Project Title:	**Sensing and predictive treatment of frailty and associated co-morbidities using advanced personalized models and advanced interventions**
Contract No:	690140
Instrument:	Collaborative Project
Call identifier:	H2020-PHC-2014-2015
Topic:	PHC-21-2015: Advancing active and healthy ageing with ICT: Early risk detection and intervention
Start of project:	1 January 2016
Duration:	36 months

Deliverable No: D4.7

Linguistic Corpus (vers a)

Due date of deliverable:	M18 (1st July 2017)
Actual submission date:	29th June 2017
Version:	1.0

Lead Author:	Christos Makris (UoP)
Lead partners:	(UoP)

Horizon 2020
European Union funding
for Research & Innovation

Change History

Ver.	Date	Status	Author (Beneficiary)	Description
0.1	01/05/2017	draft	Christos Makris, Andreas Kanavos (UoP)	Initial draft
0.2	01/06/2017	draft	Christos Makris, Andreas Kanavos (UoP)	First draft deliverable report
0.3	17/06/2017	draft	Vasilis Megalooikonomou (UoP)	Updated deliverable report sent for internal review.
0.8	23/06/2017	draft	Christos Makris, Andreas Kanavos (UoP)	Second draft deliverable report.
1.0	27/06/2017	final	Christos Makris, Andreas Kanavos (UoP)	Deliverable finalised taking into account internal review's comments.

EXECUTIVE SUMMARY

The deliverable “D4.7-Linguistic Corpus (version a)” is a document reporting on the social data collection phase. Specifically, in this phase, e-mails, Facebook posts and Twitter messages from several older people will be gathered and tagged according to each patient’s mental frailty condition. The linguistic corpus is focused on the Greek and French languages.

DOCUMENT INFORMATION

Contract Number:	H2020-PHC–690140	**Acronym:**	FRAILSAFE
Full title	Sensing and predictive treatment of frailty and associated co-morbidities using advanced personalized models and advanced interventions		
Project URL	http://frailsafe-project.eu/		
EU Project officer	Mr. Jan Komarek		

Deliverable number:	4.7	**Title:**	Linguistic Corpus (vers a)
Work package number:	4	**Title:**	Data Management and Analytics

Date of delivery	**Contractual**	01/07/2017 (M18)	**Actual**	29/06/2017
Status	Draft ☐		Final ☒	
Nature	Report ☒	Demonstrator ☐	Other ☐	
Dissemination Level	Public ☒	Consortium ☐		
Abstract (for dissemination)	This deliverable reports the first version of the Data Management Plan (DMP). It summarizes the data expected to be collected or generated during the FrailSafe lifecycle but also the specific measures to be adopted for each distinguished dataset. Finally, the deliverable illustrates the envisaged strategy to achieve open access to FrailSafe research data and results.			
Keywords	FrailSafe, Data Management Plan, DMP, open research data, frailty			

Contributing authors (beneficiaries)	Christos Makris, Andreas Kanavos (UoP)			
Responsible author(s)	Christos Makris		**Email**	makri@ceid.upatras.gr
	Beneficiary	UoP	**Phone**	+30 2610 996968

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1 Introduction

The deliverable “D4.7-Linguistic Corpus (version a)” is a document reporting on the social data collection phase.

In Task T4.4, we had to collect elderly data from social media, e.g. social media analytics, which are used to monitor and capture user’s behavior as well as the social interaction of the elderly. For example, we can consider as aspects that differentiate user behavior, the number of followers of a user, the number of contributions to the corresponding social network as well as the frequency of contributions. However, given the difficulties in collecting elderly data from their interaction with social media platforms, we have proposed a set of questions plus a Big Five questionnaire to correlate (with use of machine learning approaches) the word usage and the social media behavior of subjects to frailty symptoms thus providing an extra tool to the doctors to clarify the state of the subject. In addition, this questionnaire would provide more insight when building the profile of the elderly especially in the case of text analytics, where we are more than clear that this correlation actually exists.

The main idea behind the extra Big Five questionnaire was to connect frailty symptoms with the Big Five personality traits (i.e., Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness) and combined them with techniques that emotionally characterize elderly scripts. The plan was to collect data and plan to investigate techniques that connect text features and multiple values of the Big Five personality traits with symptoms of frailty. The training classification phase that will follow aims at predicting/characterizing frailty based on the written scripts (T4.5).

Specifically, we moved on two directions:

(i) to investigate the social interaction of the elderly and with our machine learning techniques, to locate his Big Five personality characterization. Then, based on the corresponding personality characterization, to relate this characterization to mental disorders trace frailty that can be easily captured by their connection to the two basic frailty symptoms of anxiety and depression;

(ii) combine machine learning techniques with linguistic and emotional tools for analysis such as the LingTester (T4.5). In following we aim to connect them, with use of a training classification phase, with frailty symptoms of the persons that wrote these passages. The trained classifiers can then be used to predict frailty based on the written scripts.

For the first direction, we employed a set of questionnaires and a set of crawlers that could help to download the social media interaction of the users. However the data collected from this phase though enough in size for the data collection phase had few information concerning the social media interaction, hence we enriched the data collection phase with data from anonymous users that could help overfit the LingTester built in phase 4.5.

Shown below, we present the technologies we used to build our crawlers and how we explored them plus a description of the data that we had collected.

2 Technologies

Platform and development tools

In this section we analyze the tools that were used to implement the crawling system; the tools are comprised of the **API** of **Twitter**, the **Python** programming language, and its libraries, and the file format and data format **JSON**.

Twitter API

The term API (Application Programming Interface) characterizes a set of procedures/functions, protocols, and tools that are to be used in order to create a software component or a complete application. All the well known social media have implemented and provide to the programmers an API that permits the programmers to build applications with them; in our case study we mainly focus in Facebook and Twitter, however, this holds for the other social media as well. In particular, Twitter provides two important APIs:

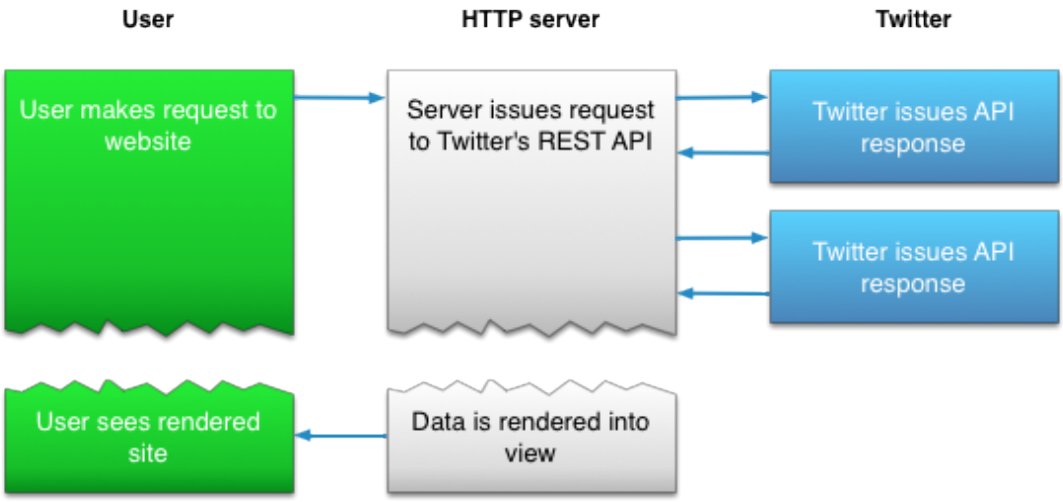
1. the REST API (https://dev.twitter.com/rest) and
2. the Streaming API (https://dev.twitter.com/streaming).

We should note that Twitter does not impose any limit concerning the amount of data that it provides through its APIs. However it sets restrictions concerning the time interval when that specific data is collected. These restrictions are applied for every programmer's account and refer to the permitted time interval of fifteen (15) minutes. Other APIs of social networks, as for example the **Graph API** of **Facebook** do not impose such kind of restrictions.

REST APIs

Using the REST APIs it is provided access to the programmer in order to read and write darta to Twitter. The programmer can retrieve information for a user using his tweets, his followers, and so he can crawl his movements in the social netwroks. The programmer first registers his application by employing the **OAth Protocol** (https://dev.twitter.com/oauth) and the provided credentials connecrted to every client and application, then the communicartion with the API takes place using HTTP requests and then we provide replies in the form of JSON files.

In the following figure we depict diagrammatically the interaction and the flow of information with the REST API.



Figue 1 - Employing and flowing of information with the REST API

Hence it is clear that REST API is suitable in cases that we want to perform specific searches related to a specific tweet or Twitter user.

Streaming APIs

With the Streaming API it is possible to provide to the programmer the real-time data, that is we can retrieve all tweets concerning a specific topic the time instant that they are published.

There is no limitation to this API, it is enough to establish a persistent connection in order to retrieve live data as long as the connection is active. In this case, also the data that are retrieved are provided in the form of JSON files. Finally, and comparing the two APIs when we use the streaming API, we do not need so many computational resources and the communication is more easy and without delays in comparison to REST API.

In the following figure, we depict diagrammatically the communication with the Streaming API.

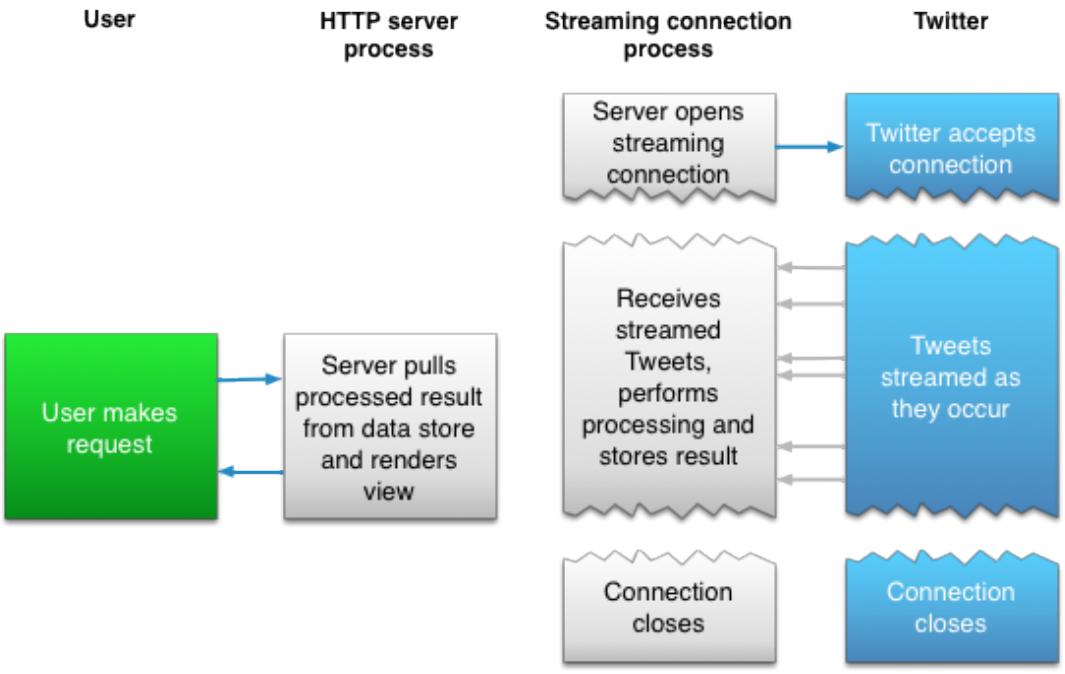


Figure 2 - Employing the Streaming API

Registration and authorization procedure for OAth

In both cases of using API it is important to firstly authorize and register the application that is to be communicated with the Twitter. In order to achieve this the programmer should have a valid account in Twitter and in the sequel creare an **App** (https://apps.twitter.com) and an **Access token** for the application.

After all these, in the sequel by using the: **Consumer key**, **Consumer secret**, **Access token**, **Access token secret** we can authorize and register the application using the protocol OAth.

Python

Python is a high-level programming language that has a basic characteristic that is more simple and easy in its usage and learning. The basic feature that distinguishes this language from other programming languages such as C++, and Java is that it has a lot of libraries that provide extra functionality and that it needs less number of code lines for implementing these functionalities.

Moreover, Python is a scripting programming language that is it uses interpreter instead of the compiler in order to translate the source code program. This means that the source code is first taken in real time to an interpreter, namely an intermediate software level which sequentuially reads and executes every line of source code. The version that is used is **Python 3.5.1**, one of the last versions of the language, in order to exploit the latest features that are added in the language. The interpreter specific to Python is the Python Virtual Machine (PVM).

The Python libraries which were used in the code of the social crawler in order to provide the desired functionality were pip, setuptools, tweepy, and matplotlib as presented below. Moreover, in order to embed and employ tweepy we used git, that does not constitute a part of Python but autonomous application instead. In the Ubuntu environment more of these tools are preinstalled, while in a typical Windows environment is necessary the use of cygwin or of a special development environment that provided capabilities analogue to that of the bash shell. In Windows 10 the basic bash shell functions have been integrated to the new standard Windows command line to a great degree, thus facilitating rapid application development.

Tweepy

Tweepy is a complete and integrated Python library whose objective is to connect and to communicate with Twitter. It provides to the end user of the host application an API encompassing the entire range of Twitter functionality. Moreover, the basic characteristic of Tweepy is that it is continuously renewed by following the changes and the new capabilities of the APIs of Twitter. It is remarkable that the specific tool is proposed by Twitter as a Python library in a relationship with the API programming interface.

The tool provides a lot of capabilities and solutions to the programmer, e.g. the partial handling of the limitations that are put on Twitter and the interaction with the APIs. Using the methods of this library, it is more easy to connect and the transactions with the REST APIs as also with the Streaming API in order to download the data that are needed from the Twitter social network during an encrypted session. Hence it is not needed from the programmer to send HTTP in order to communicate with the APIs, since they are all implemented in methods.

Tweepy still has to follow the main limitations of Twitter concerning the amount of information. Specifically, there is an upper limit of a hundred tweets and their associated information, such as likes and retweets, which can be retrieved every five minutes. This cap is enforced in order to prevent bursty requests. Such a traffic would place additional and irregularly patterned strain to Twitter servers and would create unfair access conditions for users and applications alike. To overcome this limitation, the main application code is enclosed in a loop whose termination conditions are related to the number of downloaded of tweets. This strategy has the additional benefit of keeping the secure communications channel with Twitter open because of the continuous traffic.

In order to properly and smoothly install tweepy, pip, git, and setuptools were installed prior to installing and using tweepy in order to satisfy its dependency conditions. Each software tool is analyzed thouroughly in the following sections.

pip

Pip is one of the most popular tools of package management in Python in combination with Anacodna and functions as an improved interface of the Python package index (PyPI). The main of its functionalities focus in the location, installation, and abstraction of Python packages. Moreover, it takes care in order to satisfy the interdependencies between packages and for this reason it installs multiple versions of the same package whereas this is necessary. It is compatible with all the existing versions of Python and in some cases there is a specialized version of pip in accordance to the version of the programming language. Moreover, pip cooperates with the package setuptools for managing the already installed packages in a more smooth way.

setuptools

The setuptools of Python constitute a collection of hlepful tools that are related with the already packages of Python. It functions autonomously but also in comparison with package managers with enhanced capabitlities such as pip and Anaconda. They are well known for the effective usage of network connections of low speed, and also for the fast restart of package installations that were interrupted. They periodically monitor the announcements in the digital archives for renewed package versions and they control if the new versions satisfy the interpedence criteria of already intalled packages.

matplotlib

The matplotlib is one of the most kwown scientific libraries of Python on which we base between others and a lot other libraries such as numpy and scipy in order to visualize results that are from scientific calculations, mainly composite visualization of multidimensional data. It is characterized from low memory usage for creation of high quality charts, with all the necessary parameterizations such as colors, width and kind of curves, representation of discrete and categorical data, and al the available metadata. It can easily transform the produced graphical represntations to al the well known image files.

JSON

The form of the data replies that are produced from the APIs of Twitter is JSON. It is a form all else standard that is used often as a way of reply to queries throught some API. It evolved from JavaScript, and this explains its abbreviation (JavaScript Object Notation), but it is considered indepedent from other programming languages, because there exist softaware programs written in it, for creation and analysing data in this form, to all the known and often used programming languages.

Graph API

Graph API is the standard programming interface of Facebook intended for external applications. It is currently consumed in all main programming languages as well as in some less known ones. This has contributed to the prolonged popularity of Facebook as a social medium, since from a closed platform it has evolved to the center of a rather vibrant and diverse ecosystem. As its name suggests, it eventually operates on the large social graph of Facebook, though its documentation and speculation among programmers indicate that data request from applications are filtered through certain intermediate stages. Moreover, Facebook has tigher constraints as to which and how much data are visible to the applications. For instance, an external program, even it is properly authenticated, can only access the Facebook wall of a page but not the wall of an account. This severly curbs the potential for building social analytics for accounts, most of which belong to individuals.

Nonetheless, there is plenty of room for retrieving and processing data related to the project. The application through the Graph API has access to the social graph, which comprises of the following data:

- Nodes (Users, Photos, Pages, Comments etc.)
- Edges (The links between nodes, e.g. between user Pages, the comment of an uploaded photo, etc.)
- Fields (The information that constitutes the nodes e.g. the name of the user)

Prior to accessing the social graph of Facebook, an authentication procedure takes place based on Facebook generated application credentials. Only a successful authentication provides access to reading and writing social data. The credentials are a set of unique large integers coded as strings used in order to securely communicate with the Graph API which.

- Permit calls to the Graph API.
- Locate Facebook pages
- Retrieve information

The authentication tokens also provide information for the application that created it, the application owner, and their expiration date. Currently the Facebook security scheme supports the following types of access tokens:

- User Access Token: it permits the capability of reading, modifying and writing to a user profile.
- App Access Token: it permits the reading and writing of the setting of an application.
- Page Access Token: it permits the reading, modification and writing of the data in a page.

3 FrailSafe Datasets

The questionnaire regarding social interaction has two types of questions. Initially, open type questions are given to the elderly so as to collect written text that will be used for the implementation of LingTester. For example, the participants can prepare some old texts they have, or write about a major enjoyable life event, such as wedding, child's birth, enjoyable travel experience, etc. Furthermore, one potential way to get some text is to show to the participants an attached picture and in following to ask them to describe it in written text.

With the utilization of the open type questions, we try to understand the interaction of the elderly with the social media in terms of general behavior.

In this task, we collect data from social media, e.g. social media analytics, which are used to monitor and capture user's behavior as well as the social interaction of the older people. We consider as aspects that differentiate user behavior, the number of followers of a user, the number of contributions to the corresponding social network as well as the frequency of contributions. Given the difficulties in collecting data in this age range from their interaction with social media platforms, we have proposed a set of questions plus a big five questionnaire to correlate (with use of machine learning approaches) the word usage and the social media behavior of subjects to frailty symptoms thus providing an extra tool to the doctors to determine the clinical status of the subject. In addition, this questionnaire would provide more insight when building the profile of the older people especially in the case of text analytics, where we are more than clear that this correlation actually exists.

More analytically, in order to collect information regarding social media, we proposed the Social Media Questionnaire. In this Social Media Questionnaire (see Appendix 1), we have formulated 36 questions. These questions measure the number of incoming/outgoing phone calls/sms, emails, use of social networks etc and other social and behavioural parameters (through linguistic analysis of text appearing in chat sessions or other write-ups, monitoring their location throughout the day) while respecting privacy and without becoming invasive. The outcome of the analysis of this plethora of data will be a formal and quantitative definition of a frailty metric that will be based on the aforementioned sensing dimensions. This questionnaire is administrated once, during or after the first clinical assessment in order to investigate the social interaction behavior of each participant.

The most widely known model of personality trait qualification is the Big Five. According to Big Five, the human personality is described as a vector of five values of traits. The combination of Big Five personality dimensions explains the dynamics of a personality. For example, a person may be very talkative (high Extraversion), not very tolerant and sensitive (low Agreeableness), systematic and punctual (high Conscientiousness), easily anxious (high Neuroticism) and extremely curious (high Openness).

In order to be recognized the traits Big Five model each person should answer in the following Likert scale, a questionnaire which includes 44 questions:

Strongly Disagree → 1

Disagree a little → 2

Neither agree nor disagree → 3

Agree a little → 4

Strongly Agree → 5

3.1 Data set description

Self-administrated questionnaires / text sampling

- **Social interaction**: 34-item self-administered questionnaire with both open and close questions.
- **Data collection of written texts**: Collection of previous typed or handwritten text, type or handwrite or dictate a major life event, type or handwrite or dictate the description of a standard picture.
- **Big five personality trait:** Big 5 test (self evaluating test).

The main idea is to measure social interaction of ageing people as well as social and behavioral parameters that emotionally characterize their scripts. The dataset was collected with use of questionnaires.

The plan was to collect data and to investigate techniques that connect text features and multiple values of the Big Five personality traits with symptoms of frailty. The training classification phase aims at predicting / characterizing frailty based on the written scripts.

The self-filled questionnaires are:

1. **Social interaction** (measured by the number of incoming/outgoing phone calls/sms, emails, use of social networks etc) and other social and behavioural parameters (through linguistic analysis of text appearing in chat sessions or other write-ups, monitoring their location throughout the day) while respecting privacy and without becoming invasive.
   The outcome of the analysis of this plethora of data will be a formal and quantitative definition of a frailty metric that will be based on the aforementioned sensing dimensions.
   This questionnaire is administrated once, during or after the first clinical assessment in order to investigate the social interaction behavior of each participant;
2. **Data collection of written texts** were given to the participants to fill in a second time, except if the participant is unable to write. More specifically, participants were asked for previous text, were asked to think of a major life event (prompts for life events are available such as weddings, child's birth, professional achievements etc), are asked to describe in written text an attached picture. The timing of collection of the written texts coincides with the clinical assessments.
   The natural language analysis tool, which was developed in terms of T4.5 will be able to detect signs of cognitive deficiencies in written text;
3. **Big five personality trait:** There are works that connect mental disorders with the Big Five personality traits and works that try to exploit this information employing social media.
   According to it, the human personality is described as a vector of 5 values of traits:
   i. **Openness:** This trait features characteristics such as curious, original, intellectual, creative and open to new ideas;
   ii. **Conscientiousness:** Common features of this dimension include organized, systematic, punctual, achievement, oriented and dependable;
   iii. **Extraversion:** This trait includes characteristics such as outgoing talkative, sociable and enjoying social situations;
   iv. **Agreeableness:** This personality dimension includes attributes such as affable, tolerant, sensitive, trusting, kind and warm;
   v. **Neuroticism:** Individuals high in this trait tend to experience anxious, temperamental and moody.

This personality characterization will be automatically extracted and taken into account in the frailty metric. In addition it will help in improving the services proposed by our approach to the elder according to his personality and his current emotional state.

The questionnaires that we used for the Big five personality trait are provided in the appendix. In particular for Grece and Cyprus we used the IPIP 50 item questionaires[1] that has been translated in greek[2][3] by Tsaousis, Bakola, Georgiadis. On the other hand the French questionnaire is a translation from the English[4] by John and Srivastava, published in 1999.

Moreover we implemented software for automatically recognizing the big five personality trait based on a methodology derived from study of F.Celli that use unsupervised learning techniques to recognize the Personality Traits for an individual using his/her texts. This implementation could be used in case of missing data from questionnaire. The implementation is in python 2.7 with external module of numpy[5] that is the fundamental package for scientific computing with Python .

Standards and format

The clinical web platform (including questionnaires web platform) of FrailSafe will be used. Once the data is transferred to the FrailSafe cloud facilities, it will be stored in appropriate databases.

Access policy / Dissemination level

Sensitive personal data will be handled only by the local clinical research personnel bound by local confidentiality rules. This data will not be transferred, merged or exchanged.

All other dataset's collected data will be anonymized and will contain no identifying information. Data will then pass to be used by the members of the FrailSafe consortium.

Data storage

The dataset as well as the answers to the questionnaires will be securely stored -via the clinical web platform- in the FrailSafe cloud facilities into appropriate databases. Anonymized data stored in the cloud will be encrypted.

The data will persist in the FrailSafe database at least throughout the duration of the project.

---

[1] http://ipip.ori.org/New_IPIP-50-item-scale.htm

[2] https://www.surveymonkey.com/s/NT5PC7N

[3] https://drive.google.com/file/d/0ByB685KehfWINGZ4aG1ndzYtX2c/edit

[4] http://fetzer.org/sites/default/files/images/stories/pdf/selfmeasures/Personality-BigFiveInventory.pdf

[5] http://www.numpy.org/

4 Crawlers

The dataset as well as the answers to the questionnaires will be securely stored -via the clinical web platform- in the FrailSafe cloud facilities into appropriate databases. Anonymized data stored in the cloud will be encrypted according to advanced encryption standards which are currently state of the art.

Social media crawlers are the dedicated software components which perform a search across the social media, in this specific case Facebook and Twitter, for information of interest such as wall posts in the former case and tweets in the latter case starting from a known account and traversing its friends or followers. Then this process is recursively applied to the accounts which were found during the previous step. This essentially amounts to a breadth first search until certain predetermined termination conditions are satisfied. Alternatively, a social crawler should be able to dynamically reconfigure itself according to the data it collects. The latter is necessary in order to follow events which happen in real-time.

Search patterns can be broadly divided to two categories. The first includes structural properties of the social graph such as the number of followers or the friends of an account, whereas the second consists of functional properties such as hashtags and mentions, both quite common actions across social networks. Although the social graph structure forms the backbone of any network, the latter are built precisely in order to provide functionality. Moreover, certain structural properties are implicit in the functional ones. Therefore, a carefully programmed social crawler must act according to a balanced search criterion. Topic sampling, namely the collection of accounts which reference a specific set of hashtags which is provided by the user. This search pattern is functional but also implies a certain degree of structural coherence as the crawler jumps to neighboring accounts.

Social crawlers run for extended time intervals, typically in the range of tens of hours for datasets currently considered large, rendering thus real-time human monitoring inefficient. Of course, besides this batch execution mode, a social crawler can be periodically executed in order for the dataset to be updated and expanded.

Termination criteria typically rely on the information coherence of the collected accounts using metrics such as:

- The Kullback-Leibler divergence
- The Aikaike information criterion
- The Bayesian information criterion
- The gini index
- The GEMINI metric

Moreover, there is an additional stop mechanism which is a function of the amount of the collected information. This is failsafe mechanism which prevents the unbounded retrieval of social information.

4.1 Twitter Crawler

The Twitter crawler (Table 1) was implemented in Python using the tweepy library which is free for academic purposes. In order to satisfy the software dependencies of tweepy, a number of Python packages were first installed. These packages include the pip package manager, the setuptools installation manager, and the matplotlib graphics package. Once those packages were installed, the tweepy repository was cloned from GitHub and then it was installed.

The next step was to obtain the four application authentication tokens required by Twitter. The first pair contains information about the application owner, whereas the second pair uniquely determines the Twitter application itself. All four abovementioned tokens are encoded as ordinary Python strings which facilitates their use. These tokens are generated and registered automatically by Twitter through the dedicated application dashboard.

Once the authentication tokens are obtained, they are hardcoded to the Twitter crawler. Then the method OAuthHandler is called to establish a secure connection. Since this procedure might time out for reasons related to the cryptographic protocol, the calls to this method are inside a try statement in order to ensure that the data retrieval segment of the crawler runs on valid data. Then follows an infinite while loop which asks for the Twitter screen name of the account whose data are to be retrieved. After that a second such loop asks for the number of tweets to be fetched. Finally, the crawl time in ISO format is displayed followed by the screen name, the username, the number of friends, and the tweets of the selected account.

From the above description is evident that the current crawler version is designed for batch processing and retrieves the information associated with a single Twitter account.

Table 1 - Flowchart of Twitter crawler

1: obtain access tokens from local file (4 tokens)
2: present tokens to Twitter authorization server
3: read account name n
4: read number of tweets p
5: obtain recent information about n in JSON format
6: **while** p tweets **not** read
7: parse JSON and read last tweet t
8: store t in a Python list and remove it from JSON
9: **end while**

4.2 Facebook Crawler

The Facebook crawler (Table 2) is also implemented in Python using the BeautifulSoup library which is free of charge for academic purposes. BeautifulSoup contains various methods for obtaining the data of Facebook pages by anchoring to a specific account node of the Facebook social graph as well as parsers for HTML and XML documents.

The setup of BeautifulSoup varies according to the library version. For versions numbers less than or equal to 3 there is an independent software package which can be installed at the operation system level, while for versions greater than 4 a Python package manager such as pip or easy_install is required.

Similarly to the Twitter case, appropriate authentication tokens must also be obtained from Facebook before any access is granted to the crawler. Those tokens come in the form of four strings which are subsequently hardcoded into the source code as their interactive insertion is difficult.

Initially the Facebook crawler presents a text prompt page to the end user where the page name is manually inserted. Then a similar second prompt asks for the number of public posts which are subsequently stored to a list. Both prompts are enclosed in their own separate loops which ensure that valid data are inserted. In order to retrieve the public posts, a secure connection to Facebook is established using the OAuth open protocol.

It should be noted that, due to privacy concerns, only public posts from Facebook pages can be obtained. The public information of any kind posted by the various Facebook accounts can be retrieved only after the explicit permission of this account is given to the application. Moreover, this permission should be obtained each time the application attempts to retrieve data posted by an account.

Table 2 - Flowchart of Facebook crawler

1: obtain access tokens from local file (4 tokens)

2: present tokens to Facebook graph management server

3: read page name n

4: read number of public posts p

5: **while** p posts **not** read

6: call BeautifulSoup to obtain another post m

7: store m in a Python list

8: **end while**

5 Datasets by the numbers

Until now there are data available from Greece (UoP), Cyprus (MATERIA), and France (NANCY) for the two out of the three participant groups, which are:

- Group A: Participants 1-80
- Group B: Participants 81-120

5.1 Big Five

We have conducted detailed experiments for the Big Five Personality Traits Extraction for Greek, Cypriots and French participants. The questionnaires of B5 are transferred to .csv in format (id_qi, Ai) where id_qi is the id of question i and Ai is the score of question i. The name of each .csv is the id of the subject.

Each trait is calculated by the average of the score (scale 1-5) or reverse score (6 – score) of certain questions in the questionnaire as it is implemented in our code. The output of the code has stored in folder PT where each subject has a .csv file that contains the Personality Traits for each individual and the .csv file (ALL_PT.csv) in format Subject,Extraversion,Agreeableness,Conscientiousness,Neuroticism,Openness with all results.

The implementation is in python 2.7 with external module of numpy.

The number of participants that we recognized was 119, 92, 74 for Greek, Cypriots and French cases. The cases with missing values in the questionnaire were not examined.

5.2 New and Previous Texts

In this subsection, data from their social media interaction (Twitter, Facebook and e-mail) are collected based on the social media questionnaire. Specifically, we asked those users that are active in social media and have e-mail accounts to provide the appropriate information. We have also provided users with written text describing an image and an important life event.

Utilising eCRF API, we were able to retrieve all available raw data, stored by each medical team, containing detailed answers to the questionnaire, along with uploaded files of present and past text.

The corresponding data are till now the following:

Greece (UoP): 120 Users with 248 Texts where

- Image: 118 Texts
- Important Life Event: 120 Texts
- Past Written Text: 10 Texts

Cyprus (MATERIA): 75 Users with 138 Texts

- Image: 61 Texts
- Important Life Event: 54 Texts
- Past Written Text: 8 Texts
- Facebook Posts: 15 Texts

France (NANCY): 86 Users with 262 Texts

- Image: 39 Texts
- Present Text: 73 Texts
- Past Written Text: 150 Texts

Specifically, in the following tables (Table 3, Table 4, Table 5), the total number of data per user is presented.

Table 3 - Greece (UoP) results per user

Subject	Image	Important Life Event	Past Written Text
1001	1	1	1
1002	1	1	1
1003	1	1	1
1005	1	1	
1006	1	1	1
1007	1	1	1
1008	1	1	
1009	1	1	
1010	1	1	
1012	1	1	
1013	1	1	
1014	1	1	
1015	1	1	
1016	1	1	
1017	1	1	
1018	1	1	
1019	1	1	
1020	1	1	
1021	1	1	
1022	1	1	
1023	1	1	
1024		1	
1025		1	
1027	1	1	
1029	1	1	

1030	1	1	
1031	1	1	
1032	1	1	
1033	1	1	
1034	1	1	
1035	1	1	
1036	1	1	
1037	1	1	
1038	1	1	
1039	1	1	
1040	1	1	
1041	1	1	
1042	1	1	
1043	1	1	
1044	1	1	
1045	1	1	
1046	1	1	
1047	1	1	
1048	1	1	
1049	1	1	
1050	1	1	
1051	1	1	
1052	1	1	
1053	1	1	
1054	1	1	
1055	1	1	
1056	1	1	

1057	1	1	
1058	1	1	
1059	1	1	
1060	1	1	
1061	1	1	
1062	1	1	
1063	1	1	
1064	1	1	
1065	1	1	
1066	1	1	
1067	1	1	
1068	1	1	
1069	1	1	
1070	1	1	
1072	1	1	
1073	1	1	
1074	1	1	
1075	1	1	
1076	1	1	
1077	1	1	
1078	1	1	
1079	1	1	
1080	1	1	
1081	1	1	
1082	1	1	
1083	1	1	
1084	1	1	

1085	1	1	
1086	1	1	
1087	1	1	
1088	1	1	
1089	1	1	
1090	1	1	
1091	1	1	
1092	1	1	
1093	1	1	
1094	1	1	
1095	1	1	
1096	1	1	1
1097	1	1	
1098	1	1	
1099	1	1	
1100	1	1	
1101	1	1	
1102	1	1	1
1103	1	1	1
1104	1	1	
1105	1	1	
1106	1	1	
1109	1	1	
1110	1	1	1
1111	1	1	
1112	1	1	
1113	1	1	

1114	1	1	1
1115	1	1	
1116	1	1	
1117	1	1	
1118	1	1	
1119	1	1	
1120	1	1	
1505	1	1	
1508	1	1	
1518	1	1	
1526	1	1	
1529	1	1	
1548	1	1	
1558	1	1	

Table 4 - Cyprus (MATERIA) results per user

Subject	Image	Important Life Event	Past Written Text	Facebook Posts
2001	1			
2003		1		
2004	1	1		
2005	1	1		
2006	1	1		
2007	1			
2008		1		
2013	1	1	1	
2015	1	1	1	
2016	1	1		
2018	1	1		
2026	1	1	1	
2029	1	1		10
2031	1			
2032		1	2	
2036	1	1	1	
2048			1	
2049	1	1	1	
2050	1			
2051	1	1		
2052	1	1		
2053	1	1		
2054	1	1		
2055	1	1		
2056	1	1		

2057	1	1		
2058				5
2060	1	1		
2063	1	1		
2071	1	1		
2072	1	1		
2073	1	1		
2074	1	1		
2075	1	1		
2076	1	1		
2077	1	1		
2078	1	1		
2080	1	1		
2081	1	1		
2082	1	1		
2083	1	1		
2084	1	1		
2085	1	1		
2086	1	1		
2087	1			
2088	1			
2089		1		
2090	1			
2091	1			
2092	1			
2093	1	1		
2094	1	1		

2095	1	1		
2097	1	1		
2098	1	1		
2099	1	1		
2100	1			
2101		1		
2102	1			
2104		1		
2105	1			
2107		1		
2108		1		
2109		1		
2110	1			
2111	1			
2112		1		
2113		1		
2114	1			
2115	1			
2117	1			
2118	1			
2119		1		
2518	1			
2541	1	1		

Table 5 - France (NANCY) results per user

Subject	Image	Important Life Event	Past Written Text
3002	1		1
3003		1	1
3004		1	1
3005		1	6
3006	1	3	2
3007		2	1
3008		2	1
3009		2	3
3010	1		
3011		1	1
3012	1	1	
3013		1	
3014		1	
3016		1	1
3017		1	
3019	1	2	1
3025		1	1
3026		1	1
3028		1	
3029		1	1
3030	1	1	1
3031	1	1	
3032		2	2
3033	1	1	
3035	1	1	3

3036	1	1	1
3037			1
3038			1
3039		1	
3040			5
3041			1
3042	1	1	1
3043	1	1	1
3044	1	1	1
3045			1
3046			1
3047	1	1	2
3048	1	1	
3049			1
3051	1		
3052	1	1	
3053	1	1	2
3055		1	
3056		1	1
3057			1
3058	1	1	1
3063			66
3081		1	
3082		1	
3083	1		
3084	1	1	
3085		1	1

3086	1		1
3087			2
3088	1		1
3089			1
3090	1	1	1
3091		1	
3095		1	2
3096	1	1	3
3097	1	1	
3098		1	2
3099	1	1	2
3100	1	1	1
3101		1	1
3102		1	1
3103	1		1
3104		1	
3105	1		
3106		1	1
3107	1	1	1
3108		1	
3109	1	1	1
3110	1		1
3112		1	
3113		1	1
3114	1	1	
3115		1	1
3116	1	1	1

3117		1	
3118	1	1	
3119		1	1
3120	1	1	4
3518		1	
3592	1	1	1
3611	1	1	1

5.3 Twitter Data

English: 30756 Tweets from 9917 Users

- 4 Users with 5 Tweets
- 997 Users with 4 Tweets
- 8916 Users with 3 Tweets

French: 18656 Tweets from 8531 Users

- 1594 Users with 3 Tweets
- 6937 Users with 2 Tweets

Greek: 8090 Tweets from 5717 Users

- 2372 Users with 2 Tweets
- 3344 Users with 1 Tweet

References

[1] European Commission, “Guidelines on Data Management in Horizon 2020”, version 2.1: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Last updated: 15 February 2016.

[2] European Parliament and Council, 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”: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en

[3] Celli, Fabio. "Unsupervised personality recognition for social network sites."*Proc. of Sixth International Conference on Digital Society*. 2012.

Appendix 1 - The Social Media Questionnaire

Given to the participant at the end to fill in at his/her home [Leave blank in database all unanswered questions]

Questionnaire for usage of the internet and social media

- **This questionnaire is given and answered in a second time**
- **Administrated once during or after the first clinical assessment**

Date administrated: ___/___/_____ (dd/mm/yyyy)

1. Which means do you use to be kept updated and informed? (more than one answer)

1. Television
2. Newspapers/magazines
3. Family/friends
4. Internet
5. Other……

2. Do you use the internet?

[ ] No=1

[ ] Yes =2

If the answer is no, stop here

3. Do you consider yourself to be familiar internet user ?

[ ] beginner =1

[ ] less familiar =2

[ ] very familiar=3

4. Which device do you usually use to connect to the internet? (more than one answer)

[ ] Computer/laptop 1=No, 2=Yes

[ ] Tablet 1=No, 2=Yes

[ ] Mobile Phone 1=No, 2=Yes

5. How often do you connect to the internet per week? (the answer can be from 1 up to 7)

[ ] times per week

6. How many hours per day do you usually use the internet?

[ ] Hours per day

7. Which internet services do you usually use? (more than one answer)

[ ] News/ Update/Information 1=No, 2=Yes

[ ] Communication, Social media 1=No, 2=Yes

[ ] Entertainment (games, music, TV, video) 1=No, 2=Yes

[ ] Online Transactions 1=No, 2=Yes

[ ] Other:

.....

8. Describe in a few words your 'internet' activity and the changes it has have brought upon your life

.....

.....

.....

.....

9. Do you use any social media (i.e. facebook etc)?

[ ] No, I have never used them

[ ] No, i do not use them but i used to =1

[ ] Yes I use the social media =2

Why did you stop;

……

.

If the answer is no, stop the here

10. How long have you been using social media?

I have used social media for [ ] months

11. Which of below social media you use? (more than one answer)

[ ] Facebook 1=No, 2=Yes

[ ] Twitter 1=No, 2=Yes

[ ] YouTube 1=No, 2=Yes

[ ] Instagram 1=No, 2=Yes

[ ] Personal blog 1=No, 2=Yes

[ ] Other: .....

12. How often do you use social media per week? (the answer can be 1 up to 7)

[ ] times per week

13. When you use social media, how many hours per day do you usually use them for?

[ ] hours per day

14. What made you use social media for the first time?

.....
.....
.....
.....
.....

15. Do you think social media are easy to use?

[ ] Very easy =1

[ ] Easy =2

[ ] Difficult =3

[ ] very difficult =4

16. Do you consider yourself a familiar user of social media ?

17. The numbering in this question is the opposite of the same question for internet use where beginner is 1 and very familiar is 3. Confusing.

[ ] Very familiar =1

[ ] Less familiar =2

[ ] Beginner =3

18. Which of the above information is included in your profile?

(more than one answer)

[ ] Real name 1=No, 2=Yes

[ ] e-mail 1=No, 2=Yes

[ ] telephone 1=No, 2=Yes

[ ] House Location 1=No, 2=Yes

[ ] Photographs 1=No, 2=Yes

[ ] Video 1=No, 2=Yes

[ ] Religion 1=No, 2=Yes

[ ] Interests 1=No, 2=Yes

[ ] Other:.....

19. Does the information you provide on social medial represent reality and why?

[ ] No = 1

[ ] Yes = 2

Why:

.....

.....

.....

20. Fill in only if you use twitter, or otherwise go to question 23

How many followers you have on twitter?

[ ] (fill in a number)

21. How many people do you follow on twitter?

[ ] (fill in a number)

22. How often do you tweet?

[ ] (fill in a number)

23. Fill in only if you use facebook, otherwise go to question 24

24. How many friends/contacts do you have on facebook?

[ ](complete number)

25. How many of your Facebook friends do you consider your true friends from all your friends/contacts?

[ ] Only a few =1

[ ]many of them =3

[ ] most of them =4

[ ] everyone=5

26. Do you accept friend requests from strangers at your social media accounts?

[ ] never =1

[ ] sometimes =2

[ ] always =3

27. What do you usually do during your social media visits?

(Fill in with numbers by beginning with 1 from the most frequent to the least frequent activity)

[ ] post

[ ] share

[ ] like

[ ] comment

[ ] share a photo

[ ] share a link

[ ] share video/music

[ ] other:.....

28. Do you follow politicians/organizations on social media?

[ ] No =1

[ ] Yes=2

29. Do you believe that communication between politicians and voters through social media is important?

[ ] Strongly agree=1

[ ] Agree=2

[ ] Disagree=3

[ ] Strongly disagree=4

30. What type of pages you follow at social media?

.....
.....
.....

31. Do you believe that social media affect your social life

[ ] very positively =1

[ ] positively =2

[ ] Do not affect at all =3

[ ] Negatively =4

[ ] Very negatively =5

32. Besides the activity of yourself and of your contacts, is there anything else that draws your attention in social media? (i.e. Advertisements, offers etc)? How do you respond to this? Does it affect your judgment in some degree?

.....
.....
.....
.....

33. Do you think that there is a danger for the safety of your personal data in social media?

[ ] Strongly agree =1

[ ] agree=2

[ ] disagree=3

[ ] strongly disagree =4

[ ] I don't know/doesn't concern me=5

34. Do you believe that the privacy policies of the social media that you are using are effective?

[ ] Strongly agree=1

[ ] I agree=2

[ ] i dont know/it doesnt concern me=3

35. Are you aware of who can check your profile and the information it contains in the social media you are using?

[ ] No=1

[ ] yes=2

[ ] I don't know / it doesnt concern me=3

36. Have you changed your security settings in social media in order to protect your personal data?

[ ] No=1

[ ] Yes=2

Appendix 2 - The Big five questionnaire - for French partner

(http://www.outofservice.com/bigfive/, https://www.ocf.berkeley.edu/~johnlab/bfi.htm)

In order to be recognized the traits Big Five model each person should answer in the following Likert scale a questionnaire which includes 44 questions.

Strongly Disagree → 1
Disagree a little → 2
Neither agree nor disagree → 3
Agree a little → 4
Strongly Agree → 5

The questions are the following:

I see myself as someone who …

1. Is talkative
2. Tends to find fault with others
3. Does a thorough job
4. Is depressed, blue
5. Is original, comes up with new ideas
6. Is reserved
7. Is helpful and unselfish with others
8. Can be somewhat careless
9. Is relaxed, handles stress well
10. Is curious about many different things
11. Is full of energy
12. Starts quarrels with others
13. Is a reliable worker
14. Can be tense
15. Is ingenious, a deep thinker
16. Generates a lot of enthusiasm
17. Has a forgiving nature
18. Tends to be disorganized
19. Worries a lot
20. Has an active imagination

21. Tends to be quiet
22. Is generally trusting
23. Tends to be lazy
24. Is emotionally stable, not easily upset
25. Is inventive
26. Has an assertive personality
27. Can be cold and aloof
28. Perseveres until the task is finished
29. Can be moody
30. Values artistic, aesthetic experiences
31. Is sometimes shy, inhibited
32. Is considerate and kind to almost everyone
33. Does things efficiently
34. Remains calm in tense situations
35. Prefers work that is routine
36. Is outgoing, sociable
37. Is sometimes rude to others
38. Makes plans and follows through with them
39. Gets nervous easily
40. Likes to reflect, play with ideas
41. Has few artistic interests
42. Likes to cooperate with others
43. Is easily distracted
44. Is sophisticated in art, music, or literature

Appendix 3 - The Big five questionnaire - for Greece and Cyprus partner

(http://ipip.ori.org/New_IPIP-50-item-scale.htm)

In order to be recognized the traits Big Five model each person should answer in the following Likert scale a questionnaire which includes 50 questions.

Very Inaccurate → 1
Moderately Inaccurate → 2
Neither Accurate Nor Inaccurate → 3
Moderately Accurate → 4
Very Accurate → 5

The questions are the following:

I see myself as someone who …

1. Am the life of the party.
2. Feel little concern for others.
3. Am always prepared.
4. Get stressed out easily.
5. Have a rich vocabulary.
6. Don't talk a lot.
7. Am interested in people.
8. Leave my belongings around.
9. Am relaxed most of the time.
10. Have difficulty understanding abstract ideas.
11. Feel comfortable around people.
12. Insult people.
13. Pay a ttention to details.
14. Worry about things.
15. Have a vivid imagination.
16. Keep in the background.
17. Sympathize with others' feelings.
18. Make a mess of things.
19. Seldom feel blue.
20. Am not interested in abstract ideas.
21. Start conversations.
22. Am not interested in other people's problems.
23. Get chores done right away.
24. Am easily disturbed.
25. Have excellent ideas.
26. Have little to say.
27. Have a soft heart.
28. Often forget to put things back in their proper place.
29. Get upset easily.
30. Do not have a good imagination.
31. Talk to a lot of different people at parties.
32. Am not really interested in others.
33. Like order.
34. Change my mood a lot.
35. Am quick to understand things.
36. Don't like to draw attention to myself.
37. Take time out for others.

38. Shirk my duties.
39. Have frequent mood swings.
40. Use difficult words.
41. Don't mind being the center of attention.
42. Feel others' emotions.
43. Follow a schedule.
44. Get irritated easily.
45. Spend time reflecting on things.
46. Am quiet around strangers.
47. Make people feel at ease.
48. Am exacting in my work.
49. Often feel blue.
50. Am full of ideas.

Appendix 4 - The Big five questionnaire - Personality Traits Recognition

Greek Participants

Subject	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness
1001	3.9	4.7	4.1	3.4	4.1
1002	2.8	4.5	3.7	2.6	4.2
1003	4.0	4.6	3.8	2.5	4.1
1004	2.6	3.7	2.7	2.2	3.0
1005	2.3	4.8	3.6	2.3	4.1
1006	4.3	3.9	3.3	2.4	4.1
1007	4.1	4.6	4.3	3.3	3.2
1009	3.5	4.5	2.6	2.6	3.2
1010	2.4	4.5	3.4	2.0	3.8
1011	1.8	4.0	2.4	1.8	2.8
1012	4.2	4.5	4.2	3.2	4.4
1013	3.7	3.9	3.2	2.2	3.3
1014	3.6	4.4	3.5	3.5	3.5
1015	3.4	4.4	3.5	3.4	3.5
1016	3.0	4.5	3.3	2.9	3.8
1017	3.2	4.0	4.5	3.5	3.5
1018	2.8	4.1	3.7	3.9	4.0
1019	3.3	3.5	2.6	2.4	3.2
1020	2.5	3.6	2.8	3.0	3.3
1021	1.8	3.6	2.8	2.6	1.4
1022	3.1	2.4	3.8	2.6	2.9
1023	2.3	3.1	3.3	1.8	4.0
1024	2.7	4.6	3.4	2.7	3.4
1025	2.3	3.9	2.5	2.7	2.6
1026	4.0	2.0	2.8	4.8	3.9
1027	3.1	3.7	2.1	3.3	2.8
1028	3.2	3.2	2.9	3.0	2.6
1029	2.2	4.6	4.7	2.0	3.5
1030	3.2	4.2	3.5	3.6	3.1
1031	2.2	3.7	2.0	3.6	2.7
1032	2.5	2.1	1.3	2.4	4.3
1033	3.4	4.1	4.8	3.3	3.9
1034	2.5	3.9	3.0	3.2	2.4
1035	4.0	4.1	3.5	4.3	2.8
1036	3.0	3.4	3.7	2.2	3.6
1037	3.1	3.0	3.2	3.3	2.6
1038	4.0	4.3	4.1	2.8	3.3
1039	2.4	4.2	3.8	3.3	2.6
1040	3.2	4.1	4.2	2.8	3.6

1041	2.8	4.2	4.0	3.4	2.9
1042	3.7	4.5	4.7	3.8	4.5
1043	3.0	4.3	4.2	2.4	2.8
1044	3.1	4.2	4.1	3.3	3.4
1045	3.8	4.1	4.1	4.0	4.2
1046	3.2	4.4	3.5	4.0	2.9
1047	2.9	4.0	3.6	3.7	3.8
1048	2.6	4.1	3.4	3.2	3.1
1049	1.5	3.3	1.9	2.5	2.0
1050	3.6	4.4	4.0	2.4	3.1
1051	1.6	4.3	2.8	2.4	2.9
1052	3.9	4.4	4.3	3.3	3.5
1053	2.9	3.5	4.4	4.0	4.1
1054	3.1	4.6	3.7	1.8	3.2
1055	2.1	4.8	3.9	1.9	3.5
1056	2.6	4.5	4.2	3.1	2.7
1057	1.5	2.6	2.5	1.7	1.9
1058	2.3	2.6	3.6	2.6	2.5
1059	2.9	4.6	4.5	4.6	4.2
1060	4.5	4.4	3.7	3.1	4.0
1061	3.1	3.6	4.1	1.4	3.6
1062	4.0	4.2	2.7	3.2	2.7
1063	3.0	3.7	3.7	3.5	3.0
1064	2.8	3.2	3.3	2.7	2.5
1065	1.8	3.5	3.2	2.8	3.1
1066	1.9	2.5	1.7	1.5	1.6
1067	4.0	4.7	4.1	3.4	3.3
1068	3.0	3.5	3.0	3.0	3.0
1069	2.3	3.3	3.0	3.7	2.9
1070	2.4	2.9	2.3	1.9	3.2
1071	3.0	4.4	2.9	3.4	3.0
1072	3.9	4.4	4.3	3.0	3.6
1073	1.9	3.2	2.0	2.3	2.2
1074	2.2	3.0	2.3	1.8	1.9
1075	3.2	4.5	4.2	3.1	2.5
1076	2.6	3.0	3.1	2.0	2.8
1077	3.2	2.5	2.2	2.4	3.3
1078	2.6	3.0	3.1	1.9	3.0
1079	3.2	4.8	4.3	3.5	2.6
1080	2.0	3.5	3.5	3.0	1.8
1081	3.2	4.2	3.0	3.2	3.0
1082	3.3	3.8	2.1	2.9	2.5
1083	2.7	3.1	2.8	3.1	2.3
1084	2.5	3.0	4.3	3.8	3.8
1085	4.2	4.7	4.4	3.3	3.4
1086	3.3	3.6	4.0	2.4	3.7

1087	3.3	3.7	4.7	1.9	4.0
1088	2.9	2.8	2.7	1.4	2.2
1089	2.6	4.6	4.4	2.6	3.5
1090	2.4	3.8	2.7	2.2	1.9
1091	2.9	4.2	4.1	2.9	2.9
1092	2.6	4.6	4.4	2.6	3.5
1093	4.1	4.2	4.6	3.0	3.0
1094	2.2	4.7	3.3	2.5	2.0
1095	4.0	3.6	4.4	4.8	3.9
1096	4.2	4.1	4.7	4.9	3.8
1097	4.4	4.7	4.8	4.6	4.3
1098	4.0	3.5	3.7	3.8	3.4
1099	3.3	3.7	3.4	3.4	2.4
1100	2.9	2.5	3.0	3.0	2.3
1101	4.2	4.3	4.3	4.5	3.3
1102	4.0	4.5	4.2	4.2	2.9
1103	3.5	4.2	4.6	4.5	3.2
1104	4.3	4.2	4.0	4.0	3.2
1105	2.6	3.2	3.2	3.0	2.2
1106	3.4	3.8	2.7	3.9	2.0
1107	4.3	4.4	4.4	4.3	4.2
1108	4.3	4.6	4.5	4.6	3.2
1109	1.9	3.3	3.7	2.2	2.9
1110	4.4	3.4	3.7	3.9	3.9
1111	4.4	4.6	4.7	4.6	4.0
1112	2.1	4.3	4.1	3.1	2.6
1113	1.8	4.9	4.6	3.6	3.4
1114	4.6	4.6	4.3	3.9	2.6
1115	2.5	4.8	3.9	2.9	2.8
1116	4.2	4.9	4.7	3.1	3.4
1117	3.5	4.9	4.4	3.2	3.3
1118	4.6	4.8	4.5	3.4	4.1
1119	2.0	4.6	4.3	3.1	3.0
1120	1.8	4.5	4.1	2.8	2.9

Cyprus Participants

Subject	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness
2001	2.63	3.44	3.0	3.38	3.3
2002	2.88	3.11	2.89	3.13	3.0
2003	2.75	3.33	3.22	2.75	3.1
2004	2.88	3.0	3.0	3.5	3.1
2005	3.13	3.33	3.67	2.75	2.8

2006	2.3	3.6	3.1	3.4	3.1
2007	3.63	3.67	3.22	2.5	3.4
2008	2.9	4.1	3.7	2.2	3.9
2009	3.38	3.22	3.56	2.75	3.2
2011	2.5	4.0	3.78	2.75	3.3
2012	3.13	3.44	3.22	3.13	3.2
2013	2.8	3.9	3.0	3.1	2.9
2014	3.0	2.67	2.78	3.0	2.9
2016	2.88	2.89	3.33	3.5	2.8
2017	3.38	3.11	3.33	2.88	3.1
2020	3.0	3.0	3.67	3.38	3.3
2021	2.63	3.67	3.22	2.5	2.8
2022	4.3	4.5	4.0	1.8	4.8
2023	3.63	3.0	2.44	3.0	2.9
2024	3.25	2.78	3.0	2.88	3.4
2027	3.88	3.0	2.89	3.13	3.2
2028	3.5	2.56	2.89	2.63	3.0
2029	3.5	3.9	3.4	3.7	3.4
2030	4.13	3.89	3.0	3.0	3.1
2032	3.75	2.89	2.56	1.88	2.9
2034	2.75	3.22	2.56	2.88	2.6
2035	2.88	3.11	3.11	2.63	3.4
2036	4.2	4.3	4.5	2.0	4.0
2037	3.1	3.1	3.0	3.6	3.4
2038	2.88	3.22	3.67	3.38	2.9
2039	3.5	3.11	3.0	2.88	2.9
2040	3.13	3.0	2.33	2.75	3.2
2041	3.25	3.44	2.44	2.38	3.4
2042	2.63	3.0	3.22	2.13	3.0
2043	3.38	2.89	2.78	3.25	3.4
2044	3.4	3.9	4.1	3.7	3.9
2045	3.1	4.6	4.6	4.3	4.6
2046	3.5	4.2	5.0	3.9	4.0
2047	2.6	4.7	4.5	2.9	3.9
2048	3.63	3.22	2.56	3.13	3.1
2049	3.1	4.6	4.1	2.3	3.5
2050	3.1	3.6	3.3	1.3	3.6
2051	3.63	2.67	3.67	3.25	3.3
2052	2.9	3.2	3.1	2.6	3.4
2053	3.3	4.0	2.9	3.0	3.2
2054	4.2	4.0	4.2	3.2	4.5
2055	2.75	3.11	2.78	2.63	2.7
2056	2.8	3.9	3.0	3.1	2.9
2057	3.38	3.11	4.11	2.5	2.8
2060	3.63	3.44	3.89	2.13	3.4
2061	2.7	4.9	4.6	2.2	4.0

2062	3.25	2.56	3.11	3.0	3.5
2063	2.0	4.3	4.9	1.7	3.4
2069	2.88	3.22	3.33	2.75	3.5
2071	2.4	3.1	3.8	3.0	3.4
2072	3.2	4.9	4.6	4.5	3.8
2073	3.9	4.9	4.9	3.9	3.5
2074	2.6	4.3	5.0	1.8	3.9
2075	2.9	4.6	4.9	3.2	3.4
2076	3.13	3.44	3.11	2.88	2.9
2077	2.0	3.0	4.1	3.5	3.5
2078	1.6	4.1	4.5	3.1	2.8
2079	3.0	3.33	3.22	3.38	3.6
2080	2.0	4.6	4.9	2.7	2.9
2082	4.0	4.2	4.6	2.6	3.8
2083	2.1	3.7	4.6	2.6	3.4
2084	3.5	3.11	3.67	3.0	3.7
2085	3.63	4.0	3.78	2.5	3.3
2085	3.4	4.2	3.7	3.2	4.0
2086	3.3	4.4	4.2	3.3	4.0
2087	3.25	3.0	2.67	2.88	3.2
2089	3.0	4.22	4.0	2.13	3.3
2093	2.5	2.33	2.22	3.38	3.9
2094	2.0	2.78	3.22	3.5	3.4
2095	3.38	2.44	3.44	3.0	2.4
2096	3.5	2.78	3.67	3.63	2.4
2097	2.88	3.67	2.33	2.5	3.9
2098	4.5	3.22	3.56	3.0	3.4
2099	3.9	4.2	4.6	2.3	3.8
2100	2.7	3.4	4.2	1.8	3.3
2101	4.0	3.67	3.11	3.0	3.2
2102	3.0	1.33	2.78	3.13	3.9
2107	3.5	3.22	3.56	2.63	2.9
2110	3.63	2.78	4.0	3.13	3.2
2111	3.63	2.56	3.11	2.63	2.8
2116	3.13	3.0	2.67	3.0	3.1
2117	3.0	3.0	2.89	2.88	3.0
2218	2.88	3.44	2.44	2.63	2.9
2119	2.5	3.22	3.0	3.25	3.0
2220	3.0	3.11	2.89	2.38	2.6
2518	2.88	3.44	2.44	2.63	2.9
2541	3.38	3.67	3.22	2.75	3.2

France Participants

Subject	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness
3001	3.75	4.67	3.44	1.5	2.4
3002	2.0	3.78	2.78	2.5	1.6
3003	2.88	4.44	3.33	2.63	3.3
3004	2.38	4.22	3.89	3.0	3.1
3005	3.25	4.0	3.56	2.0	2.5
3006	3.88	3.78	4.44	1.38	4.2
3007	2.38	3.22	3.0	4.38	3.6
3009	2.63	4.89	3.33	3.88	2.3
3010	3.25	4.78	3.67	3.63	4.0
3011	3.0	4.44	3.78	2.13	3.3
3012	3.88	4.44	4.11	3.0	3.3
3013	2.5	3.89	3.22	4.13	2.8
3014	3.5	3.78	3.22	3.0	2.9
3016	4.63	4.11	4.67	3.0	4.0
3017	3.25	4.22	3.56	2.5	3.4
3019	3.5	3.44	4.56	4.0	4.6
3026	3.38	3.44	4.11	3.0	4.5
3028	4.0	4.56	4.22	3.38	4.6
3029	4.13	3.22	4.22	2.88	4.5
3030	3.13	4.11	4.0	2.75	3.3
3031	2.88	4.22	4.44	2.13	3.5
3032	4.38	4.22	3.33	2.25	3.2
3033	2.38	3.89	3.78	3.38	2.9
3035	2.75	5.0	3.56	3.38	4.8
3036	3.13	4.0	3.0	2.38	3.1
3042	2.88	4.56	4.89	3.13	4.2
3043	3.75	5.0	3.89	2.25	3.1
3044	1.5	4.22	4.22	3.38	2.6
3047	2.5	3.89	4.11	3.38	2.7
3048	3.25	3.0	4.22	3.13	4.4
3051	2.63	2.89	4.0	1.88	4.5
3052	3.0	4.22	4.33	3.25	4.3
3053	3.0	4.56	3.67	1.63	3.2
3055	3.13	3.78	3.44	3.38	3.0
3058	2.75	4.22	3.78	3.13	2.9
3063	4.5	4.0	2.67	2.5	3.1
3081	2.63	4.33	3.11	4.38	3.7
3082	4.38	4.11	4.22	2.25	4.1
3083	2.38	4.56	3.78	3.5	2.7

3084	2.63	3.89	4.33	2.88	3.9
3085	5.0	4.44	5.0	1.5	4.4
3086	3.5	4.44	3.11	2.0	2.4
3088	3.13	4.0	3.0	2.88	3.8
3089	4.38	4.67	4.0	2.88	3.0
3090	3.25	3.89	5.0	2.5	4.1
3091	2.88	3.67	2.56	2.38	3.1
3095	3.38	4.0	3.44	1.75	3.8
3096	1.88	4.11	4.0	4.25	2.5
3097	2.5	4.56	3.11	1.38	3.9
3098	2.63	4.0	3.89	3.0	3.4
3099	3.13	3.0	2.78	2.88	3.3
3100	2.5	3.56	3.11	3.25	3.3
3101	3.0	4.0	4.0	3.13	2.7
3102	2.0	4.22	4.0	3.0	3.7
3103	1.25	3.11	3.33	3.0	2.7
3104	3.88	5.0	4.67	2.25	3.5
3105	3.38	4.11	2.67	3.25	3.1
3106	3.0	4.0	3.89	2.13	3.5
3107	3.13	3.89	3.78	3.13	2.9
3108	3.5	4.33	3.89	3.5	3.6
3109	3.88	4.33	4.56	1.88	4.2
3110	4.75	4.44	4.33	1.25	4.6
3112	4.63	3.44	3.56	3.13	4.1
3113	2.0	3.22	3.56	3.38	3.6
3114	2.0	4.0	3.11	3.38	4.0
3115	2.5	4.56	3.67	3.0	3.0
3116	2.88	4.0	4.0	1.63	3.6
3117	3.88	5.0	4.44	2.5	4.0
3118	2.88	3.67	3.56	3.38	3.4
3119	3.38	3.33	3.11	2.63	4.2
3120	4.63	4.11	4.0	2.75	4.7
3518	4.25	4.0	3.78	2.0	4.0
3592	2.25	4.56	3.78	2.13	4.2
3611	3.13	4.78	4.67	2.38	4.5