FANDANGO

FANDANGO DELIVERABLE

Deliverable No.:	D9.2
Deliverable Title:	POPD - Requirement No. 2
Project Acronym:	FANDANGO
Project Full Title:	FAke News discovery and propagation from big Data and artificial inteliGence Operations
Grant Agreement No.:	780355
Work Package No.:	9
Work Package Name:	Ethics requirements
Responsible Author(s):	Massimo Magaldi, Luca Bevilacqua
Date:	30.06.2018
Status:	v0.1 - Draft
Deliverable type:	ETHIC
Distribution:	PUBLIC

Revision History

Version	Date	Modified by	Comments
V0.1	31.06.2018	Angelo Manfredi	First draft
V0.2	19.07.2018	Massimo Magaldi	Full draft
V0.3	23.07.2018	Luca Bevilacqua	Internal Review
V1.0	31.07.2018	Silvia Boi	Quality check

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ABBREVIATION	DESCRIPTION
H2020	Horizon 2020
EC	European Commission
WP	Work Package
EU	European Union

EXECUTIVE SUMMARY

FANDANGO general objective is to assist users (mainly professional journalists) in evaluating the level of trustworthiness that can be put in a piece of information looking like a potential news (i.e. aiding the process of judging whether a news is fake or not).

To achieve this objective, in this document we will explain that two broad categories of issues may rise regarding the access and use of personal sensitive data:

1. Issues related to personal data belonging to research participants,
2. Issues related personal data of people that are not research participants (and are hence in no condition for being asked explicit consent); such data might be inferred while analyzing streams of data about trustworthiness of examined news.

We chose to focus the deliverable D9.1 on all aspects dealing with personal data belonging to research participants. In fact, D9.1 documents detailed information on the informed consent procedures that will be implemented regarding the collection, storage and protection of research participant's personal data.

D9.1 in particular provides templates of informed consent forms and information sheets for human research participants.

This document (D9.2) will deal with all topics related to the risk of inferring personal data of people that are not research participants.

The risk of inferring personal data will be explained in detail and all the procedures to manage any such data will be described.

As of FANDANGO partners have not yet decided to establish a formal position of Data Protection Officer, hence his opinion/confirmation that all data collection and processing will be carried according to EU and national legislation, cannot be submitted yet.

Finally, this document is somewhat similar in content to the D1.2 document, since both documents had to examine the prominent role of privacy concerns in FANDANGO.

1. INTRODUCTION

FANDANGO general objective is to assist users (mainly professional journalists) in evaluating the level of trustworthiness that can be put in a piece of information that looks like a news (i.e. aiding the process of judging whether a news is fake or not).

As part of the project WBS (Task 2.3 - User and system requirements) all project partners representing final users are working at specifying in detail the requirements of the FANDANGO platform, on the basis of the needs and priorities felt by their professional journalists. These requirements will be addressed to the extent allowed by scientific feasibility and compatibility with the approved FANDANGO DoA.

Preliminary results in this direction show that the interviewed representatives of professional users would rather see FANDANGO results help professional human journalists in reaching a conclusion about the trustworthiness of a potential news rather than making an automatic decision about it.

Furthermore, in their opinion, this decision support function would be better served by a set of functionalities.

The functionalities that are deemed most useful are:

- news text verification,
- photo/video text verification,
- alert system.

The first two functionalities would operate with the user explicitly choosing a text, or an image/video to be checked, while the last one would automatically ingest such data from many (explicitly specified) sources to perform similar checks on all the potential news being published by the specified sources.

FANDANGO in this last function would be certainly dealing with a big data class problem since the sheer number of potential news to be examined may be very large, and each potential news would need to be analyzed by considering all its text and multimedia content (potentially large and unstructured files).

However, a big data class problem will be faced also for the easier tasks of verifying a single piece of news.

In fact, FANDANGO verification components will rely on state of the art machine learning approaches and algorithms, and It is well documented that the performance that can be obtained by such an approach can be improved by employing large training datasets.

Moreover, some verification criteria will rely on the confrontation with text or multimedia content belonging to past news (real or fake ones).

This is the reason FANDANGO chose from the onset to leverage an integrated big data platform based on Open Source middleware.

2. Justification of Data Collection

We just stated that FANDANGO verification components will rely on state of the art machine learning approaches and algorithms, and It is well documented that the performance that can be obtained by such an approach can be improved by employing large training datasets and that some verification criteria will rely on the confrontation with text or multimedia content belonging to past news (real or fake ones).

In other words, the data collection in FANDANGO serves many purposes:

1. a preliminary ingestion of data, significantly annotated as belonging two at least two broad categories, that is fake news vs. genuine news, will be necessary in order to train all the machine learning algorithms that will be used to implement the FANDANGO processors;
2. when such a training will be completed (let us say conventionally define T_0 the time at which all modules are trained and ready to operate), further data will need to be collected to assess the level of trustworthiness of a specific news (FANDANGO operating as a news text or photo/video verification tool);
3. when T> T_0 further data may be collected to massively evaluate the level of trustworthiness of all news specified sources (FANDANGO operating as an alert system);
4. when T> T_0 data collected after T_0 may be used to trigger on overall retraining of some/all machine learning algorithms that will be used to implement the FANDANGO processors.

Point number 2 above might not always be necessary. In fact, some processors just need to be trained in order to reach a significant partial fakeness score (as we will describe later image video processor D4.3 is one such example). However, in most cases references to past data will be necessary, such data will need to be stored in some form. In all these cases, FANDANGO partners will apply pseudo=anonymization and appropriate security measures will be taken.

3. Justification of Data Processing

FANDANGO will not process personal data as such, but only data about news, especially data related to potentially untrustworthy (fake) news.

However, while the FANDANGO system has no intention or interest to process personal data as such, some processing steps regarding news data, may offer the theoretical opportunity potentially to infer personal data.

Assessing that a personal account on Twitter is usually writing untrustworthy news and keeping this information for future decision making about other news published by the same person is probably the easiest example showing such a circumstance.

In all such cases FANDANGO will operate by pseudoanonymizing personal data.

As an example, FANDANGO will analyse pictures confronting them to previous pictures in order to check if a past genuine picture is being repurposed out of the original spatial and temporal context to build a fake news (a quite frequent case of untrustworthy intent). For this processing FANDANGO will store only appropriate feature vectors of past pictures therefore minimising issues related to personal data of people appearing in prominent positions in the specific pictures.

To try to explain in more general terms how FANDANGO will deal with these privacy sensitive aspects we need to give some a general model of FANDANGO collection and processing of data.

In short, as already stated, FANDANGO will collect data about potential news whose trustworthiness (i.e. are they real news or fake news?) is uncertain and will generate assessment scores about the probability of each of being genuine or fake.

If we call S_i such fakeness scores (where i = 1, 2 … N and N >> 1) Fandango will generate the S_i fakeness scores by combining partial scores $S_{i,j}$, where j = 1, 2, 3, 4 are the output of specific software modules, studied and developed as part of FANDANGO original results (in WP4, tasks 4.1 to 4.5).

Each module will perform an assessment on the basis of a specific criterion, thus computing a partial fakeness score.

Fandango will generate the S_i fakeness scores by optimally combining partial scores $S_{i,j}$.

This process is depicted in the following figure:

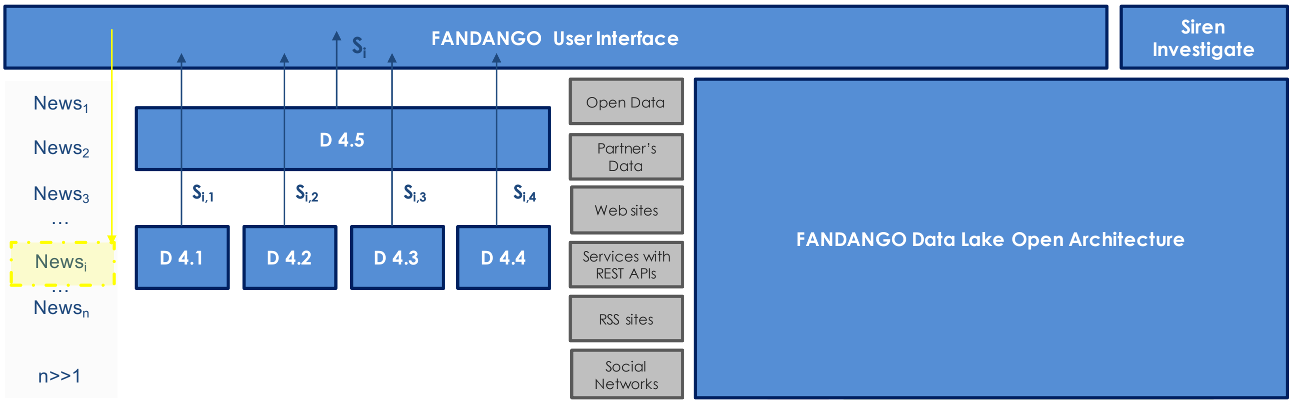

Figure 1 FANDANGO processing of a specific news

More precisely in FANDANGO, all collected data will be processed/analysed by a set of software modules to extract markers and cues in order to reveal fake or misleading news.

As already stated different (four) analysis modules will be in the FANDANGO toolset:

1. The Spatio-temporal analytics and out of context fakeness markers module D4.1 will be responsible for analysing news posts and finding duplicate or near duplicate posts in the past or referring to other geographic/physical locations or contexts. In fact, a common case of fake news is the re-posting of a real past piece of news that it is no longer relevant or is removed from its original context. Such spatio-temporal or out-of-context correlations can generate strong fakeness markers (i.e. generating $S_{1,i}$). Reference to past images is necessary to carry out meaningful comparisons for D4.1.

2. The Multilingual text analytics for misleading messages detection module will handle multilingual content and score it the text as potentially misleading or not. To establish such scoring ability, it will digest data from the public web as well as existing and well updated knowledge bases such as YAGO, DBPedia, Geonames etc.) to identify contradictions and potentially intentional errors. (i.e. generating $S_{2,i}$). Referencing, recently processed text in addition to training set data, might be useful to improve the scoring results for D4.2.

3. The copy-move detection on audio-visual content module will detect the manipulation of images and videos to modify their visual content. This module will leverage deep learning architectures to identify such content and the pool of near duplicate content and visuals that were used as sources for creating the fake object. Synthetic data and publicly available big image datasets will be used to train the models. Moreover, state of the art audio analysis algorithms will be deployed to detect modified or voice-over attacks in news videos (i.e. generating $S_{3,i}$). For D4.3 no reference to privacy sensitive images is necessary once the training will be completed (and the training data discarded).

4. The Source credibility scoring, profiling and social graph analytics module will profile the sources of news and apply graph analytics to detect paths and nodes that tend to produce fake news and spread them widely on the public web. (i.e. generating $S_{4,i}$). D4.4 is a processor that will need appropriate pseudoanonimyzation techniques.

5. D4.5 will fuse the output of all modules above for overall fake news scoring (i.e. generating S_i); a machine learnable score function will be trained to learn how to optimally weight them.

4. DATA STORAGE, PROTECTION, RETENTION AND DESTRUCTION

The measures to be taken for appropriate data storage, protection, and retention (versus destruction) need an overall understanding of the FANDANGO operating logic.

In short, FANDANGO will operate by several logic phases:

- ingesting data sources to train all system components (for $T<T_0$),
- evaluating single news or stream of news (for $T>T_0$) by computing the $S_{i,j}$ partial fakeness scores, and provide such $S_{i,j}$ scores to the human user together with the original media determining the evaluation, in a decision support approach,
- evaluating stream of news (also for $T>T_0$) by computing the $S_{i,j}$ partial fakeness scores, and providing an overall evaluation of probability of fakeness by computing the S_i fakeness scores,
- storing for some time (retention) at least some data, to be able to improve the performance of such algorithms, or to trigger an overall retraining of all system components (likely for $T>>T_0$).

As of today, (month 6 of the 36 months of FANDANGO time span) D9.2 is formally due, and has to be finalised and uploaded, while none of the application scenarios or processing algorithms has been defined in detail yet.

As a consequence, some data sources, most of the intermediate data models, and more in general all aspects related to the optimal design and training of the processing modules still have to be finalised.

Obviously, this fact precludes the possibility of defining in full detail all aspects of the procedures for data storage, protection, retention and destruction.

Further details will be given in specific sections of periodic reports.

5. GDPR General Procedural Implications for FANDANGO partners

The right to data protection is enshrined in the Charter of the Fundamental rights of the European Union (2000/C364/01). The Charter guarantees the right to protection and fair processing of personal data. Data protection provisions are also included in the primary law of the EU.

The Treaty on the Functioning of the European Union (TFEU) provides that 'everyone has a right for the protection of the data concerning them' and mandates the European parliament and the Council to make the necessary rules for the protection of people's rights (Article 16/2). This provision led the EU parliament to draft the Directive on Protection of Individuals Personal Data 95/46/EC.

At this point we can think of two kinds of information that could improve FANDANGO effectiveness:

1) Professional End User Settings: The FANDANGO system could remember certain settings, so that the user (professional journalist) will not need to enter these settings (e.g. topics or field of investigation he is most interested in, level of expertise in detecting fake news ecc.) every time s/he wants to use the system.

2) People social posts: FANDANGO aims at detecting patterns of fake news propagation and evaluating the trustworthiness of specific publishers of news potentially fake. As neither the graph analysis algorithms nor the technical solutions have been decided in near to final detail, the respective data management options cannot be described yet. At this point it seems very likely, that only in specific aspects personal data may be held. In these cases when personal or sensitive data will be kept in FANDANGO, FANDANGO itself will decouple all data that may be used to connect an identified user with any sensitive data.

As a general complement to more specific considerations presented in the previous chapters in the following table a high-level checklist for FANDANGO partners in order to ensure compliance with GDPR requirements.

Transaction location	If a transaction or service takes place inside the EU, then EU residents' personal data is protected by the GDPR. GDPR rules apply to all organisations collecting or processing such data.	Establish routines to determine location of transactions or services offered.	GDPR Articles 3, 28-31
User consent	Explicit consent must be given by users. Language must be simple and clear. Special conditions apply to children consent.	Check that requests for consent specify the purpose of data collection/processing, and that the language used is simple and clear. Provide details about exactly which data will be collected.	GDPR Articles 4 & 32
Users' right to access their data	Users have a right to know whether their data is being held and processed. Users are also entitled to receive an electronic copy of their data, free of charge.	Each organization handling user data must declare that they process user data and must indicate one or more contact points within the organization for data-related queries. Such details should be clearly communicated to users.	GDPR Articles 12, 15
Data portability	Data Subjects have the right to receive their data in a 'commonly used and machine-readable format'.	Each project partner should establish protocols for making data available in 'commonly used and machine-readable' formats.	GDPR Articles 12, 20
Right to be forgotten	A Data Subject has the right to request that his/her data be erased or withdrawn from dissemination. Under certain circumstances ('public interest') requests may be denied.	Each project partner should establish procedures for dealing with requests to delete or amend user data.	GDPR Article 17
Data integrity	Processing of data must take place with ensured security of data. Ethical principles must be observed and appropriate technical measures must be taken.	Each project partner should set up a data storage and transfer (if applicable) protocol per organization. Partners must ensure compliance with GDPR rules on data security, including encrypted transmission (if applicable), secure data transfer (if applicable), secure storage.	GDPR Articles 6, 35, 36 and 83
Privacy by design	'Privacy by design' calls for the inclusion of data protection from the onset of the designing of systems, rather than as an addition.	All partners' data management teams are obliged to adopt the principle of 'privacy by design' for all future data processing systems. This methodology must be established before data collection begins.	GDPR Article 25
Data minimisation	Only the minimum data required to achieve the specified goal may be held. Access to personal data must be limited to those with the need to access it.	Partners must establish protocols within the organisation or project to ensure that only required data may be stored. Partners must ensure that only those people necessary for processing or maintaining data infrastructure have access to collected data.	GDPR Article 23
Accountability	Data controllers must maintain clear and secured records of any data processing activities performed under their responsibility.	Partners must demonstrate that data processing activities and security measures are compliant with GDPR requirements.	GDPR Article 6
Legal basis for holding data	Organisations must now identify the legal basis for holding subjects' personal data, and must document those reasons.	Partners must justify and document the legal reasons for holding user data.	GDPR Article 5 (2) and 6

Pseudonymization	Pseudonymization of data must be carried out so that personal data can no longer be attributed to a specific data subject without the use of additional information. Such additional information must be kept separately and securely.	Partners must detail pseudonymization measures (if relevant) and ensure that they correspond to GDPR requirements. If additional data which may be used to identify users is stored, partners must document the intended storage plan and security measures for storage and transmission.	GDPR Article 5 (2)
Metadata	Metadata is to be anonymized or deleted if users have not given explicit consent for its use, unless needed for billing purposes.	Partners who store, transmit or handle metadata that relate to non-anonymous or non-pseudonymized user information must identify whether this data is needed for billing purposes. If not, partners must identify this anonymization and/or deletion procedures.	GDPR Article 5 (2)
Data Protection Officer	Data Protection Officer (DPO) appointment will be mandatory only for those controllers and processors (a) whose core activities consist of processing operations which require regular and systematic monitoring of data subjects on a large scale or (b) of special categories of data or data relating to criminal convictions and offences.	It is acceptable under GDPR to appoint a DPO for a collection of organisations. Shall FANDANGO do it ?	GDPR Article 37
Data transfer	Personal data may only be transferred outside of the EU in compliance with the conditions for transfer set out in Chapter V of the GDPR.	If partners intend to transfer data outside the EU, do such transfers comply with the new regulations?	GDPR Article 45, Recitals 103-107, 169
Re-use of personal data	Repurposing of data is only allowed when the new purpose is compatible with the initial purpose.	Partners must ensure that data purpose compatibility tests are in place before re-use of personal data.	GDPR Article 6 (4)
Breaches of data security	All breaches must be reported to the Data Protection Officer within 72 hours.	Protocols for action to be taken in the light of any breach of data security must be documented and reported to the DPO.	GDPR Article 33/34
Penalties	Up to 4% of global turnover or 20m EUR, whichever is greater. Both data controllers and data processors are subject to penalties.	Partners must ensure that relevant staff are aware of the significant penalties for breach of the GDPR.	GDPR Articles 33, 34, 83

Table 1