# FIDIS

Future of Identity in the Information Society

| | |
|---|---|
| Title: | "D12.10: Normality Mining: Results from a Tracking Study" |
| Author: | WP12 |
| Editor: | Mark Gasson (University of Reading, UK) |
| Reviewer: | Hans Hedbom (KU, Sweden) |
| Identifier: | D12.10 |
| Type: | [Other] |
| Version: | 1.0 |
| Date: | Tuesday, 30 June 2009 |
| Status: | [Final] |
| Class: | [Public] |
| File: | FIDIS_D12.10_v1.0.doc |


***Summary***

Within FIDIS, WP3 and WP12 have dealt with RFID, WP11 has investigated mobility and identity while WP6 has examined biometrics and WP7 profiling. The aim of this report is to bring these disparate threads together into a tangible study which will demonstrate privacy issues surrounding products and services which are likely to start emerging on to the consumer market.

New generations of mobile handsets, with integrated devices like GPS and internet capabilities, are becoming less like traditional phones. In fact we should stop viewing them as simply mobile phones - they are now more like mobile computers which can make phone calls. These advances in mobile technologies will inevitably lead to new services which we can enjoy anywhere, anytime. Location Based Services which utilise the phone's GPS to tell us for example where we are, or where the nearest cinema is, are an obvious first step – but what happens if the phone monitors where we go at all times? Can these new services build a picture of who we are based on where we have been? Can they use this profile of us to understand what we like and tailor their results specifically to us? And if so, at what cost to our privacy? In this report, aimed at the potential consumers of such services, we will look at results from a recent tracking study which examines these issues.




*The FIDIS NoE receives research funding from the Community's Sixth Framework Program*

# Copyright Notice:

# Members of the FIDIS consortium

| | |
|---|---|
| ***1. Goethe University Frankfurt*** | Germany |
| ***2. Joint Research Centre (JRC)*** | Spain |
| ***3. Vrije Universiteit Brussel*** | Belgium |
| ***4. Unabhängiges Landeszentrum für Datenschutz (ICPP)*** | Germany |
| ***5. Institut Europeen D'Administration Des Affaires (INSEAD)*** | France |
| ***6. University of Reading*** | United Kingdom |
| ***7. Katholieke Universiteit Leuven*** | Belgium |
| ***8. Tilburg University***[1] | Netherlands |
| ***9. Karlstads University*** | Sweden |
| ***10. Technische Universität Berlin*** | Germany |
| ***11. Technische Universität Dresden*** | Germany |
| ***12. Albert-Ludwig-University Freiburg*** | Germany |
| ***13. Masarykova universita v Brne (MU)*** | Czech Republic |
| ***14. VaF Bratislava*** | Slovakia |
| ***15. London School of Economics and Political Science (LSE)*** | United Kingdom |
| ***16. Budapest University of Technology and Economics (ISTRI)*** | Hungary |
| ***17. IBM Research GmbH*** | Switzerland |
| ***18. Centre Technique de la Gendarmerie Nationale (CTGN)*** | France |
| ***19. Netherlands Forensic Institute (NFI)***[2] | Netherlands |
| ***20. Virtual Identity and Privacy Research Center (VIP)***[3] | Switzerland |
| ***21. Europäisches Microsoft Innovations Center GmbH (EMIC)*** | Germany |
| ***22. Institute of Communication and Computer Systems (ICCS)*** | Greece |
| ***23. AXSionics AG*** | Switzerland |
| ***24. SIRRIX AG Security Technologies*** | Germany |

[1] Legal name: Stichting Katholieke Universiteit Brabant
[2] Legal name: Ministerie Van Justitie
[3] Legal name: Berner Fachhochschule

# Versions

| *Version* | *Date* | *Description (Editor)* |
|---|---|---|
| **0.1** | 01.06.2009 | • First draft structure |
| **0.2** | 28.06.2009 | • Integration of legal aspects |
| **0.3** | 29.06.2009 | • First review release |
| **1.0** | 30.06.2009 | • Internally reviewed release |

# Foreword

FIDIS partners from various disciplines have contributed as authors to this document. The following list names the main contributors for the chapters of this document:

| *Chapter* | *Contributor(s)* |
|---|---|
| **Chapter 5** | Eleni Kosta (KU Leuven) |
| **All Others** | Deliverable Editor |
| | |

# Table of Contents

# 1 Executive Summary

Developments in mobile phone devices are rapidly reforming our relationship with technology. The changes are not just technological - they are driving changes in cultural and social paradigms, and further empowering the consumer to seek new experiences and employ new services for entertainment and convenience, among others. The drive from industry to stay at the cutting edge of the technology has seen the humble mobile phone turn into a feature packed computing device within a decade. With Internet capabilities, high resolution cameras, GPS and growing selections of third party software applications available, we should stop viewing these devices as simply mobile phones - they are now more like mobile computers on which we can make phone calls. Indeed it seems that mobile handsets are the first wave of successful 'wearable computers' prophesised for years by scientists, at least in the sense that they comprise a relatively powerful computing device which people habitually carry with them. Such devices will almost certainly retrospectively be viewed as the forerunner to 'ubiquitous computing' and 'Ambient Intelligent Environment', other paradigm shifts predicted in our evolving relationship with technology. As it stands, new generations of handset are heralding a new era of information access and disclosure. In this report we are focussed on the ability for people to reveal where they are at any time in the name of safety, convenience or for social use. The concern here is that this begins to vastly ebb the divide between safety, convenience or entertainment and the invasion of personal privacy.

Throughout April 2009, four people from three different European member states were persistently tracked via GPS enabled mobile phones and their location data stored in a central database for automated and manual processing. The aim of this processing was a first attempt to mine new information from the data relevant to forming behavioural profiling of the individuals based on where they had been. It is thought that services offering customised information based on the results of such profiling will become commonplace in the very near future, and so this is an opportunity to assess the possibilities and potential risks of allowing the processing of location data. This is particularly important since from the service offered it may not immediately be apparent to the user that a wealth of information about them, potentially unrelated to the service, can be revealed. Further issues occur if the user agreed while subscribing to the service for the data to be passed to third parties where it may be used to their detriment during the provision of other services.

While it is evident that the month long period undertaken for this pilot study is not enough to draw substantial conclusions, and thus elaborate profiles of the individuals, the analysis of the user data that was conducted in the frame of this tracking study allows us to make some very interesting remarks. Very important, especially from a legal point of view, is the realisation that location data in several instances, while being useful for drawing behavioural profiles, can potentially reveal information relating to the health life of the individual, his political or religious beliefs etc. However, if the collection and processing of location data reveals information that falls under the category of sensitive data, then probably the whole legal and business model around it will need to change. A lot of questions arise, as a result of our user data analysis, with regard to location data. While the objective of this study is not to answer these questions, what is evident is that at the very least the users of these devices must be made aware of the privacy risks of disclosing their location data. This is especially when it is persistently collected, but they should in any case make sure that they are aware of how this potential wealth of information is being further exploited by those offering the services they are using.

# 2 Introduction

Advances in mobile technologies have meant that being able to track or locate people has been possible for some time, although the information is usually only readily available to mobile phone operators. More recently, the advent of data enabled mobile phones, and the emergence of popular social networking internet sites has realised a dramatic increase in the volume of information people willingly disclose about themselves, in many cases to large numbers of complete strangers.

In February 2009, Nokia forecast that 50% of its handsets sold this year will include a GPS unit[4], while the 3G iPhone, with integrated GPS, supposedly held a 4-6% share of the handset market in the UK toward the end of 2008. The miniaturisation of GPS devices and their inevitable inclusion in mobile handsets has generated a new era of information disclosure, and new services are likely to appear which encourage people to reveal where they are at any time in the name of safety, convenience or for social use. The danger here is that this begins to vastly ebb the divide between safety, convenience or entertainment and the invasion of personal privacy.

The aim of D12.10 is to harness elements from various WPs within FIDIS in order to demonstrate that the seemingly harmless data does not just reveal such things as where you have been – it exposes aspects of your private life that at first glance you may not realise. It is not just where you live, where you work or when you go for a coffee, it is possible to aggregate all these little pieces of information and use data mining techniques to extract a 'behavioural profile' from the data. The problems could come if for example this information is used by third parties to vary their services, prices or inclusion specific to the individual, and in some cases to their detriment. This is especially given that the end 'user' may have no real understanding that this is happening, or indeed how it is happening because they do not understand the mechanisms behind the service they have opted to use.

We have implemented a system which allows a mobile device to be globally tracked in real-time by having it access GPS and cell tower information which are sent for processing in a central database. The data recorded from this system (in-line with an agreed privacy policy) has been analysed to demonstrate how (rough) profiles can be drawn from such data. For a period of a month, four people have been tracked using this system to collate an amount of data suitable for an initial appraisal.

Further to this, we want to demonstrate how technology can be used to periodically authenticate a user to a device such that we can ensure that the data being collected is from the user we believe it is. The two approaches are active through biometric authentication and passive through implanted RFID tags. The associated data confidence using these techniques is also discussed.

[4] http://uk.biz.yahoo.com/04022009/323/update-2-asustek-garmin-join-gps-phone-market-foray.html

# 3 Emerging LBS technologies vs. Ambient Intelligence

Technology exists as a means to further empower people, a result that is best achieved by constructing a close synergy between man and machine. The rapid development of technology has led to new fields of research dedicated to developing new and intuitive methods by which humans can interact with machines. Essentially, the problem is no longer just one of how technology can make a task easier for us, but in addition how we can interact with machines to benefit from their functionality to the greatest extent. Basically a, the interface through which a user must interact with the machine provides a distinct layer of separation between what the user wants the machine to do, and what it actually does. Ambient Intelligence Environments (AmI) have been presented for many years as the panacea for the human / technology interaction bottleneck. The very essence of AmI is to enrich the user experience by capitalising on the potential that additional computing processing can bring. Part of this enrichment is achieved by augmenting the user in their daily lives through additional services and access to additional information. However, this is achieved whilst actually reducing the focus on the traditional explicit data input / output paradigm - a true shift in our concept of what a computer is, and how we should interact and use it.

"*It seems like a paradox but it will soon become reality: The rate at which computers disappear will be matched by the rate at which information technology will increasingly permeate our environment and our lives*"[5].

AmI has been a well researched topic in wider academia and in the FIDIS project (for example see D12.2 and D7.3) for some years. This vision is thought by many to be an accurate depiction of the future, although no-one can say for sure when this vision may be realised. What has become clear is that this does indeed seem to be a paradox since we are in fact moving away from this ideal of 'disappearing' computers. Some 600 million mobile phone handsets sold in 2004 alone, and new generations of mobile handsets, with integrated devices like GPS and internet capabilities, are becoming increasingly popular and less like traditional phones. In fact we should stop viewing them as simply mobile phones - they are now more like mobile computers on which we can make phone calls. These devices are akin to the types of Wearable Computer devices which have also been discussed and developed by scientists over the past few years, but which to date have not found any satisfactory consumer implementation. It would seem that if indeed AmI is a realistic vision, then 'wearable computers' in the form of sophisticated mobile devices may be the first wave. Interestingly, within the FIDIS project, the concept of an AmI environment in which the user had more control through the concept of 'mobile sensors' was proposed, and the coming generations of sensor augmented mobile (phone) devices seems to suggest that this is a viable prospect.

In the short term, advances in mobile technologies will inevitably lead to new services which we can enjoy anywhere, anytime. Location Based Services which utilise the phone's GPS to tell us for example where we are, or where the nearest cinema is, are an obvious first step – but what happens if the phone monitors where we go at all times? Can these new services

[5] Streitz & Nixon (2005), 'The Disappearing Computer', Communications of the ACM, Vol. 48 (3), March 2005. pp. 33-35.

build a picture of who we are based on where we have been? Can they use this profile of us to understand what we like and tailor their results specifically to us? And if so, at what cost to our privacy?

The wider issues associated with LBS and mobile identity have been explored within WP11 of FIDIS. The work here builds largely on the report D11.12 of which parts are summarised below, however it is not the intention to reiterate this material here and so the interested reader is further directed to the aforementioned documents.

## 3.1 Normality mining

Collating data and inferring new information, or knowledge, from it through the application of data mining techniques is nothing new. Database marketing for example makes use of data mining and (new) knowledge discovery to develop models of customer behaviour. These models generally abstract from the individual, but try to classify customers and products in classes and identify rules for behaviour. The rules are used to select customers to be addressed and products to address customers with. Such data mining and knowledge discovery techniques need as much data as possible about customers to increase the probability that the model developed fits the needs. The concept of 'normality mining' is drawn from the fact that *everything* we do in our normal, even seemingly benign lives, is of some interest, and value, to someone, somewhere. The typical barriers to accessing this information are technical ones – if the data cannot be collected, then it cannot be exploited. However, today's information society is making enormous headway in the ability to collect data from disparate sources. The data may, for example, be collected from details of the transaction history with one's own customers or bought from third party companies that have captured the information. Typical sources are charity donation forms, application forms for free products or contests, product warranty cards, subscription forms, customer loyalty programs, and credit application forms. Data of interest for customer profiles can be manifold, starting from name and address, history of shop searches and purchases, demographics, and the history of past communications to and from customers.

LBS offer a new dimension to classic database marketing by disclosing the physical location of the user (or rather the handset). However, data available to providers of mobile or geographic services reveals even more about their users' personal life than data collected by classical services. While this enables services to provide an enhanced service or experience, such as route guidance, tourist and weather information, more marketing activities which explicitly address people's physical places become possible. This becomes even truer for ubiquitous computing where users do not interact with the computing infrastructure explicitly and may not even be aware of marketing activities.

### 3.1.1 Profiling systems for LBS

As stated in the introduction, in February 2009, Nokia forecast that 50% of its handsets sold this year will include a GPS unit[6], while the 3G iPhone, with integrated GPS, supposedly held a 4-6% share of the handset market in the UK toward the end of 2008. Further, a research

[6] http://uk.biz.yahoo.com/04022009/323/update-2-asustek-garmin-join-gps-phone-market-foray.html

report from Berg Insight predicts that more than 960 million mobile handsets sold in 2014 will have integrated GPS receivers. The miniaturisation of GPS devices and their inevitable inclusion in mobile handsets has generated a new era of information disclosure, and new services are likely to appear which encourage people to reveal where they are at any time in the name of safety, convenience or for social use. While these services will undoubtedly be useful, the seemingly harmless data may not just reveal such things as where you are or have been – persistent collection of data can expose aspects of your private life that at first glance you may not realise. It is not just where you live, where you work or when you go for a coffee, it is possible to aggregate all these little pieces of information and use data mining techniques to extract a 'behavioural profile' from the data. The problems could come if e.g. this information is used by third parties to vary their services, prices or inclusion specific to the individual, and in some cases to their detriment. This is especially given the end 'user' may have no idea that this is happening, or indeed how it is happening.

As an example application, let's consider 'Recommender systems'. These are designed to help the user find what he is actually looking for. The idea of recommender systems is not new and a lot of systems are well known in academia as well as in practice. Some websites like Amazon.com, CDNow.com, Barnes & Nobel, MovieFinder.com, Pandora.com, TiVo.com, Netlix.com or Launch.com have made successful use of such systems. The analysis of user behaviour is accomplished by classical research separated into the active way which asks the user explicitly about his behaviour and the passive way where the behaviour is derived by indirect information collection and interpretation. With new generations of mobile handsets it is possible to download new software and services to the phone (e.g. the 'apps' from the iPhone store) which can enhance the user experience. Recommender services exist in this context already, with most allowing the user to rate the suggested content to allow the service to learn the user's preferences. This is a type of active engagement. Now consider the deployment of a recommender system which attempts to passively build a profile, to take the burden away from the user. New generations of handsets are ideal for this purpose because not only do they have the technological capability to capture and send data, but they are generally carried by the individual everywhere they go, practically at all times. This means GPS enabled devices can monitor where the user goes, data mining can be applied to find what routines they have and where they tend to go and profiles can be drawn from the data. In the case of monitoring multiple users, social and business networks can also be inferred. The advantage to the user is that, for example, when travelling to a new city, the user can ask the service individual specific, yet abstract questions: 'where are restaurants that I would like?' or 'where do people like me go to here?' and so on. The downside is that the data used to create these profiles will inevitably reveal much more about the user.

To further explore this emerging technology, and to assess the potential impact on the privacy of user in light of the European Data Protection Legislation, we have implemented a study tracking four people using GPS for a period of one month to examine the type of information we are able to glean. The aim is to investigate in what ways the data we can collect over this relatively short period supports our hypotheses that it is adequate to draw a simple profile and that the data is highly privacy invasive. This is detailed in the following Chapters.

# 4 Initial Results from a GPS Tracking System

The data collected during this study represents a very valuable resource which can be exploited in a number of ways. Within the context of this study, the analysis has been kept to processing the data to examine issues surrounding potential privacy problems. As such, the process of analysis has been simplified in the first instance to focus on this area, while more comprehensive processing based on the initial findings is underway for more academic publication at a later date. In this chapter the basics of the system are described first to familiarise the reader with the types of data collected.

## *4.1 Data collection*

The design of the system was the result of multiple iterations based on conflicting factors such as usability, cost, availability and functionality. This involved process is not detailed here – instead we briefly describe the resulting implementation. Each user was issued with a HTC 6500 handset with integrated GPS, fingerprint reader and full size SD slots for additional peripheral devices. Running Windows Mobile 6.1 (after upgrade), this unit provides an ideal development platform for this study, simplifying the hardware solution into a mainly integrated device which is less prone to user damage. Two pieces of custom software are run on all devices – the first logged each time the user authenticated with the device (typically through the fingerprint reader). The second polled the internal GPS unit every 10 seconds for updated NMEA[7] data which includes, among other data, GPS co-ordinates. Based on these the software calculated the distance travelled and if greater than 100m, it collected the cell tower details from the Windows Mobile Radio Interface layer and the last time of authentication from the log file mentioned above. The cell tower information only gives details about the id of the cell tower, so the software used the opencellid.org service to retrieve the GPS co-ordinates of the cell tower (if they are available). All of this data was then sent for central storage as detailed below.

### 4.1.1 The database and processing platform

The data is sent automatically from the devices to a MySQL database where it is securely stored. The data format takes the form below:

| Field | | Type | | Example data |
|---|---|---|---|---|
| Row | | bigint(20) | | 32974 |
| Latitude | | double | | 50.86544374 |
| Longitude | | double | | 4.65134145 |
| TowerId | | int(11) | | 41766 |

[7] See for more information e.g. http://www.gpsinformation.org/dale/nmea.htm

| | | | | |
|---|---|---|---|---|
| MobileCountryCode | | int(11) | | 206 |
| MobileNetworkCode | | int(11) | | 20 |
| LocationAreaCode | | int(11) | | 303 |
| cellLatitude | | double | | 50.872253 |
| cellLongitude | | double | | 4.660578 |
| IDTime | | varchar(14) | | 1241033658 |
| ID | | varchar(11) | | ‘unique ID’ |
| Time | | varchar(14) | | 1241035074 |
| Realtime | | varchar(40) | | Wed, Apr 29, 2009, 19:57:54 GMT |
| POI | | tinyint(1) | | 0 |
| Use | | tinyint(1) | | 1 |

‘Row’ is simply the incremental row identifier, ‘IDTime’ is the timestamp of the last authentication, ‘ID’ is the unique user identifier, ‘Time’ refers to the time the data was sent in Unix timestamp form, and ‘Realtime’ is a human readable version of that. ‘Use’ specifies whether a data point should be used in the data processing or not, and is designed for removal of spurious data points (which can occur if the GPS signal is very low).

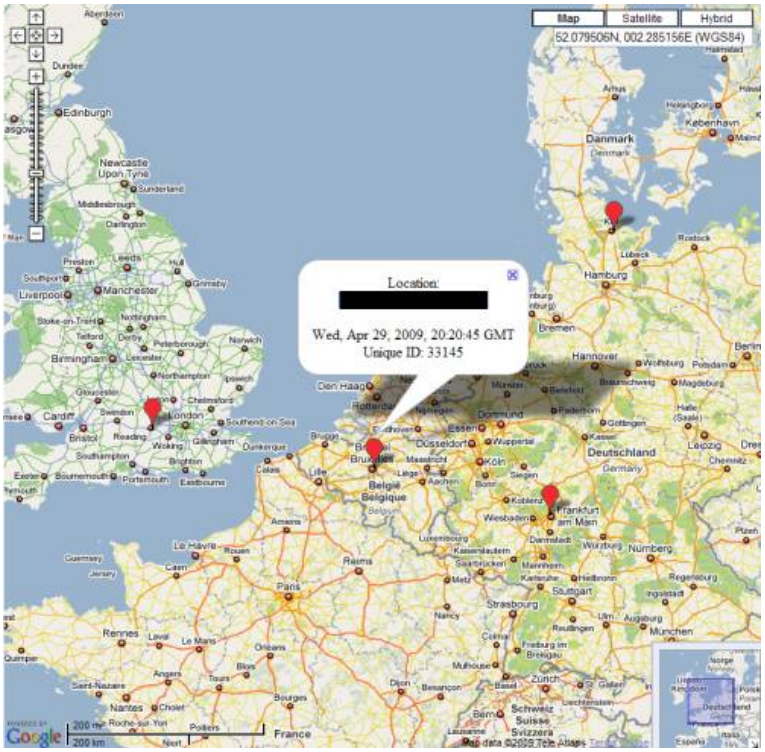



**Figure 1: Result of an algorithm finding the last known location of each user (shown as red dots), displayed via the Google Map API with additional data information shown (location data removed)**


**Figure 2: Display of one user's activity over a 24 hour period**


**Figure 3: A plot of the actual points a location has been logged (each indicated with a blue flag). Note the 'Toggle Status' link in the information panel**

The front end to the database is a bespoke PHP enabled viewer which allows user definable algorithms to be run on the data, the results of which are then displayed using the Google Map API. This allows simple functions such as displaying the last known location of every user (see Figure 1), or the movement activity of any user on any day (see Figure 2). Equally the actual location points can be plotted, with an option to toggle their 'use' indicator on and off (to remove spurious data points if need be (Figure 3)).

## 4.2 User profiling from GPS data

It should be noted that the purpose of this study is not to implement a complex and fully automated system which comprehensively profiles the participants. Instead the aim is to examine the kinds of privacy invasive information we can draw from data collected over a relatively short period, and so in what ways such systems could be privacy invasive if the user of an LBS agreed to persistent tracking. As such, during the processing phase the interactive interface was used to both run various automated tasks and enable manual inspection of the data.

One of the key tools for initial mining is the generation of 'Points of Interest' (PoI) from the data (see Figure 4). This uses a technique common in the literature of calculating the time between subsequent location points, and concluding that a PoI exists if the time is over a preset threshold. In our case a threshold of 3 minutes was deemed adequate.

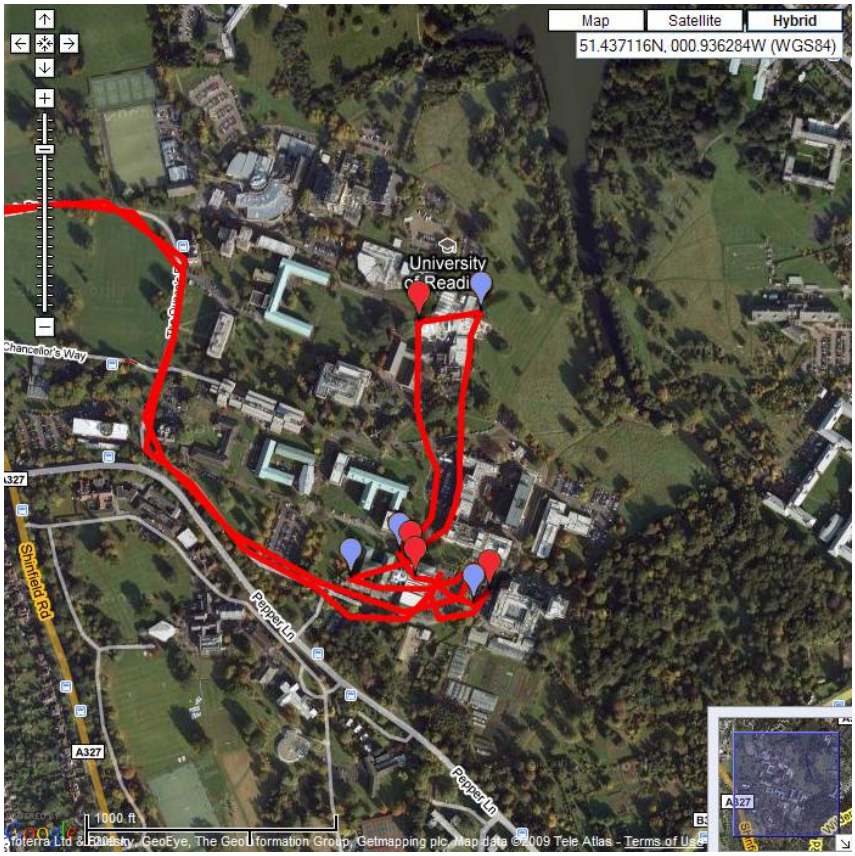



**Figure 4: Indicators marking automatically generated 'Points of Interest' with optional satelite overlay on the map**

Essentially this determines points between which the user spent some time. Figure 4 displays these as a red flag for the point at which the data stopped updating, and a blue flag for the point at which it continued. This could in fact be the entrance and exit to a building, which may not be the same point – e.g. see the upper most red/blue flag pair in the figure.

The next phase of the process is to determine the location identified as being of interest. It is expected that in a fully fledged profiling system this would be an automated process utilising multiple online resources such as the map shown in Figure 5.

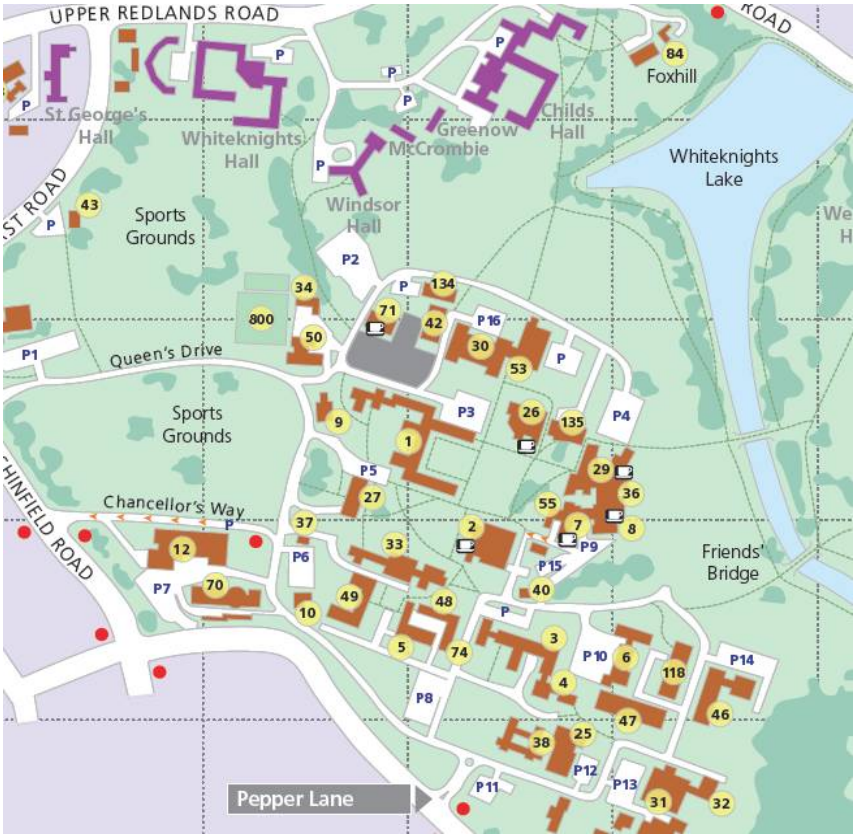



**Figure 5: Online resource detailing the points to be identified**

From this resource, and the times the points occurred, we can infer that this user arrives at work (Systems Engineering at the University of Reading) at around 09:00, takes lunch in the university cafeteria from 11:45 until 12.20, then presumably takes a drink at the cafe in the Business centre before returning to work. This person then leaves work late at 21.45. In this study the identification of PoIs has been largely achieved manually, and in some cases by survey of the users in order to expedite the process.

# 5 Data protection and location data

The study is based on the collection of GPS/location data of the participants. As all the participants are European citizens and the collection and processing of the data takes place in Europe, we need to examine how the European legislation on data protection applies. For the needs of this study, we are going to focus on the European legal framework and not on country-specific legislations.

## *5.1 The European Directives applicable to location data*

Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data (hereinafter called 'Data Protection Directive')[8] pursues two closely linked objectives: to lay down specific rights of the individual on his personal data but also to ensure that such data can move freely within the single market created between the Member States of the EU. This Directive contains the basic provisions with regard to the data protection principles that apply on the processing of personal data, defines the rights of the data subject and the obligations of the data controllers.

Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (hereinafter called 'ePrivacy Directive')[9] contains specific provisions on the processing of personal data in the electronic communications sector, regardless of the medium used. Among others, the ePrivacy Directive contains detailed provisions on the processing of location data in the context of value added services[10].

In 2006 a new Directive 2006/24/EC was adopted, which deals with the retention of specific types of traffic and location data, as well as identification data, for law enforcement purposes (hereinafter 'Data Retention Directive')[11]. Every Member State can choose to retain these data for periods of not less than six months and not more than two years from the date of the communication[12].

[8] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, O.J. L 281/31, 23 November 1995

[9] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), O.J. L 201/37, 31 July 2002. The ePrivacy Directive replaced Directive 97/66/EC of the European Parliament and the Council of 15 December 1997 on the processing of personal data and the protection of privacy in the telecommunications sector, O.J. L 53, 14 January 1998

[10] Article 2(g) ePrivacy Directive defines a value added service as "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"

[11] Directive 2006/24/EC of the European Parliament and of the Council on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC, Official Journal L105, 54, 15 March 2006.

[12] Article 6 Data Retention Directive

## 5.2 Personal data

Article 2(a) of the Data Protection Directive defines "personal data" as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his/her physical, physiological, mental, economic, cultural or social identity".

According to the Article 29 Data Protection Working Party[13], the definition of personal data contains four main building blocks[14]:

a. any information
b. relating to
c. an identified or identifiable
d. natural person

In the context of this tracking study, which is based on the processing of location information of the users, the third element on the identifiability of the user (identified or identifiable) is of great importance and will be further elaborated below. Admitting that the concept of "identifiability" plays an important role for the legal status of all not fully (or not immediately) identifiable data, the Article 29 Data Protection Working Party provided some further clarification in its Opinion 4/2007 on the concept of personal data[15], section 3.

In the case of this study, where the data controller[16] is based in the UK, understanding the UK's interpretation of the directive is essential to ensure the correct and appropriate system implementation. In the UK, the Information Commissioner's Office (ICO), responsible for promoting public access to official information and protecting personal information, issued a Technical Guidance aimed to help the data protection practitioners in deciding whether specific information falls within the definition of personal data. The Data Protection Technical Guidance contains an eight point check list to enable the quick assessment of whether particular data should be considered as personal data[17]:

[13] Under Article 29 of the Data Protection Directive, a Working Party on the Protection of Individuals with regard to the Processing of Personal Data is established, made up of the Data Protection Commissioners from the Member States together with a representative of the European Commission. The Working Party is independent and acts in an advisory capacity. The Working Party seeks to harmonize the application of data protection rules throughout the EU, and publishes opinions and recommendations on various data protection topics.
[14] Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data, WP 136, 20 June 2007, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf
[15] Idem
[16] A data controller is the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data (Article 2(d) Data Protection Directive
[17] http://www.ico.gov.uk/upload/documents/determining_what_is_personal_data/whatispersonaldata2.htm

- Can a living individual be identified from the data, or, from the data and other information in the possession of, or likely to come into the possession of, the data controller?
- Does the data 'relate to' the identifiable living individual, whether in personal or family life, business or profession?
- Is the data 'obviously about' a particular individual?
- Is the data 'linked to' an individual so that it provides particular information about that individual?
- Is the data used, or is it to be used, to inform or influence actions or decisions affecting an identifiable individual?
- Does the data have any biographical significance in relation to the individual?
- Does the data focus or concentrate on the individual as its central theme rather than on some other person, or some object, transaction or event?
- Does the data impact or have the potential to impact on an individual, whether in a personal, family, business or professional capacity?

If any of the points in the list are true, then the data has to be considered as personal data. An assessment based on this list for the location data collected from this study and indeed any such location based service is given in section 5.4.

## 5.3 Sensitive data

Article 8 of the Data Protection Directive contains provisions regarding special categories of data, commonly known as sensitive data, the processing of which is as a rule prohibited. Such data are the personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership and data concerning health data or sex life[18]. The processing of the aforementioned data is only allowed under the following conditions, stipulated in Article 8(2)[19]:

> (a) the data subject has given his explicit consent to the processing of those data, except where the laws of the Member State provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject' s giving his consent; or
>
> (b) processing is necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law in so far as it is authorized by national law providing for adequate safeguards; or
>
> (c) processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent; or

[18] Article 8(1) Data Protection Directive
[19] More specific provisions are contained in Article 8(3)-(7) Data Protection Directive

(d) processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body with a political, philosophical, religious or trade-union aim and on condition that the processing relates solely to the members of the body or to persons who have regular contact with it in connection with its purposes and that the data are not disclosed to a third party without the consent of the data subjects; or

(e) the processing relates to data which are manifestly made public by the data subject or is necessary for the establishment, exercise or defence of legal claims.

## 5.4 Location data

Location data are "any data processed in an electronic communications network, indicating the geographic position of the terminal equipment of a user of a publicly available electronic communications service"[20]. According to recital 14 of the ePrivacy Directive they are data that may refer to the latitude, longitude and altitude of the user's terminal equipment, to the direction of travel, to the level of accuracy of the location information, to the identification of the network cell in which the terminal equipment is located at a certain point in time and to the time the location information was recorded.

Although the ePrivacy directive does not make use of the term Location Based Services, article 2(g) of the Directive defines the term value added service as "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". Therefore a Location Based Service could be defined as a value added service which processes location data other than traffic data for purposes other than what is necessary for the transmission of a communication or the billing thereof. The use of the term 'location data other than traffic data' has created some confusion among the legal scholars. In simple words, this term refers to all location data that are not used for the transmission of a communication or for setting up a connection (these data are treated as traffic data).

It thus becomes evident that location data used in the frame of this study can be linked to the identification number of each participant, the identity of which is already known to the data controller and can, in any case, be deduced from an analysis of the location patterns derived from the analysis of his GPS data. The location data used in this study, as the location data used for the provision of most LBS applications, falls within the remit as a type of personal data. As such, the objective here is to mine the data gathered within this study to examine to what extent specific information can be revealed about the users.

### 5.4.1 Processing of location data

The processing of location data for the provision of Location Based Services is only allowed "when they are made anonymous, or with the consent of the *users* or *subscribers* to the extent and for the duration necessary for the provision of a value added service"[21]. In simple words when the data are not made anonymous, the user or the subscriber of the mobile device shall

[20] Article 2 (c) ePrivacy Directive
[21] Article 9 (1) ePrivacy Directive

give their consent[22] to the processing of the location data in order to enable the provision of the Location Based Service. However, even when the consent of the user or subscriber has already been obtained, the user or subscriber must continue to have the possibility, using a simple means and free of charge, to refuse the processing of such data for each individual request[23]. When the user initiates the service by calling for instance a number or sending an SMS this action shall amount to consenting to being located.[24]

The location data used for the provision of a Location Based Service shall be processed only to the extent and for the duration necessary for the provision of the service[25]. After that they should be deleted or made anonymous. As already mentioned above, the European Union recently adopted the Data Retention Directive, which regulates the retention of traffic and location data for law enforcement purposes. It is important to clarify at this point that this obligation covers only the provider of a publicly available electronic communications service or of public communications network (mobile operator, Internet Service Provider etc). Thus, in this tracking study, the data controller is not bound by the obligations of the Data Retention Directive and shall therefore delete the data upon the end of the processing of the location data for the needs of the tracking study.

### 5.4.2 Information to be given before the initiation of the Service

Before obtaining the consent, the service provider must provide the individual with specific information regarding the type of location data that will be processed, of the purposes and the duration of the processing and whether the data will be transmitted to a third party for the purpose of providing the Location Based Service.[26] More information needs to be given to the user according to the provisions of the Data Protection Directive and the ePrivacy Directive. Such information[27] deriving from articles 10 Data Protection Directive and articles 6 and 9 ePrivacy Directive is:

- "the identity of the controller and of his representative, if any,
- the purposes of processing,
- the type of location data processed,
- the duration of processing,

[22] Consent by a user or a subscriber corresponds to the data subject's consent (Art. 2(f) and Recital 17 ePrivacy directive) as it is defined in the Art. 2(h) Data Protection Directive, that is as "any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed.".
[23] Article 9(2) ePrivacy directive
[24] Data Protection Working Party, Opinion on the use of location data with a view to providing value-added services, adopted on 25 November 2005, 2130/05/EN (WP 115), p. 6
[25] Article 9(1) ePrivacy directive
[26] Article 9(1) ePrivacy Directive
[27] The list of the information to be provided can be found in Article 29 Data Protection Working Party, Opinion on the use of location data with a view to providing value-added services, adopted on 25 November 2005, 2130/05/EN (WP 115), p. 4-5

- whether the data will be transmitted to a third party for the purpose of providing the value-added service,
- the right of access to and the right to rectify the data,
- the right of users to withdraw their consent at any time or temporarily refuse the processing of such data, and the conditions on which this right may be exercised,
- the right to cancel the data"[28].

The information shall be provided by the party collecting the location data for processing. Thus it shall usually be provided by the provider of the value added service, or if this is not possible by the electronic communications operator. The information could be provided either directly each time the service is used or in the general terms and conditions for the value-added service. In the latter case the service provider should make the information available so that the individuals concerned can consult it again at any time and by a simple method, such as via a website or while using the service (e.g. dialling a toll-free number)[29]. In addition, in cases of ongoing processing of location data the individual shall be regularly reminded about the processing of his location data. However, neither the Directive, nor the Article 29 Working Party gave a clear guidance as to how "regularly" should be interpreted. It shall be a matter of justified decision of the data controller, who shall be responsible for the provision of the aforementioned information.

## 5.5 Mobile operators

The data controller is "the natural or legal person, public authority, agency or any other body which alone or jointly with others determined the purposes and means of the processing of personal data"[30] and is the one who is responsible for the processing of personal data and will be held liable for violations of the data protection legislation. Identifying the controller in a data processing operation is crucial not only for the fulfillment of the obligations imposed to the controller by the data protection legislation and the determination of liability issues that might arise, but also for the exercise of the rights of the data subject. The problem of defining the controller of the data in the new telecommunications networks is already identified by the Consultative Committee of the Convention for the Protection of Individuals with regard to automatic processing of personal data of the Council of Europe: "Nowadays […] this model in which a sole person or body is responsible for determining the parameters of the automatic processing is increasingly challenged by examples to the contrary. Several actors, among which the controller or co-controllers, the processor(s) and the service provider(s) interact in the processing. As a result, data subjects might not always know whom to turn to in order to exercise their rights"[31].

[28] Idem
[29] Idem, p. 5
[30] Article 2(g) Data Protection Directive
[31] Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data of the Council of Europe, Opinion of the T-PD in the interpretation of the concepts of automatic processing and controller of the file in context of worldwide telecommunications networks, as adopted by the T-PD at its 23rd meeting, T-PD-BUR, 08 E fin (Strasbourg, 15 March 2007); 2006.

In the most common (traditional) Location Based Service model, the controller of the data can be either the LBS Application Provider or the Mobile Operator or both of them can be co-controllers of the data, depending on who determines the purposes and means of the processing of personal data. However in practice, according to the contractual agreements between the LBS Application Provider and the Mobile Operator, it is the latter that actually determines the purposes and means of the processing of personal data. The Mobile Operator shall thus be considered controller of the data and responsible to the user and this fact needs to be included and demonstrated in the contractual agreements.

This solution is also to the benefit of the user, who will be able to exercise his rights in front of the Mobile Operator, as a single point of contact, and will not need to be involved in understanding complicated relationships between the entities involved for the provision of the service. He will sign a contract with the Mobile Operator and interact with the latter for the provision of a Location Based Service. It is also the Mobile Operator he will turn to in order to exercise his privacy rights. Such rights are the right to ask for the rectification of data, to delete them, block them, as well as the right to object to the processing of some of his data.[32]

The technological developments in the field of mobile services and applications have enabled the creation of free Location Based Services that can be downloaded by the user on their mobile phone without the involvement of their Mobile Operator, contrary to the case in the traditional LBS scenario. The LBS application provider in this model will only know an identification number for the specific device and in most cases he will have no other information about the identity of the user of this device. However, the LBS application provider will be able to draw a location profile of the user of the mobile device, based on the location data it collects. Therefore questions arise as to whether the location data collected and processed by the LBS application provider shall be considered as personal data or not. Critical in order to give an answer to this question is the concept of identifiability, which we have already presented above.

Recital 26 of the Data Protection Directive reads that in deciding whether data could be used to identify a particular person "account should be taken of all the means *likely reasonably* to be used either by the controller or by any other person to identify the said person" (emphasis added). Thus the recital sets two criteria for identifiability: the probability and the difficulty that tend to be interlinked.[33] In any case it is supported that the term "personal data" should include all data about a person (including economic, professional etc. data) and not only data about the person's personal life[34]. This breadth of the conception of personal data means that data are usually presumed to be 'personal', unless it can be clearly shown that it would be impossible to tie them to an identifiable person (that is, unless the data are truly anonymous)[35]. The interpretation of the Member States regarding the ease of identification differs significantly between them. The data protection laws of France, Germany and Sweden

[32] For a detailed analysis of the topic, see Kosta E., Zibuschka J., Scherner T. & Dumortier J., Privacy issues in location based services - Legal considerations on privacy-enhancing Location Based Services using PRIME technology, *Computer Law and Security Report*,, Volume 24, Issue 2, 2008, p. 139-146
[33] Bygrave L, *Data Protection Law – Approaching its Rationale, Logic and Limits*, Kluwer International, 2002
[34] Dammann, U., Simitis, Sp., EG-Datenschutzrichtlinie, Nomos Verlagsgesellschaft, 1997, p. 109
[35] Kuner, C., *European Data Privacy Law and Online Business*, Oxford University Press, 2003, p.51

for instance talk about "means for identification which are *reasonably capable* (as opposed to likely) of being put to use"[36].

According to the Article 29 Data Protection Working Party, "identifiability" can be interpreted in a very broad way, depending on the circumstances of the case. "In cases where *prima facie* the extent of the identifiers available does not allow anyone to single out a particular person, that person might still be "identifiable" because that information combined with other pieces of information (whether the latter is retained by the data controller or not) will allow the individual to be distinguished from others."[37]. Based on this broad interpretation by the Article 29 Working Party and given the fact that the Mobile Operator holds the identification data of the mobile phone user, at least in most of the cases, some may sustain that the location data shall be treated as personal data by the LBS application provider as well. It shall be however pointed out that such a broad interpretation of the concept of identifiability with regard to personal data has received great criticism from the industry.

[36] Bygrave L, *Data Protection Law – Approaching its Rationale, Logic and Limits*, Kluwer International, 2002, p. 44.
[37] Art. 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data, p. 13, WP 136, 20 June 2007, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf

# 6 User Data Analysis

The analysis present in this report is the first phase of analysis designed to understand the potential of the data. Although some algorithms have been utilised to help mine the data, much qualitative assessment has also been performed to aid in the exploitation, and this is presented here. The longer term objectives of the tracking study are to design complex mathematical modelling which will better inform of the potential for comprehensive profiling which could be performed as part of a Location Based Service (LBS) such as those discussed previously in this report. The data collated represents one month of four users' activity during April 2009.

### 6.1.1 Place of Residence and Work

It may not at first glance seem obvious as to why place of work is a good starting point for analysing persistently collated location data. However, if we consider how PoIs are generated, it is evident that key locations are those frequently visited and so will attract a 'cluster' of PoI points. In some cases, leaving and returning from a work building multiple times during the day may be commonplace – such as leaving and returning from lunch – and so will register multiple times per (work) day. This may well be in contrast to a residence which may only be left and returned to once a day.

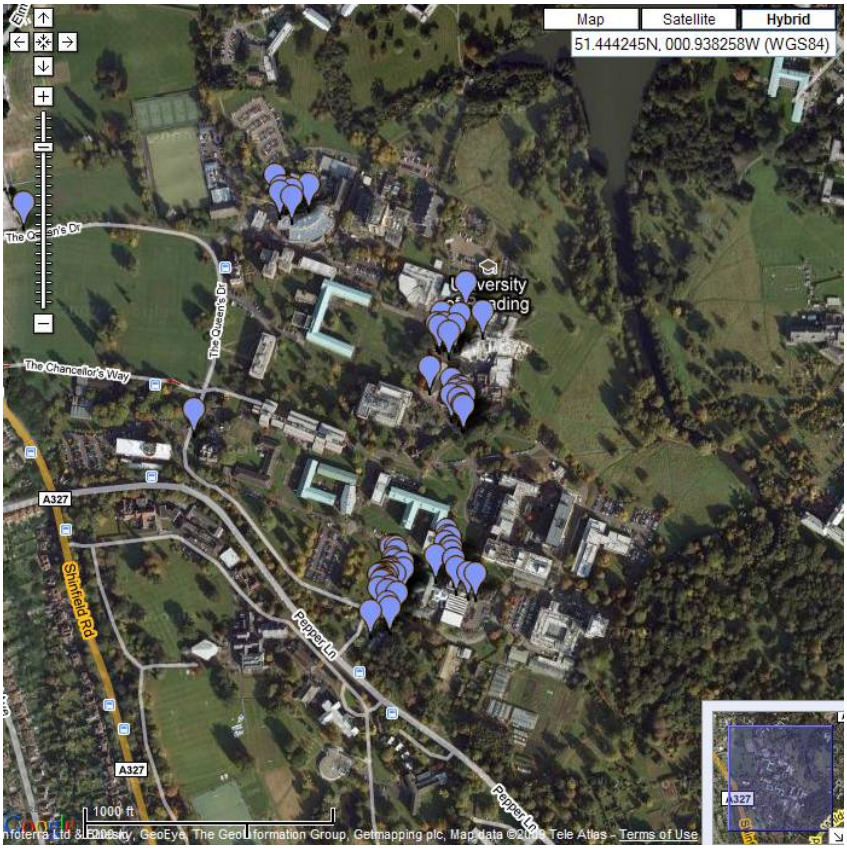
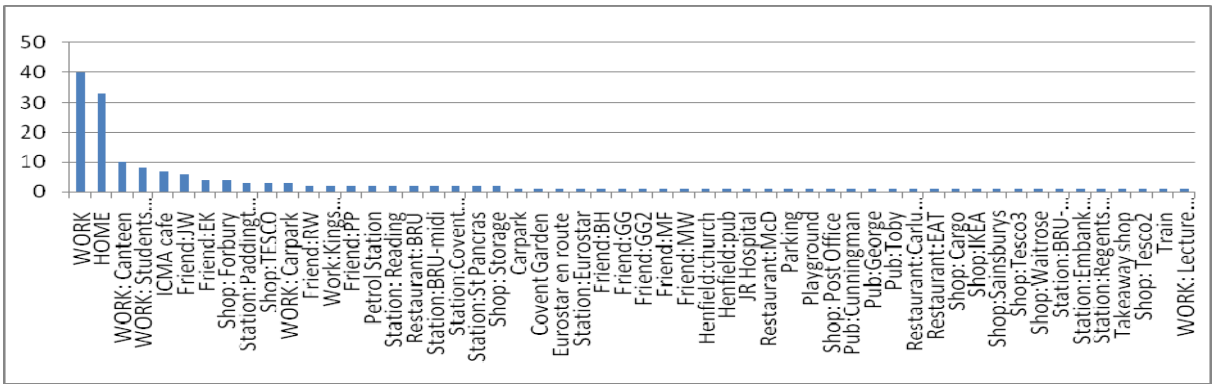


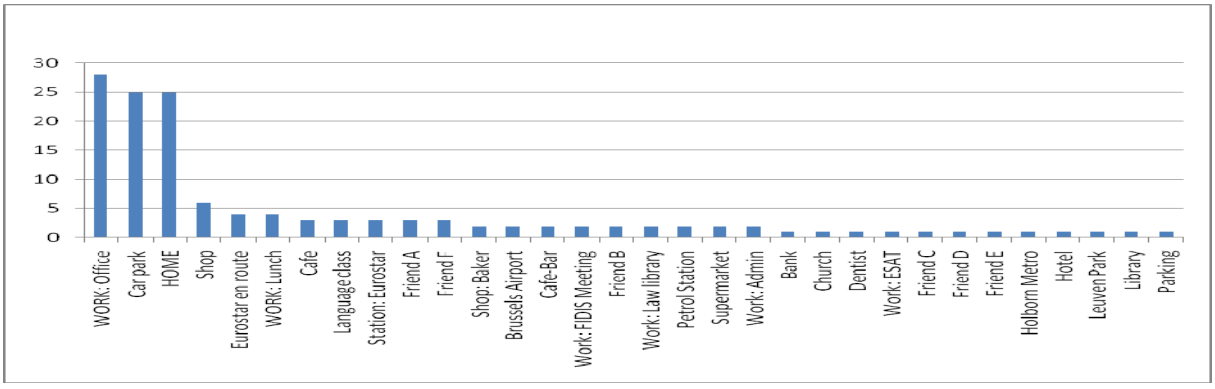

**Figure 6: Clusters of PoIs at key work locations collected from one user over the study period**

By calculating the frequency of recurring PoIs, the key locations for each user become apparent. These are given in the following graphs.

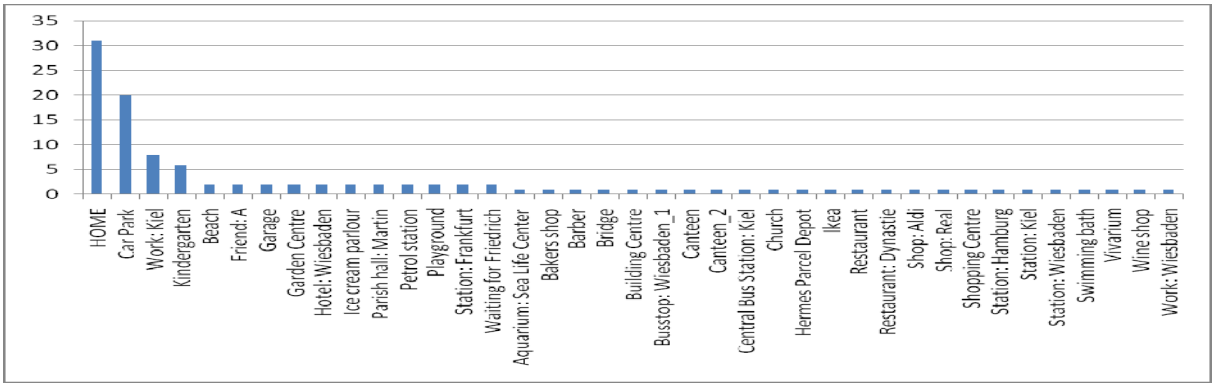
User 1:



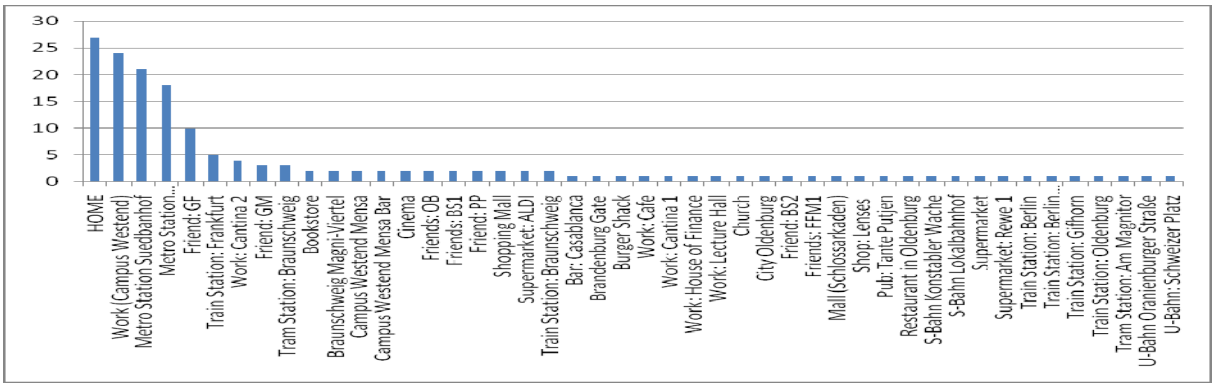

User 2:




User 3:



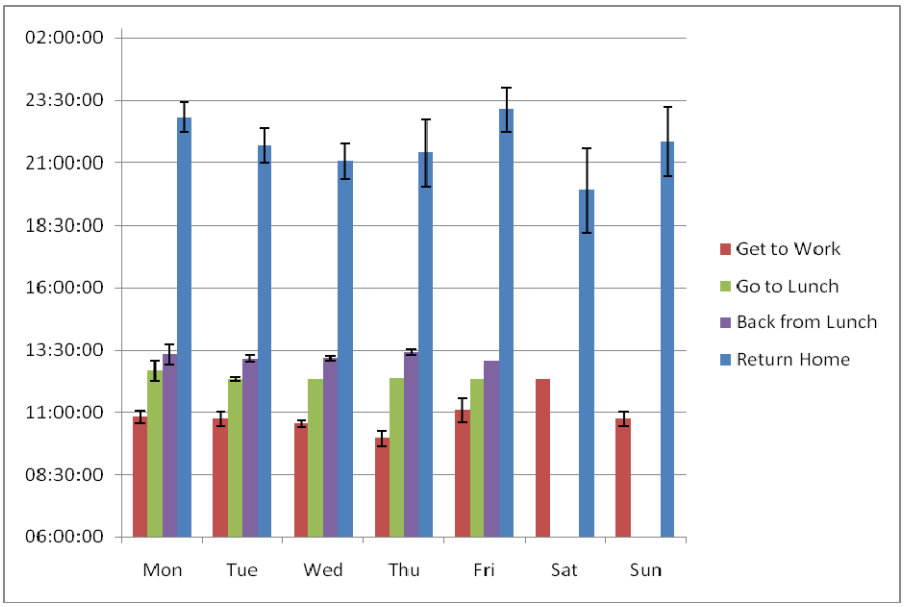

User 4:


30
25
20
15
10
5
0
HOME
Work (Campus Westend)
Metro Station Suedbahnhof
Metro Station...
Friend: GF
Train Station: Frankfurt
Work: Cantina 2
Friend: GM
Tram Station: Braunschweig
Bookstore
Braunschweig Magni-Viertel
Campus Westend Mensa
Campus Westend Mensa Bar
Cinema
Friends: OB
Friends: BS1
Friend: PP
Shopping Mall
Supermarket: ALDI
Train Station: Braunschweig
Bar: Casablanca
Brandenburg Gate
Burger Shack
Work: Cafe
Work: Cantina 1
Work: House of Finance
Work: Lecture Hall
Church
City Oldenburg
Friend: BS2
Friends: FFM1
Mall (Schlossarkaden)
Shop: Lenses
Pub: Tante Putjen
Restaurant in Oldenburg
S-Bahn Konstabler Wache
S-Bahn Lokalbahnhof
Supermarket
Supermarket: Rewe 1
Train Station: Berlin
Train Station: Berlin...
Train Station: Gifhorn
Train Station: Oldenburg
Tram Station: Am Magnitor
U-Bahn Oranienburger Straße
U-Bahn: Schweizer Platz

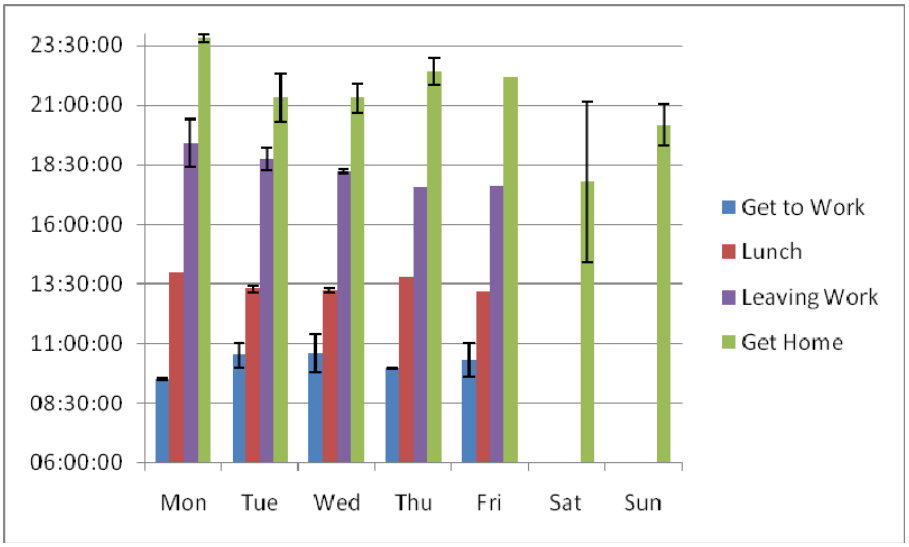
The use of temporal data can also serve to disclose the identity of some PoIs – for example the location occuring most often first and last thing during the day is likely to be the residence. Further analysis can also shed light on typical work day routine, and give an indication of job type undertaken by the users.

The following graphs show the average daily routine for three of the four users, with error bars showing the standard deviation from these averages.

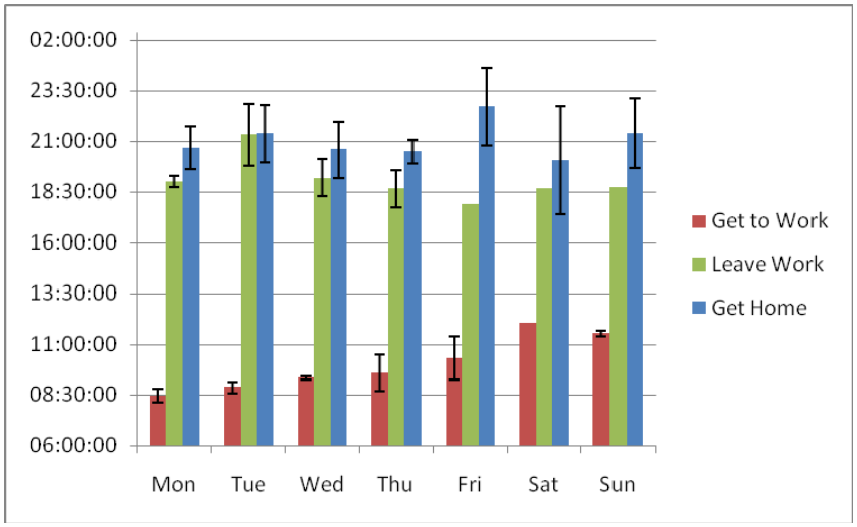
User 1:




User 2:



User 4:




Notably none of these three users have a rigid routine, demonstrated by the notable deviations from the average. They do however all keep core working hours, and in some cases this involves work over the weekend. While care needs to be taken in interpreting these results, especially due to influencing factors such as the ability to work from home – given the location can be determined, a likely job profile could be drawn for these users. Certainly, for example it is unlikely that these people are involved in a clerical role, given the flexibility in working hours.

Notably, User 3 lacks a definable daily routine, and visits an office environment only a few times during the study. This user does not conform to the profile of an office worker, however the lack of routine also does not indicate that of a housewife/husband. Given the data it is not possible to clarify the exact work status of this user.

### 6.1.2 Gender

Very little information is gender specific, and while visiting specific shops, or notable shopping patterns could be an indicator, this could also be because the user is accompanied by a member of the opposite sex. Without specific indicators – such as using a specific public toilet or entering a gender specific environment, which does not occur in this data – no gender can be assign with confidence.

### 6.1.3 Social Status

There are several indicators for social status in the data. Firstly, a key one is the area in which the users' residence lies. Since this is easily determined from the data, and given the growing availability of online geo-demographic resources to assess an area based on varying criteria,

this becomes somewhat trivial although potentially misleading. For example, User 1 lives in an area that is classified as type 13 in the ACORN[38] classification system meaning that:

*'These are predominantly well-off professional people living in urban areas of the UK. Most are well educated individuals in professional and managerial occupations, but there are also students and young singles starting out on their careers.'*

This in itself leaves great room for interpretation – this user could be a student through to a high earning professional. However, coupled with other information, for example they shop at Waitrose supermarket – considered to be a shop of choice for older, wealthier people - it is more likely the latter. In the UK the Land Registry also provide a searchable database[39] detailing when properties were sold, and what price was paid for them. In the case of persistent tracking, if it is detected that a user has changed residence, this type of extra information could also indicate whether a user had bought a property or was renting.

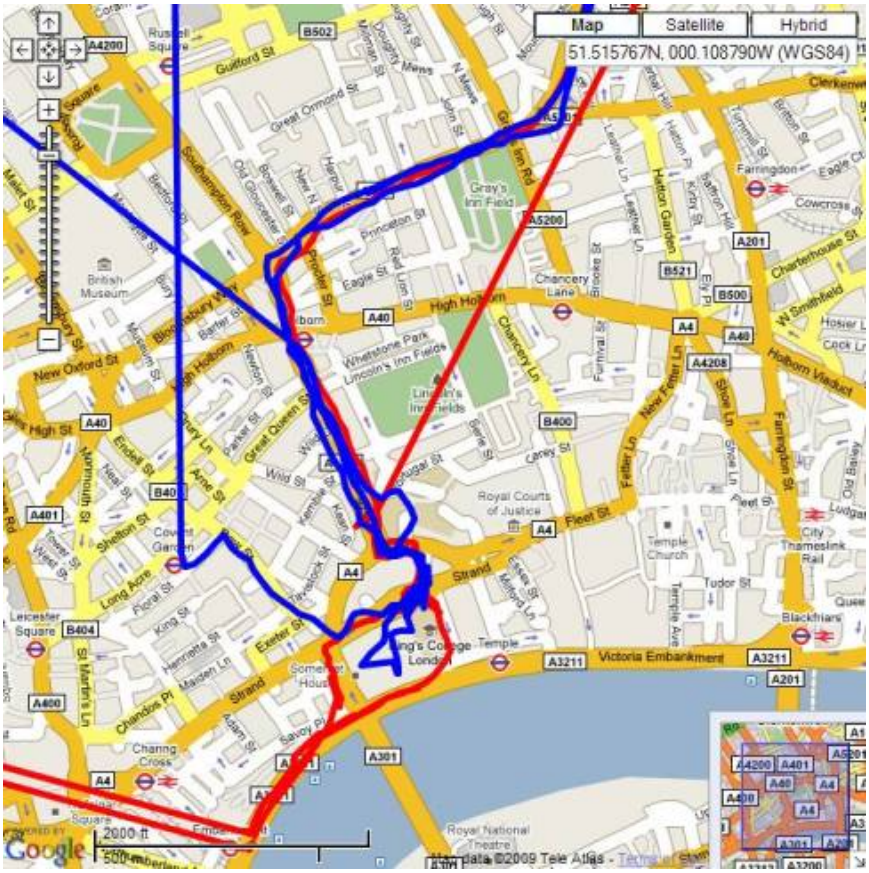



**Figure 7: Inferring a relationship by observing occasions of spatial and temporal unity**

[38] http://www.caci.co.uk/acorn/whatis.asp
[39] See, e.g. http://www.houseprices.co.uk/

Another measure of social status is the people with which a user socialises – this is especially useful if these people are also subscribers to the LBS and are being profiled. In the case of this study, by looking for instances whereby two users were in proximity of each other, a social or business relationship could be inferred. Figure 7 shows the route of two individuals - they use different tube stations, but then arrive at the same location before walking together to a second location and back again (note the time information is not shown). By observing the times at which this happened, and the end locations (points of interest), especially if it happens multiple times, we can infer whether this is a social or business relationship. In this case the users meet from 11:02 to 16:24 at King's College in London, UK. This would imply a business relationship, and would also indicate that there should be a degree of compatibility between the working lives of both users.

The occurrence of travel can also be an indicator of social status and is incorporated into the ACORN system described above. However, in this data set, of the two users who engaged in foreign travel, both appeared to do so mostly in a work context. Notably User 4 has a dependency on public transport while the others evidently own their own car which they predominantly use in preference.


**Figure 8: Disruptions in walking highlighted as PoIs**

### 6.1.4 Family life

Establishing marital status or whether there are children associated with a person is of potential interest. From GPS data this has to be mostly inferred unless specific identifiers exist. In the case of user User 3, periodic trips to the kindergarten and the park suggest young children are involved. The lack of identifiable working pattern is also suggestive of there being a spouse or partner who goes to work. For the other users, the largely unstructured and unsociable working hours is suggestive of single people. Interestingly, User 4 exhibited false PoIs at certain times during walking, see Figure 8, which could be suggestive of variations associated with accompanying young children at that time. These interesting variations in walking patterns are the subject of further research. In any case, this gives an indication that GPS data may not only reveal where you are, but potentially who you are with.

### 6.1.5 Routine

Of the three users with a definable working (and thus daily) pattern, there is also the suggestion of an underlying weekly routine as well. While this is not surprising, the duration of one month is not enough to allow a clear picture of this, especially when some days are evidently not part of the 'normal' routine (i.e. when travelling). However, there are some glimmers of routine apparent in the data. Of the two car drivers from this group, looking at their trips to the petrol station, for example, it is evident that they go at set times during the week, perhaps because of conveniently located petrol stations. This is suggestive of shopping habits which are either brand loyalty or convenience based. Given longer periods of tracking, it would be likely that further such routines would be apparent for daily, weekly, monthly and yearly periods, and these could further shed light on the nature of the person being tracked.

## *6.2 Sensitive data*

Sensitive data are data belonging to special categories of personal data, the processing of which is in principle prohibited and shall only be allowed under specific conditions and strict safeguards. Any information about a living individual that includes facts and intentions or opinions about any of the following matters is considered sensitive data[40]:

- race or ethnic origin;
- political opinions;
- religious or other beliefs;
- trade union membership;
- sexual life;

[40] Article 8(5) Data Protection Directive stipulated that "Processing of data relating to offences, criminal convictions or security measures may be carried out only under the control of official authority, or if suitable specific safeguards are provided under national law, subject to derogations which may be granted by the Member State under national provisions providing suitable specific safeguards."

- (physical or mental) health life;

Given that the processing of sensitive data is in principle prohibited and shall only be allowed under specific conditions and strict safeguards, see section 5.3, it is interesting to examine whether location data can potentially reveal information about any of these aspects.

### 6.2.1 Religion

Religion, like much of the information to be inferred from location data, is one whereby it may be possible to establish with some degree of certainty that a person is of a specific religion if certain indicators exist – whereas if they do not, it is simply inconclusive. In other words – it may be possible to classify a person as being religious, but not possible to classify a person as having no religion. Because most mainstream religions have a defined routine – be it e.g. Sunday morning church services or periodic calls to prayer – which are held in specific and identifiable locations, be it e.g. a church or a mosque, associations between a person and a religion are easy to make. In the cases of the data recorded for this study, three of the users did attend a church – however not routinely. It is not so unusual these days for people to be casual observers of a religion, and so only attend key religious events. The recording for this study included the Easter period, which is such a time, and so could explain these anomalies – additionally the variation in celebration dates, e.g. Greek Orthodox Easter Sunday - April 19$^{th}$ and Western Easter Sunday - April 12$^{th}$, could inform further. However, in this case only one of the users attended a church at a time consistent with an Easter celebration, and so the other two may have e.g. been at a wedding or christening which is of no help. For the one user who did attend an event, it is still very much inconclusive as several other factors, including family inclusion, may explain such a one off.

### 6.2.2 Sexual Life

Location specific indicators and social interactions are a potential source of information regarding inferring someone's sexual preferences. In the case of the data in this study, no such obvious indicators exist. In any case, simply attending a gay-bar for example is not sufficient to draw conclusions - rather it would have to be a wider picture drawn from disparate elements. Notably drawing the conclusion that a person is heterosexual is as relevant as inferring homosexuality or bisexuality, but is in itself potentially still non-trivial.

### 6.2.3 Health

Attending a doctor's office or specific health clinic routinely is potentially an obvious indicator of a health issue – although not necessarily that of the person being tracked as they may be accompanying another. Conversely, *not* attending a health professional, such as a dentist or optician, on a routine basis could be an indicator of general apathy regarding health. Movement patterns may also be an indicator to underlying medical problems. More concretely, the amount of physical exercise a person takes has been directly linked to risk factors associated with ailments such as heart disease. Exercise can be quite apparent in location data through walking or running – although attending a gym is another albeit less

measurable sign. The chief medical officer in the UK recommends that adults should do a minimum of 30 minutes moderate-intensity physical activity, five days a week. On this basis, it would seem that none of the four users in this study take enough exercise, and so their health could be at risk. This is of course making the assumption that they do not e.g. exercise at home, or have access at work to a gym that would not be apparent in the location data.

### 6.2.4 Commission of an Offence

The data relating to offences, criminal convictions or security measures are a specific type of data that can be considered as sensitive depending on the national data protection legislation of the Member States.[41] Having proof of location could be useful in confirming an alibi or as proof of being at a location during the time of an offence (also see section 6.3). However, in some cases it may be a direct evidence of an offence having been committed. While localised variation in GPS data is probably too large to indicate a speeding offence over a small distance, averaged over a large distance it can be quite accurate.

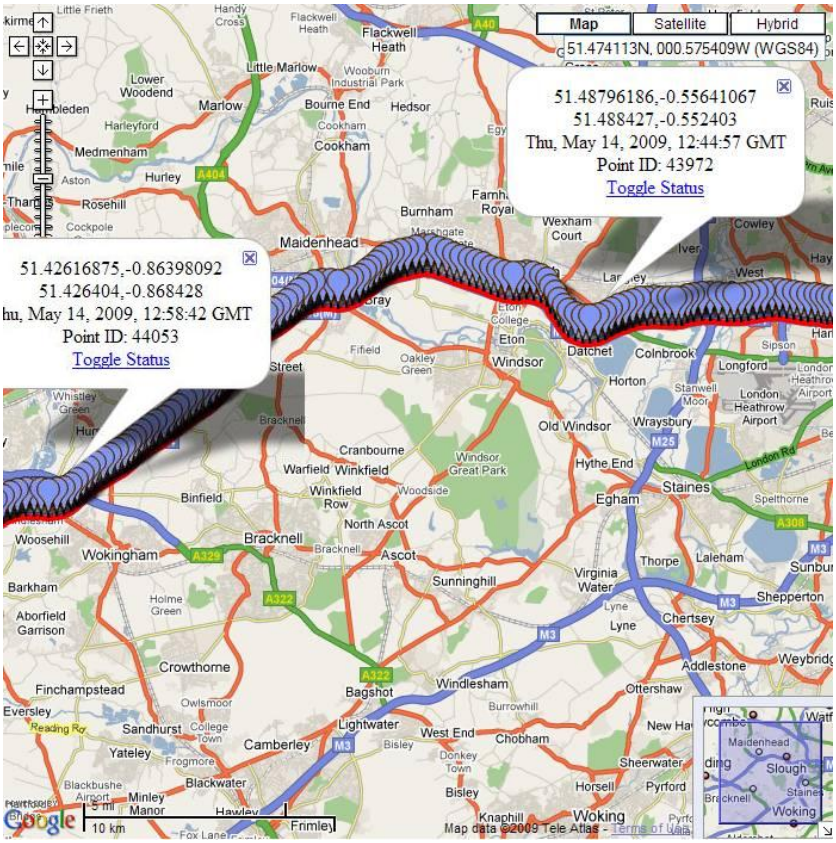



**Figure 9: Two specific points indicated on the M4 motorway with times the user was at these locations**

[41] Article 8(5) Data Protection Directive stipulated that "Processing of data relating to offences, criminal convictions or security measures may be carried out only under the control of official authority, or if suitable specific safeguards are provided under national law, subject to derogations which may be granted by the Member State under national provisions providing suitable specific safeguards."

Figure 9 shows part of a car journey along the M4 motorway in the UK. Indicated are junction 5 and junction 10 which are 16.1 miles apart. The time for this part of the journey can be calculated as taking 13 minutes and 45 seconds meaning an average speed of 70.25 miles per hour (MPH) was achieved over this distance. With the speed limit on such a road being 70 MPH, this driver was approximately legal *on average* on this occasion.

## 6.3 Data confidence

A mobile handset is quite a unique piece of technology since it is very likely to be carried by the same person everyday, all day. However, there are instances where the owner will forget their phone, and, although less likely, lend their phone to someone else. As such, it would be useful to be able to link the handset to an individual in a direct way such that there is more confidence in the data being directly relevant to the person in question. In this study we have employed two techniques to achieve this: the first is an active approach, which requires the user to authenticate to the device using their fingerprint. However, as can be seen in Figure 10 requiring action on the part of the user means that large periods of time can elapse between authentications, and so large periods of data with poor confidence occur. On average the three users who utilised this technique authenticated with the device 4.5 times per day. One user in particular was notable for periodically authenticating with their device regularly, sometimes up to 15 times per day. However, this still leaves considerable periods between authentications.

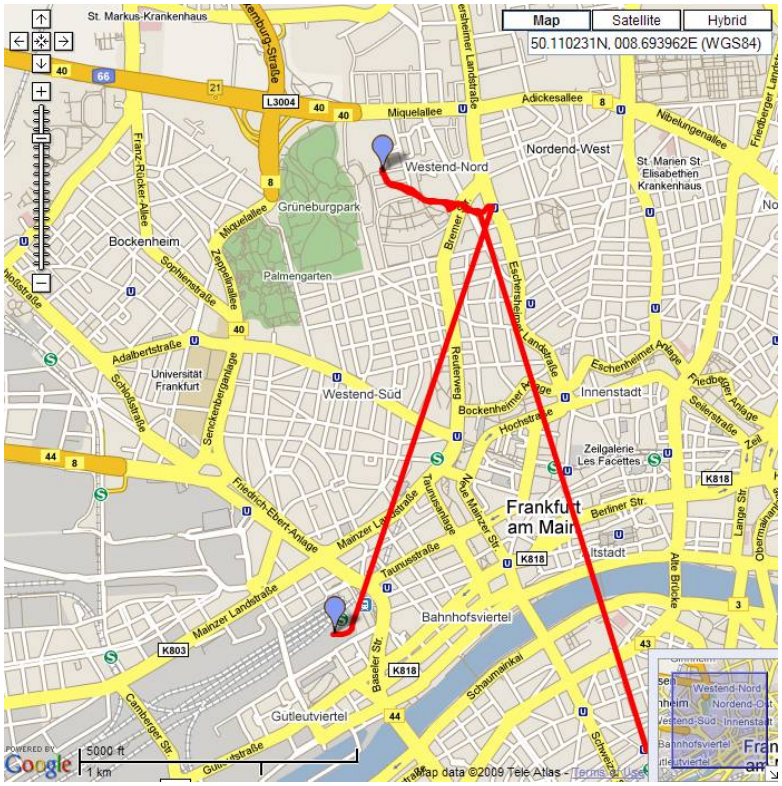



**Figure 10: The blue flags mark the locations where the user authenticated with their device – some 10 hours elapsed between these times**

An alternative method of passive authentication with the device was via an implanted RFID tag. By adding an RFID tag reader onto the handset, the handset was able to periodically try to read tags within range. With a range of only a few centimetres, reading was sporadic, but could be attempted until a read was successful. This worked especially well when the handset was being held, or when the user's hand was near to the pocket with the handset inside.

**Figure 11: An x-ray showing an RFID tag implanted near the base of the thumb in the non-dominant hand of one of the study participants**


**Figure 12: Blue flags indicate points where the handset authenticated the user via the implanted RFID tag**

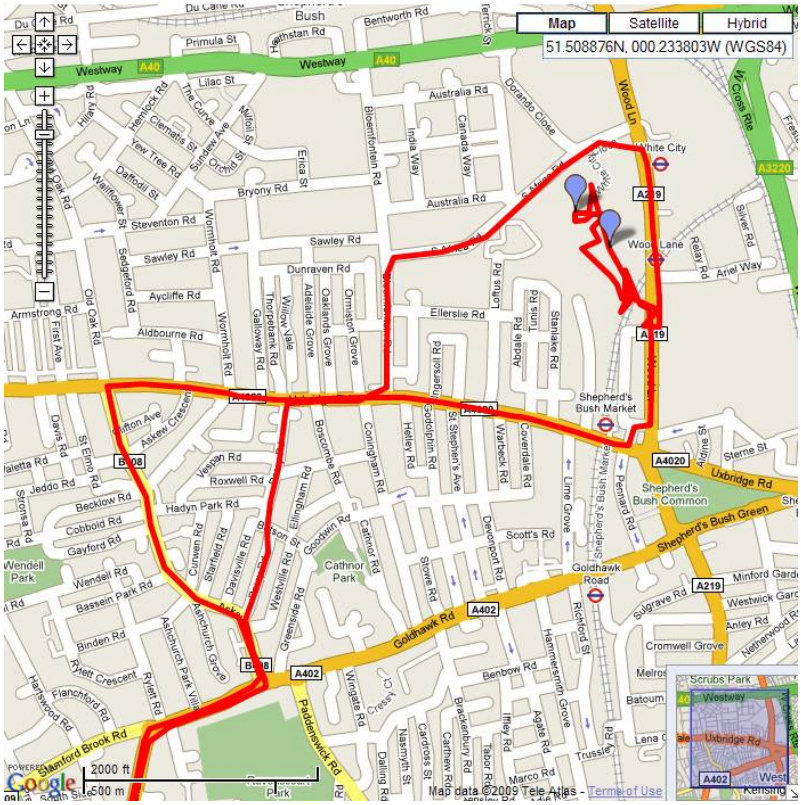



**Figure 13: The route taken by two different people using the same handset. The blue flags in the top right mark the building where the users exchanged the handset with each other**

Figure 12 shows the user taking a 25 minute walk, in which time the tag is authenticated 13 times. Notably during the periods the user is in their car – the two traces coming in from the left hand side, the tag is not read at all as the hand on the steering wheel of the car is too far away from the phone.

Another method by which the user could be authenticated is an 'intelligent' approach, using previous information regarding the user. Figure 13 shows the route taken by one individual in a car from the bottom left to a building at the top right marked by the blue flags, and then another route taken by the a second person in a car holding the same handset from the building back down to the same location off the bottom left. It can be seen that one person uses local knowledge of the area in order to navigate smaller side streets, while the other person takes the main roads between the two locations. This type of deviation could potentially be used to indicate unusual behaviour of a user, the interpretation of which would depend on the application area.

## 6.4 Conclusions

It is evident that an enormous amount of information is buried in the data available from persistently tracking people – however, it is also clear that a one month period is by no means long enough to draw a rich profile of the person. If we consider the potential application of a recommender system, one of the typical recommendations is likely to be restaurants in an unknown town or city. In the month long study conducted, very few occasions of eating out

actually occurred and so it would be difficult to draw good conclusions from that. In any case, the behaviour of the individual may well differ greatly when they are in an unknown location –for example they may prefer local cuisine rather than whatever they go out to eat normally. Equally they may prefer to spend more when travelling, and so their profile could become quite different. This type of idiosyncratic behaviour will be difficult to model and profile, and really could only come together after vast amounts of time monitoring the individual. What is clear is the real potential for incorrect conclusions being reached based on the data, the impact of which on the individual could be significant.

In any case, it has been demonstrated that over relatively short periods of time, personal, and in some cases sensitive information can be revealed, and so it is certainly within the interests of the individuals to make sure that they are aware of how this information is being further exploited by those offering the services they are using. One interesting point of discussion is the basis for considering these elements as "personal data" as defined in the Data Protection Directive. Personal data is *"any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his/her physical, physiological, mental, economic, cultural or social identity"*. In the case where such a service is offered by the mobile phone provider, then it is clear that since they hold on record for billing purposes the identification (e.g. real name) of the individual, they can trivially link the phone and thus the data to an individual – that person as the source of the data is clearly identifiable. However, if the service is offered for free, without registration as a downloadable additional service on an internet enabled handset, then this is less clear. In any case, the individual *could potentially* be identified from the data eventually and according to the broad interpretation given by the Article 29 Data Protection Working Party, it will most likely be considered as personal data in this case as well.

# 7 Conclusions

While it is evident that the month long period undertaken for this pilot study is not enough to draw substantial conclusions, and thus elaborate profiles of the individuals, the analysis of the user data that was conducted in the frame of this tracking study allows us to make some very interesting remarks. The location data, besides the actual location of the person, can be used to infer a number of information about the person concerned. Figure 7 is a clear example of two users, who were in proximity of each other for a long time and thus a business relationship could be inferred, after taking a closer look into their moving behaviour. This can be in itself problematic, if we take into account the number of GPS enabled devices that already exist and will soon be available in the market.

Very important, especially from a legal point of view, is the realisation that location data in several instances can potentially reveal information relating to the health life of the individual, his political or religious beliefs, etc. The data that relate with these matters, as it was already elaborated above, are considered as sensitive data and their processing is in principle prohibited. As location data can be linked to such matters and can eventually reveal information about the health life or the religious beliefs of an individual, they shall be treated as sensitive data. Without implying that all location data should be treated as sensitive data "just in case", the implications arising from such a realisation cannot be neglected.

A research report from Berg Insight predicts that more than 960 million mobile handsets sold in 2014 will have integrated GPS receivers and many if not all of these devices will be data enabled. As already stated above, the miniaturisation of GPS devices and their inevitable inclusion in such mobile handsets has generated a new era of information disclosure, and new services are likely to appear which encourage people to reveal where they are at any time in the name of safety, convenience or for social use. In the realisation of Ambient Intelligence environments, location data will also play a crucial role. However, if the collection and processing of location data reveals information that falls under the category of sensitive data, then probably the whole legal and business model around it will need to change. A lot of questions arise, as a result of our user data analysis, with regard to location data. Is awareness raising with regard to the privacy dangers arising from having continuously enabled a GPS device enough for the protection of the users? Shall the legislation include location data in the categories of sensitive data, in order to maximise the protection to be offered to the users? Shall location data be regulated in a specific way, varying from the existing rules on personal and sensitive data?

While the objective of this study is not to answer these questions, what is evident is that at the very least the users of these devices must be made aware of the privacy risks associated, especially since giving consent can be as simple as ticking a box on an end user agreement which experience has shown usually does not actually involve the user reading the text.