QROWD - Because Big Data Integration is Humanly Possible

Innovation Action

Grant agreement no.: 732194

D11.1 – POPD - Ethics

Due Date	30th May 2017
Actual Delivery Date	30th May 2017
Document Author/s	Luis-Daniel Ibáñez (University of Southampton)
Version	1.0
Dissemination level	CO
Status	Final
Document approved by	Enrico Bignotti (University of Trento)
	Mattia Zeni (University of Trento)
	Claus Stadler (InfAI)

Project co-funded under Horizon 2020 Research and Innovation EU programme

TABLE OF CONTENT

Page

EXECUTIVE SUMMARY

QROWD integrates humans and their data to the Big Data value chain for smart transportation. Ethical and privacy-protective treatment of these data is crucial to the development of the project. This deliverable describes the foreseen ethics considerations of the collection and processing of the data required for research and development of the project. It is a confidential deliverable intended for the members of the consortium and the Ethical reviewers of the European Commission.

Human participant data in QROWD is divided in two sets: (i) Data collected from external crowdsourcing platforms to help with specific Big Data processing tasks and (ii) data collected from humans actively participating in one of the two pilots, through tools developed by Consortium Members. At the time of delivery, the specific experiments to be carried out are not yet defined, a requirement for obtaining the authorisations from the relevant bodies. As such, we also describe the ethics procedures of the relevant members of the consortium, all complying with the appropriate European directives.

Experiments using (i) will be submitted to the approval of the University of Southampton Ethics committee. For experiments using (ii), those concerning the Trento use case will be submitted to the approval of the Ethics committee of the University of Trento, while those concerning the TomTom use case will abide to the privacy and data protection directives of TomTom.

1 ETHICS CONSIDERATIONS

The QROWD project aims at supporting participation and feedback of various stakeholder groups to foster data-driven innovation in cities. In particular, the integration of citizens -- whether it is through their direct engagement, or through the analysis of their data -- in the data-driven innovation loop is of utmost importance for the effective development of the project. As such, the consortium aims at preserving the privacy and personal data of the human participants in the project's life. In the following, we describe the ethical considerations for each aspect of the experiments where ethics considerations need to be taken in account.

1.1 Recruitment of Participants

The human participants in QROWD will be engaged on volunteering basis through a variety of channels.

For the crowdsourcing components developed as part of the Crowdsourcing services, recruitment will be through (i) paid crowdsourcing marketplaces, e.g., CrowdFlower; and (ii) not-paid crowdsourcing platforms, e.g., Wikidata Games, to perform tasks related to hybrid computations along the Big Data value chain. No personal data will be collected.

For the hackathon in WP1 and the ideas competition in WP2 through the advertisement targeted at the relevant stakeholder groups. These groups include companies, researchers, municipalities, interested citizens and social entrepreneurs. The advertisements will appear in social media, relevant mailing lists and further communication channels available to the QROWD partners. Personal data will be collected for informative purposes only, and won't be processed.

For the QROWD tools and services that require user input and feedback (e.g. with respect to the road and mobility conditions in business cases WP1 and WP2), human participants will be recruited among the users of these services and tools on volunteering basis. Participation will in all cases be cleared by the participants' informed consents.

1.2 Secondary use

Use of social network data and data from TomTom in the context of the business cases of WP1 and WP2. Data will be anonymised and processed according to standard-compliant procedures. The authorization for using these data in QROWD will be included in the consent forms to be signed by participants. Data from TomTom will follow their privacy and personal data

policies[1].

1.3 Scope and extent of participant tracking

We will capture behavioural patterns, geo-locations and responses of the human participants that agree to provide enhanced road and mobility information based on geo-location information, and to carry out analytics for identifying deadlocks, bottlenecks, mismatches, populations dynamics, etc. Information about this will be provided to the participants in the information letter, along with the purpose and treatment of data, and will be cleared by the participants' informed consents.

1.4 Storage, protection, retention and destruction

Raw data will be collected and stored in secure facilities provided byt the leading research partners (Southampton, UniTN and InfAI). Non-personal data that is deemed important for replication, repeatability and generalisation of the outputs of the project will be published according to the principles of Findability, Accessibility, Interoperability and Re-usability outlined by the European Commission (cf. Data Management Plan - Deliverable 10.1)

Personal data (e.g. responses during participation in the pilots, and geo location data required for business cases WP1 and WP2) will be dealt with on the basis of personal data arrangements outlined above (including volunteering principle, consent form and privacy considerations). Ethical approval (including privacy and legal issues) will be obtained prior to the user studies, as required by the corresponding institution(s).

2 ETHICS PROCEDURES

Human participant data in QROWD is divided in two sets: (i) Data collected from external crowdsourcing platforms to help with specific tasks and (ii) data collected from humans actively participating in one of the two pilots, through tools developed by Consortium Members. At the time of delivery, the specific experiments to be carried out are not yet defined, a requirement for obtaining the authorisations from the relevant bodies. In the following we describe the procedures to be carried out by the consortium for each data type.

---

[1] https://www.tomtom.com/en_gb/privacy/general/#Ourtenpromisestoyou

2.1 Crowdsourced data

Collection of data through crowdsourcing platforms and their processing will be coordinated by the University of Southampton. All tasks developed will be designed for computer based performance, and will be targeted at healthy, non-vulnerable, volunteer individuals.

The University is committed to providing a competent and rigorous ethics review for all studies involving human participants. Each study is registered in the Ethics and Research Governance platform, where is reviewed by the Faculty's Ethics Committee (FEC) to assess if the study requires approval from the general Research Governance Office. QROWD's research requires only FEC approval.

The FEC ensures through its review process that all such projects adhere to applicable Faculty and University policies, legislation, professional guidelines and best practice. FECs include a minimum of 6 academic staff members and one student representative from the relevant Faculty (with one senior member of the Faculty management committee or appropriate nominee to act as Chair), and one independent or "lay" member from outside the University. They meet face to face at least once in every semester but otherwise may conduct meetings and discussions via email or other electronic media.

2.2 Pilot participants data

Trento Business Case

Data collection and processing for this use case will be conducted by the University of Trento. Trento's Ethics committee refers to the national, European and international framework of juridical, deontological and ethical nature, with a view on pluralism of ethical approaches. It is inspired by the principles established on the basis of the good clinical practice and experiments involving human participants, in particular with reference to the current edition of the Declaration of Helsinki.

QROWD's experiments will be carried out using the I-Log application[2]. I-Log was developed in the frame of the Smart UniTN project for research in the field of Computational Humanism. In particular, the use of I-Log was approved for collecting data to investigate how the spatio-temporal organization of the everyday life of students may affect their academic performance.

I-Log collects data from smartphones' internal sensors, as detailed in Table 1.

---

[2] http://trams.disi.unitn.it/

Collected Data	Frequency
Acceleration, Linear Acceleration, Gyroscope, Gravity, Rotation Vector, Magnetic Field, Orientation, Temperature, Atmospheric Pressure, Humidity	20 times per second
Proximity, Wifi network connected to, incoming/outcoming calls (no audio), incoming/outcoming sms (no text), screen status (on/off), flight mode (on/off)	On change
Position, Wifi networks available, Bluetooth devices available	Each minute
Application running	each 5 seconds

Table 1: Data currently collected by the I-Log application

Users are invited to provide feedback on what they are doing, helping researchers to interpret the output of the sensors. The feedback is asked in form of a questionnaire designed by a group of sociologist of the University of Trento. Each question is generated at fixed time intervals and is composed by three subquestions. The user can reply as soon as he/she sees the system notification (as showed below) or he/she can decide to reply in another moment. A maximum of 5 questions are stored and are available for answering.

When the user registers to the system it generates a random string called salt (UUID). From that moment on this unique identifier will be used to write the data to the database. Every time the user pushes data from the smartphone to the backend it uses his/her username and password to login. This information is used to identify the UUID of the user and save the data to the database. There is no direct link between the username and the data.

The Information sheet and consent form for the use of I-Log for Smart UniTN are included in Annex 1 and Annex 2 respectively. The informed consent for QROWD's research using I-Log will be modeled after them. Additional sensor data may be collected for QROWD. All data collection using I-Log will be submitted to the approval of the Ethics Committee of the University of Trento

Additionally for this use case, Consortium partners processing data under the responsibility of the Municipality of Trento must comply with the Italian Data Protection Code[3] and the municipal "Data Protection and Confidentiality Regulation"[4]. Both regulations are inline with the European Data Protection act, for which the infrastructure of all partners complies.

---

[3] http://www.garanteprivacy.it/home_en/italian-legislation#1
[4] http://www.comune.trento.it/Comune/Atti-e-albo-pretorio/Regolamenti/A06-Regolamento-per-la-tutela-della-riservatezza-dei-dati-personali

TomTom Business Case

Data Collection and processing in this use case will follow TomTom's privacy and data protection directives. TomTom's policy[5] complies with the EU Data Protection Act. It guarantees that personal data is:

1. Processed fairly and lawfully
2. Collected for specified, explicit and legitimate purposes
3. Adequate, relevant and not excessive
4. Accurate and up-to-date when required
5. Data that allows identification is only kept for the time is necessary
6. Process personal data only
    a. the user has unambiguously given his consent; or
    b. processing is necessary for the performance of a contract to which the user is party or in order to take steps at the request of the user prior to entering into a contract; or
    c. processing is necessary for compliance with a legal obligation to which TomTom is subject; or
    d. processing is necessary for the purposes of the legitimate interests pursued TomTom, except where such interests are overridden by the interests for fundamental rights and freedoms of the user and in particular his right to privacy.

For the particular case of QROWD, where data about location will be used for improving current services and develop new ones. Users are informed through the following excerpt of the user agreement:

"

To work correctly, TomTom software uses location information from your device.

Some features of TomTom software have to send this location information to TomTom. The software also sends information that tells us your MyTomTom account name, identifies your device, tells us how you use the software and includes information you enter.

Some features will also allow you to send information, including your location and who you are to others such as friends, family, or people in your contacts or on social media.

We need your permission before we can use your information and enable these features.

We use the information so we can do the following:
• Send information to your device that is relevant to where you are.
• Send information that you choose to send about yourself or your location when using a specific feature.
• Improve the quality of our products and services, but only after we have made sure the information cannot be linked back to you.

The information is sent using the internet. Depending on your device, it can use its own

---

[5]

https://www.tomtom.com/en_gb/privacy/general/#TomTomspolicyonprivacyandprocessingofpersonaldata

connection, the connection provided by a smartphone or your computer's internet connection when you attach your device to it. If an internet connection is not available, the information is stored securely on your device to be uploaded later.
"

Concerning the management of location-identifying data, we cite the following excerpt from TomTom's policy:

"*Within 24 hours of you shutting down your device or app, TomTom automatically and irreversibly destroys the data that would allow you or your device to be identified from the location data we received.*

For Traffic, SpeedCameras, Danger Zones and Weather we delete the information within 20 minutes after you have stopped using the service by shutting down your device or app. We do not know where you have been and cannot tell anyone else, even if we somehow were forced to.

This, now anonymous, information is used to improve TomTom's products and services, such as TomTom maps, Traffic, products based on traffic patterns and average speeds driven, and for search queries to inform businesses how well-received their information is. These products and services are also used by government agencies and businesses."

ANNEX 1. INFORMATION SHEET FOR SMARTUNITN

INFORMATION FORM FOR THE SMART UNITN PROJECT

Smart Unitn is jointly developed by the Department of Information Engineering and Computer Science and the Department of Sociology and Social Research of the University of Trento.

The research objectives of the Department of Sociology and Social Research are:
• Addressing methodological issues of data collection of tools used in social research, e.g., questionnaires;
• Investigate how the spatio-temporal organization of the everyday life of students may affect their academic performance. To do so, software for the visualization of spatial data (e.g., QGIS) will be used to analyse data from sensors via clustering algorithms. Moreover, by adopting statistical models such as sequence analysis, we will try to identify characteristics of the everyday activities (order and duration) to understand whether and how the activity distribution in time can affect academic performance.

The research objectives of the Department of Information Engineering and Computer Science are:
• Validation of data collection methods of smartphones combined with questionnaires for ground truth;
• Developing algorithms for matching sensor values and descriptions used by people in their everyday life;
• Extracting analytics on the interactions between people and mobile devices, in a statistical and anonymous way.

To do so, an application called i-Log has been developed to be installed on students' smartphones. It consists of two main components:
1. The first one collects sensor data unobtrusively, without requiring the user to interact with it. At the same time, the user is notified of the data collection process via a dedicated notification. Every smartphone has a given number of sensors (subset of the list below);

A. Sensors for smartphone movement (e.g., Accelerometer)
B. Environmental sensors (e.g., Temperature)
C. GPS
D. Wi-Fi networks in range
E. Rete Wi-Fi connected to the smartphone
F. Bluetooth devices in range
G. In and out calls (NO AUDIO)
H. In and out Sms (NO CONTENT)
I. Applications running (NO DATA FROM THE APPLICATIONS)
J. Screen status: ON/OFF
K. Battery Status: Charging/Not charging
L. % of battery
M. Whether the Flight mode is on
N. Whether the phone is on silent or normal

O. Earplugs in

P. Whether a song is being reproduced (TRUE/FALSE, NO DATA ABOUT THE SONG)

2. The second component administers a questionnaire similar to time use surveys with a 30 minutes interval between questions within the whole 24 hours of the day. Each question is composed by 3 sub-questions:

1. What are you doing (time of the question)?
    a. Lesson
    b. Study
    c. Eating
    d. Self care
    e. En route
    f. Social life
    g. Social media
    h. Cultural activity
    i. Sport
    j. Shopping
    k. Hobby
    l. Work
    m. House work
    n. Volunteering
    o. Other

2. Where are you (time of the question)?

    a. Home
    b. Outdoors
    c. Class
    d. Study hall
    e. Library
    f. Other university place
    g. Canteen
    h. Bar
    i. Other private locations
    j. Workplace
    k. Gym
    l. Shops
    m. Other

3. Who are you with (time of the question)?
    a. Alone
    b. With friends
    c. With roommate(s)
    d. Partner(s)
    e. Relative(s)
    f. Classmate(s)
    g. Other

The project will last two weeks: during the first one, students are asked to answer a time diary on their smartphone about their time use, while the application collects sensor data. During the second week they are only required to have the application running for collecting sensor data. ☐
.

Incentives
Students will receive a fixed money compensation, as an incentive to participate, with additional three final prizes assigned to random users that were considered appropriate based on three parameters:

1) How much data their smartphones records in via GPS, Bluetooth, and Wi-Fi. ☐

2) How many answers students will give;
3) How long they will keep the application running

The average to consider the students "appropriate" will be calculated based on the whole sample.

Some technical tips of the application and how to use i-Log:
- I-Log does not compromise the smarthone usability.
- I-Log consumes at most 7% of battery per hour, which is within normal behaviour.
- When the battery reaches 5%, i – Log will automatically turn off.
- i-Log is multilingual and chooses the language used on the smartphone.
- i-Log will start automatically when the smartphone is turned on. The data collection can be stopped at any by clicking the Stop button.
- While the experiment is going on, the users is always notified of its status via a smartphone notification. If the notification is not visible, then the collection is not going on.

What to do..
- If you turn off the phone, first stop I-log otherwise you will lose the last 30 minutes of data.
- Be sure to always have the following sensors on: Wi-Fi, GPS and Bluetooth.
- Due to privacy concerns, data are only downloaded on the server when you are connected with the Unitn Wi-Fi. Be sure to be connected to the Unitn-x network.

The questionnaire:
How it works:
- It is composed of 3 questions with a list of predefined answer: you can only give one answer.
- If any mistake occurs, just press the back button and you ca start from the first sub-question.
- Notification have no sound or vibration, but will simply appear on your phone.
- Questions will appear every 30 minutes.
- You can answer to the questions while they appear
What to do if you can't answer:
- Questions stack for a total of 5 questions. If so, answer starting from the oldest one, if possible.
- Questions will be on 24/7: when sleeping, of course, you cannot answer. This is because we wanted to be fair and avoid conditioning you behaviour.

-        In the morning, you will receive the questions that you could not answer, but don't worry. Just answer "self-care" to the last one, which we know covers also sleeping.

Important:
Data are collected, stored and used anonymously.
Every student will receive an ID to avoid linking it with his or her true identity. With this ID, application data will be merged with data provided by the University of Trento concerning academic performance.
All data provided are used under compliance with the art. 13 of D. Lgs. 30/06/2003 n. 196.

I have read this form:

_____   _____
Date and Place   Signature

ANNEX 2 - REQUEST OF AGREEMENT AND CONSENT TO THE EXPERIMENT FOR SMARTUNITN

By giving your consent to participate to the experiment, you authorize the University of Trento (Prof. Fausto Giunchiglia and Prof. Ivano Bison) to use your data in accordance with the privacy policy described below.

I authorize the use of my data:

☐Yes   ☐No

Acknowledgement of acceptance:

_____   _____
Date and place   Signed

Privacy policy

In compliance with the art.13 of the Italian legislative Decree no. 196 dated 30/06/2003, all personal data given will be treated with tools for the management of web services related to the University website for academic and research purposes.

Topics:

1. **Data controller**
2. **Place of data usage**
3. **Type of data used**
4. **Provision of data**
5. **Mode of use**
6. **Rights of the parties**

Data controller

The data controller is the University of Trento, with registered office in Via Calepina 14, 38122 Trento (TN).

Place of data use

The usage of data will mainly take place in the Department of Engineering and Computer Science and Department of Sociology and Social Research of the University of Trento.

Type of data used

Personal data voluntarily provided by the user upon registering (as part of the questionnaire) and data from smartphone sensors.

Provision of data

Registering an account on the application and sending one's own personal data imply giving consent to the use of data for the abovementioned purposes.

Mode of use

Personal data are used manually and/or via electronic devices for the time strictly necessary to reach the aims of their collection. Specific security measures have been followed to avoid an data loss, illegal or malicious usage, and unauthorized access.

Data collected may be disseminated anonymized and without any possible way of identifying the subjects.

Rights of the parties

The subject to which the data belong to have the right to receive confirmation of the existence or nonexistence of said data at any time, to know their origin and contents, to verify their accuracy or request their integration or update, or their amendment (art. 7 of D.Lgs. 196/03).

In compliance with the same article, the subject also has the right to ask for the cancellation, transformation in anonymized from or block of the data used unlawfully, in addition to oppose to their use, under reasonable motivations