform or on an integration that is not feasible. In that case, the system may provide a pointer or external link to another data source that can be utilised when necessary. For instance, some social media sites do provide access to the data through their own APIs but prohibit storage in external databases, such as the case of Foursquare. This might be also the case where a certain observation is linked to a feature in OpenStreetMap's database. In this context, the notion of linked open data provides a promising approach to implement these links. An observation contributed to a participatory mapping project can be directly linked to a Foursquare venue or OSM feature. This is accomplished by treating both the observation and the Foursquare venue as a web resource, which is identifiable through an individual URL, and a link between both resources encoded as a RDF tuple.

In addition, GeoKey provide means for importing data or linking to external systems, such as ArcGIS Online. Another possibility is the use of citizen science platforms such as CrowdCrafting to carry out a data classification or analysis task on data that was collected on the main platform. Further, on top of the core API client applications can be implemented that focus on visualisation and analysis of the data, such as Mapping for Change Community Maps.

5. DISCUSSION AND CONCLUSION

Within the framework of participatory mapping, the process of learning is done through participation in the activity and sharing information with other participants, as well as through participation in the analysis of the information. For example, Figure 3 shows the feedback session at the end of community-led air quality monitoring in London. While, the full case study is described elsewhere (Citizen Cyberlab 2015), of relevance here is the fact that the community mapping system that was used to share the information with participants is built using GeoKey. GeoKey has been used to manage all the information that was collected by a whole range of communities across London (see https://communitymaps.org.uk/#/project/26), including the group that has participated in the meeting.

As participatory mapping aims to be an inclusive practice, it is important to ensure the engagement of participants with varying levels of technical capabilities. The types of learning that can be expected from users of GeoKey include learning about the use of applications such as EpiCollect, learning about the process of data collection in a systematic way, learning about the use of the community mapping systems and the development of spatial literacy regarding the area in which the observations were taken. This is coupled with knowledge about sharing the information with other people. As Figure 3 vividly demonstrates, participants are animated and focus on the information that is presented to them and are able to link the information that came from the scientific sensing of their environment with their own knowledge of the area, perceptions about traffic and other sources of pollution and their understanding of the mapping.

The paper maps, which are based on information that was exported from GeoKey, provide another form of visualisation and discussion of the information.

The opportunities to annotate information that was recorded by sensors, as well as make decisions related to where to position and use them, are ways in which the practice of participatory mapping invites any participants to explore what they are interested in and what are the research questions that they would like to explore. Although GeoKey primary aim is to support participatory and community mapping, the practices within which it is embedded and used supports creativity and learning in informal settings, and the platform provides the flexibility to support learning activities in many ways. There is also an opportunity to link GeoKey to learning platforms through the creation of connectors which will share information that was collected on GeoKey on systems such as virtual learning environments (VLE).

In summary, we have described in this paper both the rationale and the architecture for an infrastructure for a participatory mapping and science platform, and demonstrated how the objectives can be implemented in a modern, three tier architecture that supports both qualitative and quantitative information. The development of GeoKey as an Open Source platform which was open from the start is aimed to allow it to be suitable to a wide range of applications and contexts. The future directions of GeoKey development include a comprehensive implementation of the functionality that was described above, as well as a dedicated effort to increase the range of applications in which it is used and increase the developer community that use the system. To achieve this we need to ensure that the system is well documented and easy to use.

Figure 3. Participants use Community Maps (at the back) and paper maps to understand air quality in their area

6. REFERENCES

Becker, M., Caminiti, S., Fiorella, D., Francis, L., Gravino, P., Haklay, M., Hotho, A., Loreto, V., Mueller, J., Ricchiuti, F., & Servedio, V.D (2013). Awareness and learning in participatory noise sensing. PLoS ONE 8(12): e81638. doi:10.1371/journal.pone.0081638

Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., & Parrish, J. K. (2014). Next steps for citizen science.*Science*, *343*(6178), 1436-1437.

Chambers, R. (2006). Participatory mapping and geographic information systems: whose map? Who is empowered and who disempowered? Who gains and who loses?. *The Electronic Journal of Information Systems in Developing Countries*, *25*. http://www.ejisdc.org/ojs2/index.php/ejisdc/article/view/238

Ellul, C., Haklay, M., Francis, L., & Rahemtulla, H. (2009). A mechanism to create community maps for non-technical users. In *Advanced Geographic Information Systems & Web Services, 2009. GEOWS'09. International Conference on* (pp. 129-134). IEEE.

Ellul, C., Francis, L., & Haklay, M. (2011a). A Flexible database-centric platform for citizen science data capture. In e-Science Workshops (eScienceW), 2011 IEEE Seventh International Conference on (pp. 39-44). IEEE.

Ellul, C., Francis, L., & Haklay, M. (2011b). Engaging with local communities: a review of three years of community mapping. Urban and Regional Data Management, UDMS Annual 2011 - Proceedings of the Urban Data Management Society Symposium 2011 pp. 165-177.

Esri (2015) Community Mapping Program http://www.esri.com/landing-pages/community-maps

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), 211-221.

Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. Pervasive Computing, IEEE, 7(4), 12-18.

Haklay, M., Singleton, A., & Parker, C. (2008). Web mapping 2.0: The neogeography of the GeoWeb. Geography Compass, 2(6), 2011-2039.

INSPIRE (2011). INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 [online] Available from: http://inspire.jrc.ec.europa.eu/documents/Metadata/INSPIRE_MD_IR_and_ISO_v1_2_20100616.pdf, [Accessed 23rd March 2011].

Lewis, J., Freeman, L. and Borreill S., Free, Prior and Informed Consent and Sustainable Forest Management in the Congo Basin, SECO available at http://www.rightsandresources.org/documents/files/doc_841.pdf

Maisonneuve, N., Stevens, M., Niessen, M. E., & Steels, L. (2009). NoiseTube: Measuring and mapping noise pollution with mobile phones. In Information Technologies in Environmental Engineering (pp. 215-228). Springer Berlin Heidelberg.

Perkins, C. (2007). Community mapping. The Cartographic Journal, 44(2), 127-137.

Poore, B., and Wolf, E., (2010), The Metadata Crisis – Can geographic information be made more usable? U.S. Geological Survey Proceedings of the Second Workshop on Usability of Geographic Information, London, March 23, 2010. Available from: http://www.virart.nottingham.ac.uk/GI%20Usability/Workshop%20papers% 20pdfs/ [Accessed March 1, 2011].

Rajabifard, A., Kalantari,M., and Binns,A., (2009), SDI and Metadata Entry and Updating Tools" in B. van Loenen, J.W.J. Besemer and J.A. Zevenbergen (Eds.) SDI convergence: Research, Emerging Trends and Critical Assessment, Netherlands Geodetic Commission, Delft.

Regalado, C., Haklay, M., & Francis, L. (2015) *D4.6 Creative Learning in Extreme Citizen Science pilot activities*, Citizen Cyberlab project, London.

Sieber, R. (2006). Public participation geographic information systems: A literature review and framework. Annals of the Association of American Geographers, 96(3), 491-507.

Starbird, K. (2011, May). Digital volunteerism during disaster: Crowdsourcing information processing. In Conference on Human Factors in Computing Systems (pp. 1-4).

Stevens, M., Vitos, M., Lewis, J., & Haklay, M. (2013). Participatory monitoring of poaching in the Congo basin. GISRUK 2013, 21st GIS Research UK conference, April 2-5, 2013, Liverpool, England, UK.

Tudge, P., Sieber, R., Wiersma, Y., Corbett, J., Chung, S., Allen, P., & Robinson, P. J. (2012). Collaborating towards innovation: Lessons from the participatory GeoWeb GEOIDE network. The Added Value of Scientific Networking: Perspectives from the GEOIDE Network Members 1998–2012, 103-118.

van Wart, S., Tsai, K.J. & Parikh, T.S (2010). Local Ground: A Paper-Based Toolkit for Documenting Local Geo-spatial Knowledge. ACM Symposium on Computing for Development (DEV), December 17-18, 2010, London, UK

Verplanke, J., McCall, M. K., Uberhuaga, C., Rambaldi, G., & Haklay, M. (2016). A Shared Perspective for PGIS and VGI. *The Cartographic Journal*, 53(4), 308-317.