D3.4

Coordination of Legal Aspects in USEMP – v1

v 1.0 / 2015-05-19

Coordination by Katja de Vries, Niels van Dijk and Mireille Hildebrandt (iCIS-RU).
Contributions by Hervé Le Borgne and Adrian Popescu (CEA), Giorgos Petkos (CERTH), Noel Catterall and David Lund (HWC), Laurence Claeys (iMinds), Ali Padyab (LTU), and Theodoros Michalareas (Velti).

This document presents the results of the legal coordination and the integration during the first half of the USEMP project. This deliverable is the fruit of intense interdisciplinary collaboration with all partners and shows how legal requirements are interfaced with the technical design. The deliverable offers a legal qualification of all the data that is handled by the USEMP system and the legal requirements this implies. While also looking at anti-discrimination law and intellectual property law, the main focus of this deliverable is on the legal requirements which follow from data protection law. Two types of data protection requirements are presented: those based on EU data protection law (compliance) and those which aim to strengthen the freedom of the user towards OSNs and browsers and help her to exercise her fundamental right to data protection (empowerment). By presenting a hyperlinked version of the Personal Data Processing Agreement (PDPA) and the Data Licensing Agreement (DLA), this deliverable clarifies how these two contracts embody all relevant data protection requirements, how they link to the qualification of data as personal data and to the full listings of personal data processed by the USEMP system. The deliverable also presents a first version of the flow charts which should be available behind the information button on the USEMP Platform, together with the PDPA, DLA, with an overview of the personal data processed in the USEMP project.

Project acronym	USEMP
Full title	User Empowerment for Enhanced Online Presence Management
Grant agreement number	611596
Funding scheme	Specific Targeted Research Project (STREP)
Work program topic	Objective ICT-2013.1.7 Future Internet Research Experimentation
Project start date	2013-10-01
Project Duration	36 months
Workpackage	WP3
Deliverable lead org.	ICIS
Deliverable type	Report
Authors	Katja de Vries, Niels van Dijk, Sari Depreeuw and Mireille Hildebrandt (iCIS), Hervé Le Borgne and Adrian Popescu (CEA), Giorgos Petkos (CERTH), Noel Catterall and David Lund (HWC), Laurence Claeys (iMinds), Ali Padyab (LTU), Theodoros Michalareas (Velti)
Reviewers	Tom Seymoens (iMinds) Symeon Papadopoulos (CERTH)
Version	1.0
Status	Final
Dissemination level	PU: Public
Due date	2014-12-31
Delivery date	2015-05-19

Version	**Changes**
0.1	Initial Release, Katja de Vries and Niels van Dijk (iCIS) with

	input from Noel Catterall (HWC), Adrian Popescu (CEA) and Giorgios Petkos (CERTH) 17 April 2015
0.2	Revisions Mireille Hildebrandt (iCIS) 17 April 2015
0.3	Revisions Hervé Le Borgne (CEA) 21 April 2015
0.4	Revisions Giorgios Petkos (CERTH) 21 April 2015
0.5	Revisions David Lund (HWC) 22 April 2015
0.6	Revisions Laurence Claeys (iMinds) 22 April 2015
0.7	Revisions Ali Padyab (LTU) 23 April 2015
0.8	Revisions Theodoros Michalareas (Velti) 27 April 2015
	Internal review by Symeon Papadopoulos (CERTH)
	Internal review by Tom Seymoens
1.0	Final version 19 May 2015 (by De Vries, Van Dijk, Hildebrandt)

Table of Contents

1.Introduction

This deliverable (3.4) presents a 'report that provides a description of the harmonised legal constraints applicable to USEMP data, algorithms and platform' (DOW, p. 55), and a second version (3.9) will be delivered by the end of this year. The task (3.6) continues throughout the project, because 'the legal requirements that will be developed within this task will have to be interfaced with the tasks at hand in the other WPs. Without mutual understanding of the relevant constraints the legal requirements would develop in a vacuum and the social and technical WPs may not be capable of integrating them into their operation' (DOW, p. 54).

This first version of the report is the result of intense mutual collaboration between the technical partners and the legal partner, in order to gain a precise understanding of what data are processed how, and by which partner. This enabled us to qualify the data in terms of the legal framework and to map the legal effect: the applicability of the specific rights (for the DataBait users or data subjects) and obligations for the USEMP Consortium Partners, as joint data controllers, that follow from the legal qualification.

The report starts out by explaining the classification of data in the context of USEMP (section 1.1), referring to the raw input data, and the output data in the form of data derivatives that have been inferred from different types of datasets (both internal and external). Next to this functional classification the project works with a number of technical classifications, depending on the format and the data model used. In this report we consider the classification that follows from the legal qualification of the data involved, in order to elicit the relevant legal requirements.

Next, the report re-introduces the Personal Data Processing Agreement (PDPA) and the Data Licensing Agreement (DLA) that form the core of the legal framework within USEMP (section 1.2). The difference between sensitive data in the legal sense and other uses of the term sensitive data is reiterated, emphasizing that, in this report, the focus is on the legal qualification and empowerment (section 1.3), and not on enabling the user to engage in her own perception management.

The central element of this report is formed by a detailed elaboration of the relationship between the PDPA, the DLA and the legal requirements that follow from the legal qualification of the data processed in the backend of the DataBait tools. The architecture that brings together the data, their qualification in terms of the different legal domains, the ensuing legal requirements and the DLA/PDPA is first introduced (section 1.4) and then presented (section 1.5). The interconnections are easily followed due to extensive hyperlinking, preventing endless searching and scrolling between the lists with data, the tables with relevant legal requirements and the roots in the DLA/PDPA. In Annex A, the legal requirements are further specified as to data protection and non-discrimination. The multi-dimensional contraption is followed by a discussion of the information and withdrawal buttons that must be implemented on the USEMP platform to comply with the information obligations, thus empowering users to gain insight into the backend of the DataBait tools (section 2).

In sections 3 and 4 the requirements concerning non-discrimination and IP law are discussed, mainly indicating that they will be further developed and integrated into the next versions of the DLA and the DataBait tools, resulting in D3.6-9. In Annex B a first indication is provided of the types of legal requirements that must be integrated regarding IP rights.

Section 5 presents concluding remarks and summarises the research to be done for the next versions of the deliverables.

2. Legal qualification of the relevant data. Data types & legal requirements

2.1. Classifying and qualifying data in the USEMP project

The USEMP project processes a multitude of data. Firstly there are three types of raw data which are collected through the DataBait tool from DataBait users: OSN (Facebook and/or Twitter) data, browser (Mozilla and/or Chrome) data, and in some cases (when users participate in the USEMP pre-pre-pilot) data from the DataBait survey.[1] From a sub-set of the first two types of raw data, (OSN and browser data) additional data are derived with software (so-called "data-driven modules") developed within the USEMP project. These derived data are so-called 'data derivatives': they are data which are inferred from the original OSN and browser data. Next to the OSN, browser, survey, and derived data, the USEMP project also processes data from external data sets. Together with the survey data these data from external data sets (e.g. a set of pictures from Flickr or a set of Wikipedia pages) are used to train and test the algorithms in the USEMP data driven modules which transform the USEMP input data[2] (e.g. a set of Facebook pages that a user has liked or the URLs that a user visits in their browser) into USEMP output data ('data derivatives'). A schematic overview of these five types of data can be found in table 1.5.3. The first four types of data (OSN, browser, survey and derived data) relate directly to individual DataBait users. The OSN and browser data can be considered the input data for the DataBait tools, the derived data are the output data and the survey data and data from external data sets are the training and testing data which are necessary for the software which transforms input into output data (see figure 1).

The classification of data processed in the USEMP project into the aforementioned five classes is based on the source from which they are derived. Next to this functional classification, one could also classify the data based on a different criterion such as their format (e.g., is it an image or is it text?), their type (does the data constitute sensitive information or not?), or their mode of creation (is this data created in an automated way, is it made creatively with a distinct author, has a substantive investment been put in the

[1] This is a data typology based on the source of the data. It will be used to answer whether these data can be legally qualified as personal data or sensitive data, as protected anti-discrimination grounds or as protected intellectual property. In order to disambiguate, we can compare this data typology to the 4 categories mentioned in section 2.1 of F.6.1: 1) general (socio-ethical) sensitive data; 2) sensitive data according to legal criteria; 3) sensitive data as perceived by users, 4) types of data according to their source. We could thus say that the data typology used here, bares most resemblance with the fourth category (data source) in order to get at the second category (legal qualification). The other two types of D6.1 will not be considered here.

[2] See table A.1, A.2 and A.3 in Annex A for all the data collected by the DataBait tool. See table A.4 for the subset of these collected data which are (1) used for training classifiers (i.e. mathematical models which allow to transform the collected raw input data into inferred output data or so-called 'data derivatives') and, (2) as input for the derivation of the output data. See also D2.3 for the data which are used as input for deriving the data derivatives. It should be noted that it is not completely set in stone yet which of the collected data will be used as input data for deriving 'data derivatives' (output data) and which will not be used for any further inferences.

making?[3]). When the data are classified into legal categories (such as ‘personal data’, ‘sensitive data’ or ‘copyrighted content’) this is called *qualification*. When data are legally qualified according to a particular legal denominator, this has *legal effects*. For example, when a piece of data is qualified as ‘personal data’ in the sense of Art. 2 of Data Protection Directive 95/46/EC, this means that the processing of this data has to be done in accordance with the requirements set out in this law. When data processed in USEMP is legally qualified as personal data, the legal effect is that a bundle of legal rights applies to the end-user and a bundle of legal obligations applies to the service providers (i.e. the USEMP Consortium Partners). Thus, the legal qualification of data and the ensuing legal effects, affect the way in which the system processing such data should be designed. Consequently, the legal qualification of data processed within the USEMP project has to result in *legal requirements* for the design of the DataBait tools developed in this project. This relates to the concept of *Legal Protection by Design* (of which *Data Protection by Design* is one particular type) described extensively in D3.1.

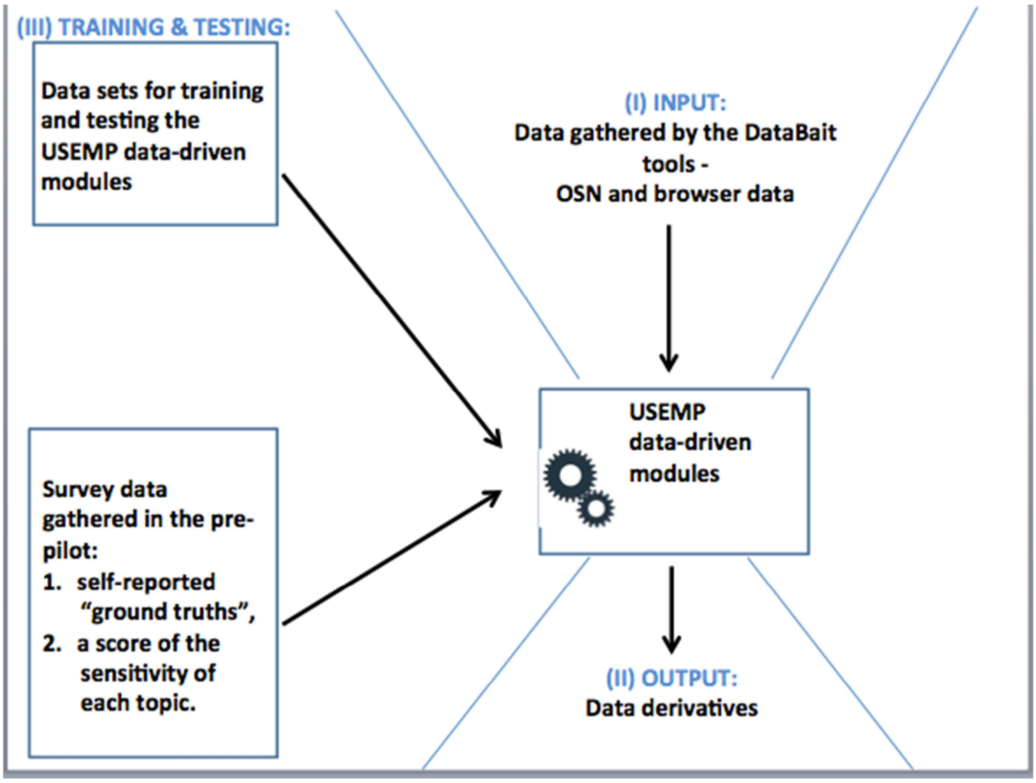

(III) TRAINING & TESTING:
Data sets for training and testing the USEMP data-driven modules
(I) INPUT:
Data gathered by the DataBait tools - OSN and browser data
USEMP data-driven modules
Survey data gathered in the pre-pilot:
1. self-reported "ground truths",
2. a score of the sensitivity of each topic.
(II) OUTPUT:
Data derivatives

Figure 1. Four types of data relating directly to individual DataBait users: OSN, browser, survey and derived data. The OSN, browser and survey data are directly collected from the user and the data derivatives are indirectly calculated from a subset of the OSN and browser data.

The data processed in USEMP are qualified from the perspective of several legal fields: EU data protection and privacy law, anti-discrimination law and intellectual rights law. As explained in D3.1-D3.3 these qualifications are not mutually exclusive: several qualifications can apply. For example, a picture posted on a Facebook profile can constitute personal data from the perspective of EU data protection law (it relates to an identifiable person), a protected ground from the perspective of EU anti-discrimination law (e.g., it depicts the racial origin of a person, and this racial aspect of the picture is used as a ground to deny the

[3] These are the legal requirements from intellectual rights like copyright and *sui generis* data base rights.

depicted person certain services; for example, one could imagine a commercial business which tries to filter its customers based on race – this is clearly prohibited) and copyrighted content from the perspective of IP law (the image is made by an author who decided on the composition, the framing, the light, etc.; this implies that this content cannot be reproduced or distributed without a license to do so).

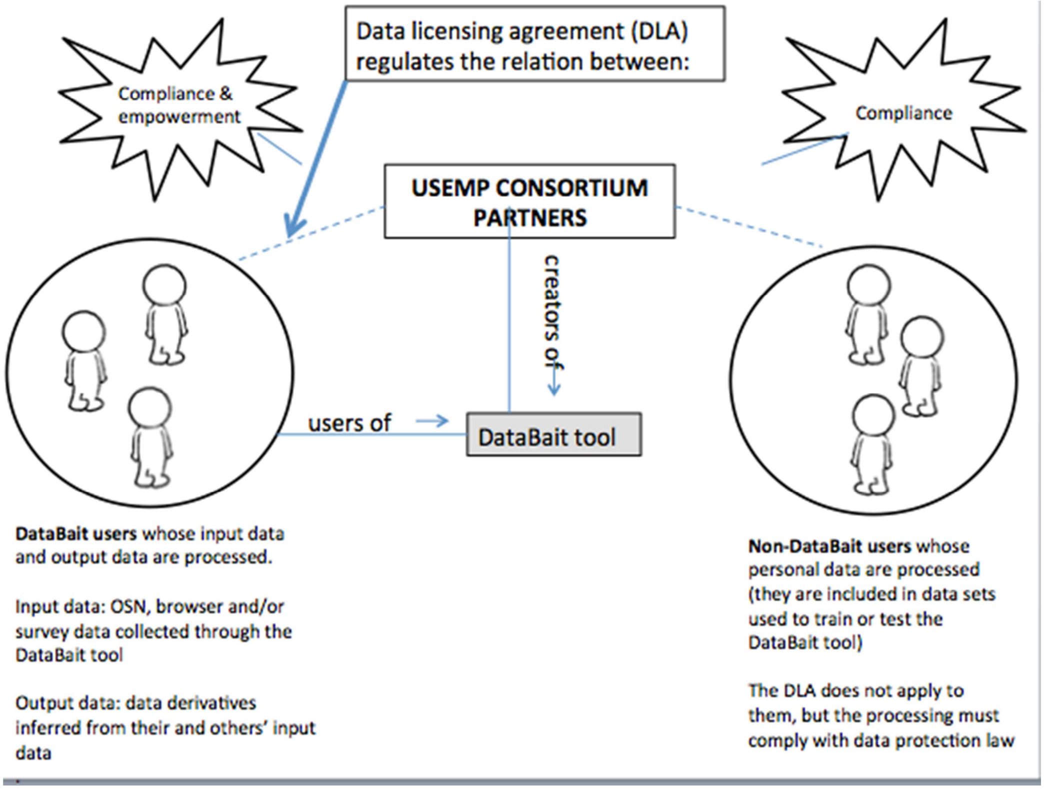
In this deliverable the main focus is on the legal qualification of data from the perspective of data protection law and the legal requirements which are derived from such qualifications. Some preliminary qualifications of the data from the perspective of anti-discrimination law and intellectual rights law are also made. Due to the fact that much of the information required to figure out the relevant legal implications was only available recently, it was not feasible to elaborate the integration for all the relevant law. For this reason, more detailed qualifications and the requirements with regard to anti-discrimination law and IP law will be presented in the final version of deliverable D3.9 and in D3.6-8. This is also related to the fact that the implications for non-discrimination and IP law are more speculative and less researched.

Overall, this is part of the interactive and reiterative nature of data protection by design. The legal work of qualification is continuously dependent on the description of the technical specifications that are still in flux. In turn, the legal requirements continuously inform and tweak the design of the technical system. In this sense it is a process of mutual specification.

2.2. The Data Licensing and Personal Data Processing Agreements as a source of the data protection requirements

The Data Licensing Agreement (DLA) and Personal Data Processing Agreement (PDPA) regulate the legal relation between the users of the DataBait tool and the USEMP consortium partners; they incorporate the legal requirements which follow from EU data protection law (*compliance* of data processing within the USEMP project with the law) and add some additional requirements for the socio-technical architecture of the DataBait tool in order to strengthen the freedom of the user towards OSNs and browsers and help her to exercise her fundamental right to data protection (*empowerment* to enable the exercise of the data protection rights granted by the law in relation to actors which track and profile individuals when they use OSNs and browsers). Thus, the DLA regulates the relation between the USEMP consortium and the DataBait user both from a compliance and an empowerment perspective (see figure 2).

In addition, the PDPA and DLA also incorporate the fact that personal data from non-DataBait users (which might be contained in the external data sets) have to be processed in compliance with EU law. However, in contrast to the legal requirements in the PDPA and the DLA with regard to the relation between the USEMP consortium and DataBait users, the legal requirements with regard to the relation towards non-DataBait users are based on the data protection law and not on the DLA. The latter is obviously not an instrument of empowerment for those whose data are processed as part of an external dataset (see figure 2).


Compliance & empowerment
Data licensing agreement (DLA) regulates the relation between:
Compliance
USEMP CONSORTIUM PARTNERS
creators of
users of
DataBait tool

DataBait users whose input data and output data are processed.

Input data: OSN, browser and/or survey data collected through the DataBait tool

Output data: data derivatives inferred from their and others' input data

Non-DataBait users whose personal data are processed (they are included in data sets used to train or test the DataBait tool)

The DLA does not apply to them, but the processing must comply with data protection law

Figure 2. The DLA regulates the relation between the USEMP consortium and the DataBait user – both from a perspective of compliance and empowerment. The relation between Consortium Partners and non-DataBait users is not based on the contract, but on data protection law. We note that the USEMP Consortium Partners are also the operators and administrators of the DataBait tools.

In this deliverable we clarify which legal requirements are incorporated in the articles of the DLA and PDPA, in what legal qualifications they are rooted (personal data, sensitive personal data, or personal data which form the input for or the output of profiling and location data). This report lists all data which fall under these legal qualifications (see Annex A). In the next section we clarify the legal qualification of (sensitive) personal data and distinguish it from the common sense understanding of what makes a piece of data sensitive.

2.3. Personal data – when are they sensitive and when not? The legal requirements and effect

When are data so personal, private or sensitive that they should be treated with extra care, only processed under special conditions and with sufficient safeguards in place or maybe even not processed at all? As shown in D6.1 this is a question which cannot be answered in an unequivocal way. What is considered personal, private or sensitive does not only vary from one culture to another but might also be appreciated differently by each individual. Moreover, various disciplines have different ways of studying this question: a social scientist might interview people, a statistician (who might also be a social scientist) might try to infer the answer by looking at the type of information which people reveal (assuming that the information which people shield is probably considered as more private or sensitive) and a legal scholar will turn to the law for an answer. Data Protection law distinguishes between personal data and sensitive data, qualifying the latter as a subcategory that requires specific safeguards. Both are defined in a precise manner and do not depend on what a specific person believes to be sensitive. The extra protection provided for sensitive data aims to prevent specific types of discrimination based on similar grounds as those in non-discrimination law, as discussed in D3.1. This means that sensitive data in the legal sense need not be the same as those which are inferred to be perceived as sensitive based on socio-statistical calculations (see D6.1, chapter 2, on various ways of calculating a 'disclosure score'[4] for a piece or set of data which are exposed on an OSN). In the context of D5.2, USEMP has defined certain data or content as 'private', and in the context of D6.1 USEMP has developed the aforementioned disclosure score – neither of which should be equated with the legal qualification of a data being either personal or sensitive or with the legal qualification of privacy. In the USEMP user studies done in WP4 (T4.2) the user's perception and definition of sensitive data by Facebook users will be studied in more detail via a card sorting exercise and qualitative individual interviews. While studying perceived privacy is important toto the USEMP project, it is crucial to acknowledge that legal protection of fundamental rights such as privacy and data protection do not depend on whatever a person wishes to hide or perceives as an invasion; notably because of the invisibility of the consequences of sharing data and the need to compensate power inequalities. In the context of USEMP, this means that the DataBait tools aim to provide profile transparency beyond mere perception management, they aim to help the user of an OSN to gain a clear picture of how she may be targeted, in order to exercise her data protection rights. This could, for instance, mean that she requires the data controller to stop processing her personal data, based on either art. 12 or 14 of the Data Protection Directive. This will require special provisions to be made in the user interface.

As explained in 1.1 of D2.3 it is important to distinguish between what is perceived as personal data or sensitive data by end-users and the *legal qualification* of an activity as 'personal data processing', or even as 'processing of sensitive personal data'. The difference between perceived and legal (sensitive) personal data can be nicely illustrated by taking a closer look at some of the 'private concepts'[5] explored in D5.2: a) smoking, b) drinking alcohol, c) extreme sports (climbing), d) political beliefs (participation in demonstrations), e) luxurious living (yacht). Whether these topics are considered as personal or sensitive will vary from individual to individual. Legally speaking, each of them data is personal when they can be linked to an identified or identifiable person. Whether data can be qualified as sensitive personal data in the sense of EU data protection law (Art. 8 DPD 95/46) depends on whether they reveal any information relating to racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, health or sex life, or data relating to offences, criminal convictions or security measures. Thus, from a legal perspective, only 'political beliefs' is clearly an instance of sensitive data in the subset[6] of private concepts discussed in D5.2, when it is applied to a Databait user. Smoking, drinking and extreme sports could be considered sensitive data if they are considered as health data – which will depend on a number of circumstances.[7] Luxurious living is clearly not sensitive in the sense of Art. 8 DPD. We note that only those data that are applied to an identifiable individual – whether volunteered, observed or inferred – can be qualified as sensitive data in the sense of

[4] In the current version of D6.1 this is referred to as the 'privacy score'. In order to avoid confusion with the legal understanding of 'privacy', this notion was changed to 'disclosure score'. In the next deliverable (D6.4) the terminology will be updated.

[5] These relate to the list of privacy dimensions explored in D6.1: A) Demographics, B) Psychological Traits, C) Sexual Profile, D) Political Attitudes, E) Religious Beliefs, F) Health Factors & Condition, G) Location and H) Consumer Profile.

[6] It should be underlined that here we only refer to the subset of five private concepts discussed in D5.2. The full list of private concepts discussed in D6.1 contains many more data types which qualify as sensitive in the sense of Art. 8 of DPD 95/46.The full list can be found in Annex A in table A.5.

[7] See Art. 29 WP Annex on health data in apps and devices, where the concept of 'health data in Directive 95/46/EC' is explained, http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf.

the DPD. In this report we will not use the terms P-Sensitivity for perceived and L-Sensitivity for legal sensitivity, as proposed for the technical deliverables, since here we only consider the legal aspect of the data.

Clearly, the way that the USEMP-DataBait presence tool creates more awareness about one's online presence, does not coincide with the obligation of 'profile transparency' following from EU law. The obligation to provide 'profile transparency' falls on the data controller[8], which in the case of data available on an OSN (Online Social Network service) is most likely to be the OSN (e.g., Facebook). The 'profile transparency' provided by Databait-tools is a simulated transparency offered by a third party (the USEMP consortium) which does not follow from any legal obligation nor exempts the data controller from her obligation. By creating this type of profile transparency, however, the user (data subject) need not trust the OSN (data controller) which may have an own interest in hiding algorithms and data aggregation (an interest that is protected by trade secret and IP rights). In that sense the profile transparency provided via DataBait is not based on an existing legal requirement. It nevertheless helps the data subject to exercise her fundamental right to data protection (*empowerment* to strengthen the data protection rights granted by the law towards actors which track and profile individuals when they use OSNs and browsers). Legally speaking, the provision of this transparency is relevant for two reasons: (1) OSN providers (data controllers) may have a legal duty to abstain from disabling such 'counter profiling' and (2) provision of potential profiles by which users may be targeted may be seen as a right to information, falling under the horizontal effect of the fundamental right to freedom of expression. At some point one could even argue that targeting people on the basis of their behavioral data points – based on inferences from other people's data, social graphs, etc. – should be conditional on the existence of independent providers of profile transparency. In that sense it could become part of the state of the art for Data Protection by Design (DPbDesign), requiring data controllers to do their business in a market that incentivizes independent provision of counter profiling. Due to the fact that the legal obligation of DPbD is not part of the existing legal framework, it is difficult to predict how the obligation will be operationalized under the upcoming framework.

2.4. Introducing the relation between the PDPA/DLA and the tables enumerating the legal requirements and processed personal data

With regard to data protection the DLA and PDPA contain all legal requirements which are relevant for the processing of personal data in the USEMP project. As explained above, legal requirements follow from a particular legal qualification (notably, data are qualified as personal data, as sensitive personal data or as location data;[9] some of these personal data form the input for profiling, others are the output of profiling). A legal qualification implies specific legal effect, based on legislation or on contractual obligations.

[8] «the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data»

[9] Under the ePrivacy Directive, location data fall under a specific legal regime. We will explore this below.

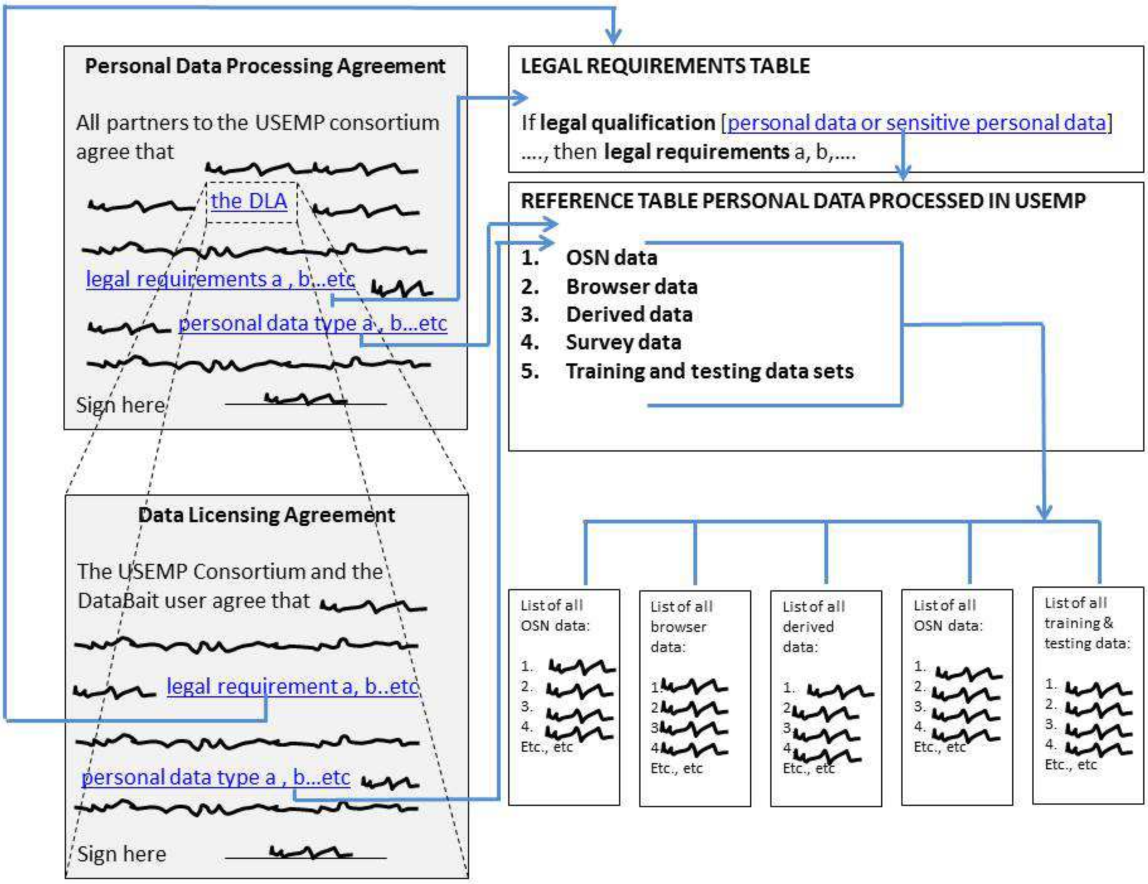

Personal Data Processing Agreement
All partners to the USEMP consortium agree that
the DLA
legal requirements a , b...etc
personal data type a , b...etc
Sign here
LEGAL REQUIREMENTS TABLE
If legal qualification [personal data or sensitive personal data] ...., then legal requirements a, b,....
REFERENCE TABLE PERSONAL DATA PROCESSED IN USEMP
1. OSN data
2. Browser data
3. Derived data
4. Survey data
5. Training and testing data sets
Data Licensing Agreement
The USEMP Consortium and the DataBait user agree that
legal requirement a, b..etc
personal data type a , b...etc
Sign here
List of all OSN data:
List of all browser data:
List of all derived data:
List of all OSN data:
List of all training & testing data:
1. 2. 3. 4. Etc., etc

Figure 3. The hyperlinks between the DPA/PDPA and the various tables in this deliverable

In table 1.5.5 the relation between these legal qualifications, the legal effect and the ensuing legal requirements is presented. The legal requirements in table 1.5.5 are compliance requirements: they follow from EU data protection law (compliance of data processing within the USEMP project with the law). Next to the compliance requirements, the DLA also contains legal empowerment requirements (articles D and E) which aim to strengthen the freedom of the user towards OSNs and browsers and help her exercise her fundamental right to data protection. They are also a matter of compliance, but now based on the contractual obligations in the DLA.

In tables 1.5.1 and 1.5.2 we present annotated and hyperlinked versions of the PDPA and DLA. The hyperlinks refer the reader to the relevant requirements in table 1.5.5 and to the type of personal data (OSN, browser, derived, survey, or test & training data) to which reference is being made. These types of data are listed in table 1.5.3. In turn, this table contains hyperlinks to the full lists of personal data in Annex A of this deliverable. Thus, when one clicks on the word “inferences” in the DLA, one is referred to the category ‘output data (data derivatives)’ in table 1.5.3. There one clicks on a second link which refers the reader to table 6.1.5 in the Annex, which gives a full list of all the personal data inferred from the raw input data. The basic structure of the hyperlinks is explained in figure 5. This hyperlinked structure allows the reader to easily go back and forth between the PDPA (incorporating the DLA - which is underlined by a hyperlink in art.B of the PDPA), the DLA and the table with legal requirements (1.5.5), the tables listing the aforementioned five types of data (1.5.3 and 1.5.4), and the lists of personal data and their legal qualifications in Annex A. This allows the reader to trace the “pedigree” of the various articles in the PDPA and the DLA, or to find out how the legal requirements are fulfilled by the various sections of the PDPA and the DLA (by clicking the hyperlinks in the requirements table and being referred to the relevant article in the PDPA or DPA).

Next to the possibility of retracing the pedigree of the legal requirements, the reader of this deliverable and the DataBait user can also retrace the pedigree or 'life story' of the data processed in the USEMP project. Firstly, table 1.5.4 shows for each type of personal data where they are stored, for how long, what the technical goal of their processing is and how they are stored (anonymized/pseudonymized). Secondly, in table A.5 of Annex A for each inferred data category (the output data or 'data derivatives') it is explained how the inference was made. Table A.5 is not fully populated yet, which is due to the way in which the USEMP data-driven modules (which transform the raw DataBait input data in output data) are developed: this is an empirical process where the best method (e.g., is it better to infer "ethnic origin" from images or status updates? And is it better to use algorithm a or b?) is only discovered along the way of the Databait tool implementation through experimentation. An updated version of table A.5 will be presented in the next version of this deliverable (D3.9).

A final remark should be made about the relation between personal data reference table presented in subsection 1.5.3 and the personal data lists in Annex A. Table 1.5.3 only distinguishes five types of data, and yet Annex A contains 6 lists. Two of the lists refer to OSN data: table A.1 is an exhaustive and up-to-date list and table A.3 presents a list of OSN data made in the very beginning of the USEMP project. The data listed in A.3 do not map precisely on those presented in A.1 which is due to the fact that the Facebook API defines which data can be used and only gives limited permissions. Table A.3 is nevertheless interesting as it shows more concisely what kind of information the USEMP consortium derives from Facebook. It is also useful in the sense that it contains codes (C1, C2, etc.) which are referred to in the list of data derivatives (table A.5). Furthermore, there are two lists in table A.2 referring to browser data: one list of browsing behaviour and one containing data with regard to trackers which track the browsing behaviour. Then there is a list of data derivatives (A.5) and a list of data sets (A.6). It should be noted that there is no list of the survey data. The full survey can be found in deliverable in D4.5 (User Categorisation of Digital Footprint - V2). However, the DataBait survey is based on the categories of data derivatives listed in table A.5 (asking for the true values or so-called ground truths of the values which the data-driven USEMP modules try to infer: age, gender, nationality, racial origin, ethnicity, etc.)

2.5. The PDPA and DLA, linked to the tables with legal requirements,[10] and to the reference table regarding the five listings of personal data in the Annex

2.5.1. PDPA

	Why is this clause important?
USEMP Personal Data Processing Agreement (PDPA) The parties: (1) CEA-France, (2) iMinds-Belgium (3) CERTH-Greece (4) HWC-UK (5) LTU- Sweden (6) VELTI-Greece (7) SKU Radboud University-the Netherlands having concluded the USEMP Consortium Agreement, being providers of the USEMP platform and the DataBait tools and services, and being joint data controllers, Hereby agree:	This PDPA regulates the legal relation between the partners in the USEMP consortium.
(A) Each party will comply with and perform in accordance with the USEMP Data Licensing Agreement (DLA, as attached to this contract) when processing the personal	This links the PDPA to the DLA.

[10] The PDPA and the DLA are hyperlinked, where relevant, to the other tables, to make moving back and forth between the tables more easy on the reader and to clearly indicate how the tables relate to the legal framework of the DLA.

data of DataBait Users, who are defined as the USEMP end-users who have signed the Data Licensing Agreement with the USEMP Consortium Partners.	
(B) Each party will comply with their national and EU data protection law, including notification of their national Data Protection Authority if necessary under their national law, when processing the personal data of DataBait Users or any other personal data processed in the context of USEMP.	This ensures that all input, output and training & testing data are processed in compliance with EU data protection law.
(C) Each party will provide precise information on what type of personal data they process concerning DataBait users, how it is processed and which data-flows they enable. This information will be available for DataBait users after clicking the button on the USEMP platform, and include an email address for each partner that processes personal data, to make further inquiries. The information will be updated whenever the relevant processing of personal data change. Each party will also provide an email address to be contacted in case a user wants to withdraw her consent for processing her sensitive data; this is preferably the same email address as the one used to gain further information, but will be available behind a separate button on the USEMP platform.	This ensures that the DataBait tool will have two buttons which are necessary in order to be compliant with EU data protection law: (1) a button to all the information which should be accessible, and (2) a button to withdraw the consent for the processing of sensitive data
(D) All parties shall carry out a personal information assurance risk assessment from their own context concerning their own collection, storage and/or processing of personal data, prior to deployment of the live service when personal data will be collected, and at any point through the operation of the system where there is a relevant change to either hardware installation, software versions, and/or software interfaces. Such a risk assessment shall follow information assurance principles covering, at least, hardware installation, software development processes, software validation and approval, software execution and backup processes. Each partner is liable for inappropriate security at its own premises.	This ensures that a risk assessment of the security of all data processing is done before processing any personal data. This needs to be done in order to be compliant with EU data protection law
(E) Parties agree that the following processing of personal data will be performed by the following parties: **CEA-France** will conduct the following processing of personal data: via image recognition and text mining techniques CEA will infer potential preferences for specific objects, places and brands. No personal data of DataBait Users will be	This provides the DataBait user with some general transparency about the personal data processing performed by each USEMP partner and who is liable if any data are unlawfully processed. Such transparency is mandatory in order to be

stored at the premises of CEA, that will be authorized to run its algorithms on the data stored at HWC. **iMinds Belgium** will conduct the following processing of personal data: together with CERTH and LTU, iMinds will prepare a survey asking registered users of the USEMP platform and the DataBait tools to answer a set of questions about their lifestyle preferences, selected health issues and personality traits, religious and political beliefs, sexual orientation, gender, age, place of residence and ethnic background. iMinds will conduct the survey to enable testing of how the inferences drawn from DataBait Users' postings, social graphs and behavioural data match their real preferences and background. The outcome of the survey feeds into the database that is stored at HWC. iMinds can access the result of the survey based on secured authorization. The transmission of these sensitive data will be done in a secure way by means of appropriate security protocols. iMinds will also conduct user interviews which contain personal user's information. Interviews will be anonymized, transcribed and stored in an appropriately secured server, only accessible to authorized iMinds personnel. **CERTH-Greece** will conduct the following processing of personal data: via image, text mining and behavioural profiling techniques (involving the 'likes' and sharing of Facebook pages and visits to URLs) CERTH will make inferences about undisclosed demographic characteristics (gender, age, origin), place of residence, sexual orientation, personality and health traits, as well as potential lifestyle preferences, including those that may interest specific types of brands and enterprises. When developing the DataBait tools, a small portion of DataBait User data will be stored at CERTH. In that case appropriate security protocols will be in force, considering the nature of the data. Data will be deleted or fully anonymized once they are no longer necessary for developing the DataBait tools. CERTH will be authorized to run its algorithms on the data stored at HWC. **HWC-UK** will conduct the following processing of personal data: all data collected through the DataBait tools are directed to and stored at HWC, who will secure the data and provide secure access to the USEMP partners for the sole purpose of scientific research as specified in the DLA contract and the description of work that is part of the Grant Agreement with the EU. During	compliant with EU data protection law.

storage at HWC appropriate security protocols will be in force concerning storage and access. Data will be deleted or fully anonymized as soon as the scientific purpose as stated in the DLA agreement is fulfilled.

LTU- Sweden will conduct the following processing of personal data: together with CERTH and iMinds, LTU will prepare a survey asking registered users of the USEMP platform and the DataBait tools to answer a set of questions about their lifestyle preferences, selected health issues and personality traits, religious and political beliefs, sexual orientation, gender, age, place of residence and ethnic background. LTU will conduct the survey to enable testing of how the inferences drawn from DataBait Users' postings, social graphs and behavioural data match their real preferences and background. The outcome of the survey feeds into the database that is stored at HWC. LTU can access the result of the survey based on secured authorization. The transmission of these sensitive data will be done in a secure way by means of appropriate security protocols. LTU will also conduct user interviews which contain personal user's information. Interviews will be anonymized, transcribed and stored in an appropriately secured server, only accessible to authorized LTU personnel.

VELTI-Greece will conduct the following processing of personal data: based on the inferences made by CEA and CERTH, VELTI will conduct further processing operations to visualize information on potential inferences to be provided to the DataBait users. Velti will also use historical Facebook and behavioural data of DataBait users, stored at HWC, for the estimation of the (monetary) value of the personal data of the DataBait users. Some of this data may be retrieved from HWC and stored temporarily at VELTI for preliminary testing. In that case appropriate security protocols will be in force, considering the nature of the data. Data will be deleted or fully anonymized as soon as the purpose of such testing is achieved.

SKU Radboud University-the Netherlands will not conduct any processing of personal data.

(F) Each party that processes personal data hereby exempts all other parties from liability for any unlawful processing of personal data, and from processing personal data in violation of the USEMP DLA or this PDPA. Thus parties will not be severely liable for violations committed by other parties.	Because all the USEMP partners are joint data controllers, each partner is severally liable for any unlawful processing in the USEMP project, this clause aims to limit such liability.
(G) Belgium law will be applicable to this contract.	
Signature page USEMP PDPA Date Place Name/function Signature (1) CEA-France (2) iMinds-Belgium (3) CERTH-Greece (4) HWC-UK (5) LTU- Sweden (6) VELTI-Greece (7) SKU Radboud University-the Netherlands	

Table 1. Text of the USEMP PDPA.

2.5.2. DLA

	Why is this clause important?
USEMP Data License Agreement (DLA) The parties: (1) [………………………………………………………….], user of the USEMP platform and services, from hereon called 'You' and (2) [CEA-France / iMinds-Belgium/ CERTH-Greece / HWC-UK/ LTU- Sweden /VELTI-Greece/ SKU Radboud University-the Netherlands],[11] provider of the USEMP platform and services, joint data controllers, from hereon called 'USEMP consortium partners'.[12] Hereby agree:	This DLA is the legal ground (art. 7 DPD) for all processing of personal data in USEMP. Establishing such ground is necessary in order to be compliant with EU data protection law
(A) You will install the USEMP DataBait tools, the DataBait-Facebook app and the DataBait web browser plug-in and the DataBait graphic user interface (GUI). The DataBait-Facebook app and the DataBait web browser plug-in will provide access to Your Facebook profile and Your browsing behaviour on Your device(s). These tools will be used by the USEMP consortium partners to collect data that You share on Facebook as well as data collected by the web browser. This data can be data You posted (volunteered data), or data captured by the USEMP tools (observed data). The latter concerns online behavioural data (storing what You did on the Internet and on FaceBook).	This establishes the legal ground (art. 7 DPD) for the collection of OSN and browser data (input data) through the DataBait tools. This is necessary in order to be compliant with EU data protection law

[11] Each partner will provide a hyperlink, such that users can click and check who is involved. CEA: http://www.kalisteo.fr/en/; iMinds: http://www.iminds.be/en/about-us/organizational-structure/research-departments/digital-society-department/iminds-smit-vub; CERTH: http://www.iti.gr/iti/index.html http://www.iti.gr/iti/index.html LTU: http://www.openlivinglabs.eu/node/125; VELTI: http://www.velti.com/; SKU: http://www.ru.nl/icis/.

[12] Click through on "USEMP Consortium Partners" will show the following: "The USEMP consortium partners have entered a separate agreement, obliging themselves and each other to act in accordance with this contract, their national data protection law and EU data protection law, in which agreement they clarify which partners processes what personal data. This contract can be accessed here."

(B) You license the use of Your volunteered and observed personal data by the USEMP consortium partners, as gathered by the DataBait-Facebook app and the DataBait web browser plug-in *for the sole purpose of scientific research* and – within that context – to provide You through the DataBait graphic user interface (GUI) with information about what third parties might infer based on Your sharing of information, and on Your online behaviour. The said data may be combined with publicly available personal data gained from other sources to infer more information about Your habits and preferences (inferred data).	This specifies the purpose of the data processing within the USEMP project. This is necessary in order to be compliant with EU data protection law
(C) This license agreement confirms Your explicit consent to store the DataBait tools on Your devices.	This establishes the legal ground (art. 7 DPD) for placing the DataBait tools on the device of the user. This also includes tracking cookies or similar tracking mechanisms (as described in art. 5.3 ePrivacy Directive) which are necessary to fulfil the functionality of the DataBait service.
(D) The USEMP consortium partners will do scientific research to predict what kind of information Facebook or other third parties with access to Your postings and online behavioural data *could or might* infer from the said data. These inferences will be shared with You in an intuitive manner, thus providing an online presence awareness tool, embedded in the “DataBait-GUI”.	This expresses how empowerment through profile transparency is achieved in the DataBait tool. It regards the transformation of your OSN and browser data (input data) into so-called data derivatives (output data).
(E) The USEMP consortium will also do scientific research to estimate the monetary value of Your data, based on the said data and their inferences. The “DataBait-GUI” will alert You that some of Your online behaviours *may* be monetisable, for example in the context of personalized advertising or in the context of selling Your data or profile to data brokers, credit rating companies or others willing to pay for access to the data or inferred profiles. This way the DataBait-GUI also acts as an economic value awareness tool.	This expresses how empowerment through profile transparency is achieved in the DataBait tool. It regards the transformation of your OSN and browser data (input data) into so-called data derivatives (output data).
(F) You agree to participate in surveys and/or focus groups, to enable the consortium to gain insights in how users engage with social networking sites and how they evaluate (1) various scenarios regarding the use of their personal data and targeted profiles and (2) the effectiveness, usability and utility of the USEMP tools.	This establishes the legal ground (art. 7 DPD) for the collection and processing of the survey data. This is necessary in order to be compliant with EU data protection law.

(G) You hereby grant Your consent to process Your sensitive personal data, notably those revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and those concerning health or sex life.	This explicit consent is the legal ground (art. 7 DPD) for the processing of the sensitive personal data of the DataBait user. Because the legal ground (art. 7 DPD) is consent (and not contract, as for all other personal data) the user can also withdraw this consent at any moment.
(H) The USEMP consortium partners will treat all Your personal data, especially Your sensitive data, with care and delete or anonymize them as soon as possible. Because one of the main goals of the USEMP project is to create awareness about the possibility to infer sensitive data from trivial data trails, it is important to alert You to such inferences and thus to process them.	This expresses that all processing is done according to the principle of data minimization. This is mandatory in order to be compliant with EU data protection law. It also expresses how empowerment through profile transparency is achieved in the DataBait tool, by transforming of your OSN and browser data (input data) into so-called data derivatives (output data).
(I) The USEMP consortium partners will process Your personal data in a secure way and not keep them any longer than necessary for the purpose of the USEMP study. In order to provide You with access to Your personal data and the inferences drawn from them, the data may be kept until the end of the project. Within 3 months of the ending of the research project (1 October 2016), all personal data will be either deleted, anonymised or processed for related scientific research. In the latter case the relevant USEMP consortium partner will ask You for Your consent.	This expresses that all processing is done according to the principle of data minimization. This is mandatory in order to be compliant with EU data protection law.
(J) The USEMP consortium partners will not provide Your personal data to any third party other than the *Future Internet Research and Experimentation Initiative* (FIRE) infrastructure, which is a multidisciplinary scientific infrastructure funded by the EU in which novel internet related tools can be tested and validated. The transfer of the data will happen in a secure way and only in as far as strictly necessary for the scientific goals of the USEMP project.	This ensures that the transfer of personal data to the FIRE infrastructure is compliant with EU data protection law and the principle of use limitation: that data should not be further processed in a way incompatible with the initial purpose of the processing. Use limitation is part of the principle of data minimization (art. 6 DPD 95/46).
(K) The national law of Your country of residence (at the moment of registration) is applicable to this contract, assuming you are a resident of the EU.	

By clicking the box below You become a party to this agreement: ☐	

Table 2. Text of the USEMP DLA.

2.5.3. Table of Types of Personal Data Processed in the USEMP project – with hyperlinks to the full listings in Annex A

PERSONAL DATA OF DATABAIT USERS WHICH ARE PROCESSED IN THE CONTEXT OF USEMP	INPUTDATA	A. Personal data collected with the DataBait OSN app	See table 6.1.1 for the most up-to-date list (and table 6.1.3 for the original list with codes) Also see annex D in D7.1 and D2.3
		B. Personal data collected with the DataBait browser plugin	See table 6.1.2 Also see annex D in D7.1 and D2.3
	OUTPUT DATA ("DATA DERIVATIVES")	C. Personal data *inferred* from a subset of the data collected through the OSN app [A] and the browser plugin [B]	See table 6.1.5 the list with derived data list (and table 6.1.4 for the OSN and browser data used for training the USEMP data-driven modules which produce these data derivatives) Also see D6.1 and D2.3

	TRAINING AND TESTING DATA FOR THE CONSTRUCTION OF THE SOFTWARE USED TO INFER THE OUTPUT FROM THE INPUT DATA	D. Personal data collected in the DataBait surveys in the pre-pilot. These data are used to establish so-called *ground truth* (what are the true values of a user – e.g. age, sexual identity, economic status, etc.? only when the true values are known, is it possible to establish how well the classifiers constructed in the USEMP project manage to predict these values) and assign scores to how sensitive a certain type of information is considered by the "average" user.	See the survey in "D4.5: User Categorisation of Digital Footprint - V2" (ready in : M24)
PERSONAL DATA OF *NON-*DATABAIT USERS WHICH ARE PROCESSED IN THE CONTEXT OF USEMP		E. Personal data in training and testing sets, used to train and test classifiers (i.e., models used to predict and infer data from the input data collected in the DataBait tools). While most data in these training and testing data sets are not personal data (they are anonymized or do not relate to an identified or identifiable person), each data set has to be screened for the presence of personal data. Also, it should be noted, that the fact that most of these data are *not* derived from DataBait users does not mean that the scrutiny in terms of data protection (in as far as these data sets contain personal data) should be any less.	See table 6.1.6 (and table 6.1.4 for the OSN and browser data used for training the USEMP data-driven modules which produce these data derivatives) Also see D2.3.

Table 3. Overview of USEMP personal data ordered according to source – reference table.

2.5.4. Table of Data Types processed: specification of the premise, technical goal, storage period and method

Personal data processed in the USEMP project, ordered according to source:	**Premise**	**Technical goal of processing?**	**Storage period**	**Anonymization/pseudonomization? (if, when)**

A. Personal data collected with the DataBait OSN app	HWC, CERTH, VELTI	1. Representing the data in the DataBait GUI to give the DataBait user more insight in her digital trail 2. Inferring other knowledge from the data to give the DataBait user more insight in her digital trail	At most until three months after the end of the USEMP project. Temporary access to train models	1. At the end of the USEMP project HWC deletes all data - outside the project they have no use for such data, and even with anonymisation or pseudo-anonymisation there still would be a risk in holding such data. 2. During the project there are no plans to anonymise or pseudonymise the data kept at HWC,[13] though much of the data is stored in a segregated state - e.g. imagery data is kept separate from user profile data, and without the profile data, the information they provide is simply the image itself. Similarly, survey data is segregated from profile data and OSN data although the survey and OSN data of course have personally identifying data within them. 3. HWC will, however, provide CERTH and VELTI with pseudonymized data for temporary usage in the pilots, pre-pilots and pre-pre pilots at their own premises.

[13] The European Parliament's version of the proposed General Data Protection Regulation introduces this notion of pseudonymous data, which it defines as "personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution." (art. 4.2a GDPR)

B. Personal data collected with the DataBait browser plugin	HWC, CERTH, VELTI	1. Representing the data in the DataBait GUI to give the DataBait user more insight in her digital trail 2. Inferring other knowledge from the data to give the DataBait user more insight in her digital trail	At most until three months after the end of the USEMP project. Temporary Access to train models	See above
C. Personal data collected in the DataBait surveys in the pre-pilot.	HWC, CERTH, VELTI	1. Finding the 'true values' (ground truth). These declared data help to assess how well the classifiers developed in USEMP are able to predict/infer these values. 2. Exploring which values users consider to be sensitive.	At most until three months after the end of the USEMP project. Temporary Access to train models	See above
D. Personal data *inferred* from a subset of the data collected through the OSN app [A] and the browser plugin [B]	HWC, CERTH, VELTI	Providing inferred knowledge to the DataBait user in the GUI to give her more insight in her digital trail and possibilities to control this information.	At most until three months after the end of the USEMP project. Temporary Access to train models	See above

E. Personal data in training and testing sets, used to train and test classifiers	HWC, CERTH, VELTI	Needed to build classifiers which can infer/predict certain attributes and their values based on the data gathered through the DataBait tools	Varying (needs to be further explored)	See above

Table 4. Detailed usage of USEMP data ordered according to source.

2.5.5. Data protection requirements for data processed in USEMP

If data is *legally qualified* as……,	**….,then the *legal effect* is….**	**…which results in this *legal requirement* :**
PD : Personal data as defined in DPD 95/46	The regime of data protection directive 95/46 applies. I. Following from Art. 12 of DPD 95/46 (in conjunction with Art. 15) and anticipating the proposed new EU data protection law (Recital 51 and Art. 15 of the pGDPR), the DataBait user has the following « Informational rights » (which includes the so-called right to « profile transparency »), which entail he or she should be informed about : • the purpose for which the data are processed • what categories of data are processed, • for what estimated period, • which recipients receive the data, • what is the general logic of the data that are undergoing the processing, • what might be the consequences of such processing, • the existence of the right to request rectification or	I. A button which the DataBait user can click with all the information that needs to be given following the informational rights from directive 95/46. The button on the USEMP platform, and include an email address for each partner that processes personal data, to make further inquiries. The information will be updated whenever the relevant processing of personal data change.

	erasure of the data concerning the data subject and of the right to object to the processing, • the right to lodge a complaint to the supervisory authority and the contact details of the supervisory authority? II. A purpose for the processing has to be specified (Art. 6 DPD 95/46) III. The processing has to be based on a ground legitimizing the processing. The ground used in USEMP is « contract » (Art. 7(b) DPD 95/46) IV. The data should not be kept longer than necessary and be deleted or completely anonymized (no re-identification possible) when no longer needed (i.e. at the end of the USEMP project). V. Security of the processing needs to be adequate using information assurance principles VI. Anticipating the new EU data protection law (Art. 8, pGPDR) : a mechanism which checks	II. Purpose described in Data Licensing Agreement and also available under informational button III. Data Licensing Agreement : The legal ground used in USEMP is 'contract' (Art. 7(b) DPD 95/46), for downloading the DataBait tools consent (art. 6.3 e-Privacy Directive 2002/58) and for processing LPD again consent art. 8 DPD 95/46 IV. The data should not be kept longer than necessary and be deleted or completely anonymized (no re-identification possible) when no longer needed (i.e. at the end of the USEMP project). A risk assessment investigating the assurance that storage, processing

	the age of DataBait users and does not allow children (below the age of 13 ?) to use it. VII. Anticipating the new EU data protection law (pGPDR): implement legal protection by default and by design as much as possible VIII. Anticipating the new EU data protection law (preamble of the pGPDR, stating that data protection is not an absolute right but that it should be balanced with other rights). IX. Notification of national data protection authority of processing of the data	and transfer of information is carried out with appropriate and agreeable technical, logical and physical security measures. V. Anticipating the new EU data protection law (Art. 8, pGPDR): a mechanism which inquires after the age of DataBait users and does not allow children (below the age of 13) to use it and gives a warning to anyone aged 13-18. VI. Anticipating the new EU data protection law (pGPDR) : implement legal protection by default and by design as much as possible: all of the above but also [following current law and the principle of data minimization] for example check default settings and try to pseudonymize, anonymize etc. when it is not strictly necessary to have fully identifiable personal data. VII. The contract (DLA) provides a more balanced approach – creating mutual duties and rights - than mere consent.

		VIII. Notification of national data protection authority of processing of the data
LSD : Legal sensitive data as defined in Art. 8 DPD 95/46 = personal data revealing : – racial or ethnic origin, – political opinions, – religious or philosophical beliefs, – trade-union membership, and the processing of data concerning health or sex life, and – the processing of data relating to offences, criminal convictions or security measures	– Specific consent - Exploring whether sensitive data (Art. 8 DPD 95/46) are used as the sole ground for profiling and preferably avoid it [This is not current law and it is up for debate whether a prohibition of such profiling solely based on sensitive data will make it into the pGPDR]	- Making sure that the DataBait tool asks the users for explicit consent [Clause G of the DLA takes care of this.] - A button where this consent can be withdrawn : Each party will also provide an email address to be contacted in case a user wants to withdraw her consent for processing her sensitive data; this is preferably the same email address as the one used to gain further information, but will be available behind a separate button on the USEMP platform. - Check whether any of the inferred data in the USEMP project are solely based on sensitive data
PROFILE-INPUT : Data used as input for profiling	- Exploring whether sensitive data (Art. 8 DPD 95/46) are used as the sole ground for profiling and preferably avoid it [This is not current law and it is up for debate whether a prohibition of such profiling solely based on sensitive data will make it into the pGPDR] - Making sure no measures which have a significant or legal impact are taken based on the	Check whether any of the inferred data in the USEMP project are solely based on sensitive data Although the profiling performed through the DataBait tools is not likely to result is measures which have a significant or legal impact in a narrow sense, we interpret "significant" in a broad sense.

	profiling, unless there is a contract or consent.	The DLA (contract) provides the legal ground.
PROFILE-OUTPUT : data which result from profiling	This data subject has the right to obtain knowledge of the logic involved in any automatic processing which significantly affects him or her (Art. 15(1) in conjunction with Art. 12(a) of the DPD 95/46). It is not completely clear how "significantly" should be defined, but to be on the safe side we give the term a broad interpretation. Making sure no measures which have a significant or legal impact are taken based on the profiling, unless there is a contract or consent. It is not completely clear how "significant" should be defined, but to be on the safe side we give the term a broad interpretation.	- The informational button and the DataBait GUI should provide insight in the logic involved in the profiling (which knowledge in inferred from which data, how is this done, how reliable is this knowledge, etc.) - Although the profiling performed through the DataBait tools is not likely to result in measures which have a significant or legal impact in a narrow sense, we interpret "significant" in a broad sense. The DLA (contract) provides a legal ground.
LD: location data as defined in e-Privacy Directive 2002/58.	- The legal status of location data is the subject of some controversies, but to be on the safe side we assume that the regime as applicable to personal data (PD) applies. Thus, see above.	- See above, same requirements as with PD.

*Table 5. **The "answer" to almost all these requirements is the PDPA (which includes the DLA).** The legal requirements are based on the legal qualification of data processed in USEMP as **personal data** – which includes (a) "ordinary" personal data, (b) personal data which are sensitive (Art. 8 DPD 95/46), and (c) personal data which are the input or output to profiling, i.e. data used to infer other data or inferred data; where "profiling" (defined in the pGPDR) is a particular type of "automated processing" (see DPD 95/46) - or **location data** (as defined in e-Privacy Directive 2002/58)*

3. The information & consent withdrawal buttons: Data Life-Cycle Management

Following from the legal requirements, the DataBait tool will contain two “buttons” which embody two specific legal requirements described in the PDPA:

1. A button which embodies the informational rights of the DataBait user and which, when clicked, provides the information required by data protection law.
2. A button which allows the DataBait user to withdraw consent – behind this button the user will be instructed how to remove the DataBait tools (including any cookies) from her devices and the USEMP Consortium Partners will terminate the further processing of her personal data.

Despite the fact that EU data protection law is quite explicit about what a data subject should be informed about (Art. 10 and 11 DPD 95/46) and to which information access should be granted (Art. 12 DPD 95/46), there is often no unequivocal rule how this information should be presented. In collaboration with the other partners we develop ways to present this information in a way that is compliant with data protection law and should be truly empowering. The first version of the information buttons will be discussed with the users of the DataBait tools in the (pre-)pilot stage during the users interviews performed by iMinds and LTU.

While the exact format and presentation will need to be discussed in more detail, the following information will be available behind the informational button:

1. The DLA and PDPA with hyperlinks to the requirement and personal data tables;
2. Information about the possible tensions between data protection and IP rights of profilers and those of the end users. It explains that the PDPA is a mutual obligatory agreement which creates mutual legal duties and rights, providing for a more balanced approach in terms of power equality between the contracting parties than mere consent – which is the legal ground used most frequently in such situations.
3. A flow chart (like the ones presented in this deliverable; see e.g. figure 1 and 2) explaining the data flows in the USEMP project. This will be accompanied by a concise but clear explanation of how the raw OSN and browser data collected through the DataBait tools are transformed into data derivatives with the data-driven modules. Section 5 of Deliverable 2.3 will be of particular use to visualise the relevant data flows.

As to the withdrawal of consent, as required by the e-Privacy Directive for installing the DataBait tools (including any cookies or other tracking mechanisms), and as required by the DPD for processing the sensitive data, we have decided that withdrawal of either consent should result in the de-activation of both the DataBait tools and the termination of the processing of the user’s personal data. The reason is that any other option would be far too complex, since the purpose of the DataBait tools is to produce data derivatives that will often qualify as sensitive data in the legal sense. The form of de-activation will be decided by the technical team for both the OSN application and the Web plugin and should result in a complete stop in the collection of the corresponding data (for example the Databait Web plugin will cease its operations of collecting data if user has withdrawn his/her consent even if the user retains an account with the Databait platform).

In point of fact the actionable information behind these buttons visualises the USEMP Personal Data Life-Cycle Management, while enabling engagement in such management by the user. In view of upcoming legislation on personal data portability, this is an example of how to provide multilevel, clickable and actionable tools that provide easy to use comprehensive means of engagement for end-users. Basically the buttons combine front-end transparency with the opportunity to gain a detailed form of back-end transparency.

4. Anti-discrimination law and data processed in USEMP

All data processed in USEMP have been qualified in terms of EU-discrimination law in the tables in Annex A. In D3.1 some preliminary suggestions for legal requirements following from these qualifications have been made. This will be further elaborated in D3.6.

5. IP rights on data processed in USEMP

Apart from data protection and anti-discrimination law, intellectual rights might also be applicable to different aspects of the work performed in USEMP. There are different types of intellectual rights that might be relevant here, like: copyright, *sui generis* data base rights, patents, trademarks, trade secrets and portrait rights (see Deliverables D3.2 & 3.3). These might also apply to different objects like: individual user data, database structures, profiling algorithms and perhaps the user profiles ("the data derivatives") themselves.

(1) Copyright in User Content & Databases

In Annex B, tables B.1 and B.2, of this deliverable we make a first inventory of the user data processed in USEMP which might be protected by copyright. We could think here of news, text, videos and images accessed through the browser and videos, statuses, photos, and posts on the OSN profile (table B.1). When such data are processed in the USEMP project, their processing is likely to entail that some (technical) copy is made. Such reproduction of copyrighted content requires permission (a 'license') from the author or another license holder which has the right to transfer the license to others (such as the OSN – see D3.2 for a discussion of the transferrable license each Facebook user grants to Facebook). Furthermore, the content in the external data sets used for training and testing of the classifiers in the data-driven modules are also likely to be protected by copyright (see table B.2 in the Annex). Moreover, the structure of the data set might also be protected by copyright (when the selection and arrangement of contents is original and does not aim to be exhaustive). The database, or substantive parts of it, might also be protected by the *sui generis* database right (when a substantial investment in the obtaining, verification or presentation of the contents has been made). Questions that need to further be explored in D3.9 with regard to data processed in USEMP and that are potentially protected by intellectual property rights, notably copyright are:

- What contents are (partially) copied?
- How long will the content be stored? (possibly relevant with respect to the exception for temporary technical copies)
- For what purposes will the content be used? (This is particularly interesting when the content is not exploited for its original content – which is the case in traditional exploitation of IP-protected rights – but in an aggregated way and/or to stimulate data traffic)
- Who has access to the stored content? Is the content shared with third parties (outside your organisation)?
- What is the source of the IP protected content? Did you find it on publicly accessible sources (such as websites or search engines)? Or did you acquire the content from a "private" source?
- How are they derived? Did you technically acquire the files containing the content? Did the source hand over the files or did you acquire these at your own initiative (e.g., through an API or through scraping?)
- Is there a license to use the IP protected content (e.g., an image, the database structure of a database with images or a status update)? If yes: what are the conditions of use?
- How are the data deleted (manually or automatically)? Or are any of the data re-used for internal or external use?

(2) Copyright & patents in the data-driven software modules:

A second issue that needs to be further explored in D3.9 is whether the software used in USEMP, which were adapted from previous work or entirely developed in the context of USEMP, are covered by intellectual rights like copyright or patents.

This issue is important both from the perspective of compliance with IP law by USEMP partners, but also from the perspective of user empowerment. Profile transparency is one of the key goals of USEMP. This also implies showing the user how the profiles have been made, for instance by offering them information about the 'logic of processing' utilized in the user profiling (art. 12 of the current Data Protection Directive). If the algorithms cannot however be tested by third parties due to trade secrets or IP rights on the software, transparency becomes a pure matter of trust on the side of the user, which is problematic. Furthermore, the Data Protection Directive in this context states that this right of access to the logic of profiling "must not adversely affect trade secrets or intellectual property and in particular the copyright protecting the software; whereas these considerations must not, however, result in the data subject being refused all information" (recital 41 of the Preamble). A balance should thus be struck here. It is thus also necessary to gain a better view on the extent to which the profiles we will show users have been inferred with algorithms originally trained/developed outside the context of USEMP (e.g. in a proprietary setting).

At this point in time we point out that the **CEA**'s algorithms were developed internally and have IP rights that can make them reusable/adaptable in USEMP.

In more detail we note in relation to D5.1 (Text mining):

- Text similarity (described in 2.2.1 to 2.2.4) is adapted from previous work in French projects in order to work for the USEMP languages and domains.
- Location detection from texts is adapted from some work that is not directly related to projects and can be reused.

For D5.2 (Image mining) all algorithms were developed within USEMP, relying on previous works in French projects or not directly related to a project, and can be reused in all cases. For D5.3 (Multimedia fusion) all algorithms were developed within USEMP. D5.2 (Image mining) and D5.3 (Multimedia fusion), all algorithms were developed within USEMP.

Regarding the modules that are developed in T6.1 and that perform a number of inferences (e.g. behavioural prediction based on likes), the used methods are adapted from previous work that has appeared in the relevant research community and can be reused.

CERTH's algorithms were mostly adapted from previous work and some were entirely developed in the context of USEMP. In all these cases, there has been an implementation by CERTH of a variation of an existing (published) algorithm, or a combination of different algorithm implementations into a more complete analysis process. In particular:

- The implementation of the text-based location detection module (described in D5.1) is an adaptation of work from the SocialSensor and REVEAL projects.
- The implementation of the visual-based location detection module (described in D5.2) is an extension of work from the SocialSensor project. Due to its unsatisfactory performance, we will not use this module in the system.
- The implementation of the supervised relevance and diversity reranking module (described in D5.3) is a (significant) extension of work from the SocialSensor project.

- The implementation of the likes-based inference mechanism (the method by Kosinski, PNAS 2013) that is described in D6.1 was fully implemented and evaluated in the context of USEMP.
- The implementation of the topic-based user classification (described in D6.1) was fully implemented and evaluated in the context of USEMP.
- The implementation of the network-based user classification (described in D6.1) is an adaptation of work from the REVEAL project.
- The implementation of the social circle detection module (described in D6.2) was fully implemented and tested in the context of USEMP. [most probably this module will not be possible to use in the actual system]
- The implementation of the private/public image classification (described in D6.2) was fully implemented and tested in the context of USEMP.

The training for all the above modules was conducted in the context of USEMP.

Furthermore, if needed, any detail about these algorithms and even the source code can be disclosed to improve the trust of users. An even more ambitious idea (which would require additional development on the front-end and documentation effort) is to give the possibility to users (e.g., via pop-up dialogs) to read very detailed technical information regarding how a particular inference (that they currently see on the interface) was made. This could provide an interesting instantiation of the right to “knowledge of the logic involved in any automatic processing of data” concerning the data subject, granted by European data protection law (article 12a DPD).

6. Concluding remarks and planned further research

This report has presented the results of the legal coordination and the integration during the first half of the USEMP project. With regard to data protection law it offers legal qualification of all the data that is handled by the USEMP system and the legal requirements which follow from this. The main achievement of this deliverable is the hyperlinked architecture of the PDPA and DLA that links them to the tables, lists and flow charts. To clarify the processing operations in the backend of the USEMP platform, a user friendly version of this architecture will be presented behind the so-called 'information button' in the DataBait tool. This is how USEMP will ensure that all processing of the data is compliant with EU data protection law but – even more importantly - that the freedom of the user towards OSNs and browsers is strengthened. Profile transparency such as the one provided behind the information button can help data subjects to exercise their fundamental right to data protection (empowerment).

In the tables of Annex A and B some (preliminary) qualifications of the data in terms of anti-discrimination law and intellectual property law are presented but further research is needed. This will be presented in deliverable D3.6 (data protection and antidiscrimination law), D3.7 (intellectual property rights of the OSN or browser), D3.8 (intellectual property rights of the end-users of OSNs and browsers) and after integration with the technical partners the implications will be presented in the second version of the current report, D3.9. Further research is also needed with regard to Twitter, the second OSN which will be studied in the USEMP project.

The main challenge will be to develop a modular DLA that provides different options to DataBait users when employed in a commercial context: (1) if the DataBait tools are provided by an independent commercial enterprise it must develop a business model, which means that the purpose of processing will not involve more than research but generating economic value, (2) if the DataBait tools are provided by the OSN service provider the tools will become the means to comply with the legal obligation to provide profile transparency, but the provider may call upon its trade secrets and IP rights to restrict the transparency.

Annex A: Legal requirements of data protection and anti-discrimination law

Table A.1. PD collected with DataBait OSN app[14]

1. Automatically Allowed Permissions	**An app may use this permission without review from Facebook.**	**To which USEMP (inferred) data listed in table A.3 (see below) does this data contribute? The codes (C1, C2, etc.) are explained in table A.3.**	**Used for inferences?**	**Legal qualification: EU data protection (DP) law EU anti discrimination (AD) law[15]**
Public profile	Access to a subset of items that are part of a person's public profile. A person's public profile refers to the following properties on the user object by default:	This contributes to C7 (User Profile* and Interests).		
	Id (the number of the profile, e.g. “"1424672444497579")		No	*DP*: PD
	Name (full name of the user)		No	*DP*: PD ; Does the name reveal race or ethnic origin? Then it could be

[14] In the Annex the term Personal Data will be abbreviated to PD and will always refer to PD in the legal sense.

[15] The protected grounds according to EU data protection law are: sex, racial or ethnic origin, religion or belief, disability, age, sexual orientation and nationality. See chapter 6.2 of D3.1.

				LSD; *AD*: If this data reveals race or ethnic origin: differentiation based on race or ethnic origin is prohibited in the fields of employment, access to good and services, social advantages, social protection and education
	first_name (first name of the user)		No	*DP*: PD ; Does the name reveal race or ethnic origin? Then it could be LSD; *AD*: if this data reveals race or ethnic origin: differentiation based on race or ethnic origin is prohibited in the fields of employment, access to good and services, social advantages, social protection and education
	last_name (last name of the user)			*DP*: PD ; Does the name reveal race or ethnic origin? Then it could be LSD; *AD*: if this data reveals race or ethnic origin: differentiation based on race or ethnic origin is prohibited in the fields of employment, access to good and services, social advantages, social protection and education.

	link (link to the Facebook profile, e.g.: https://www.facebook.com/app_scoped_user_id/1424672444497579/)		No	*DP*: PD
	gender (gender of the user)		No	*DP*: PD *AD*: Differentiation based on gender in the field of employment and the access to goods and services is prohibited
	locale (locale/language, e.g. "en_GB", which stands for British English)		No	*DP*: PD
	timezone (timezone of the user)		No	*DP*: PD
	updated_time (the time of the most recent update) verified (is the Facebook		No	*DP*: PD
	verified (is the Facebook account linked to a verified phonenumber and/or email address?)		No	*DP*: PD
user_friends	Access the list of friends that also use	C4 (Friends-list)	Yes See: C4/D6	*DP*: PD ; PROFILE-INPUT ;

	your app. (this is commonly used to create a social experience in your app.)			the PROFILE-OUTPUT based on these data *could be* LSD – depending on the content of the inference made. *AD*: the PROFILE-OUTPUT data *could* include protected grounds – depending on the content of the inference
Email	Access to a person's primary email address.		No	*DP*: PD
2. **Requested[16] extended permissions**	These permissions are not optional in the login dialog during the login flow, meaning they are non-optional for people when logging into your app. If you want them to be optional, you should structure your app to only request them when absolutely necessary and not during initial login.			
user_about_me	Access to a person's personal description	This contributes to C7 (User Profile* and Interests)	Maybe	*DP*: PD ; these data *could be* LSD –

[16] Facebook still has to give permission

	(the 'About Me' section on their Profile) through the bio property on the User object.			depending on the content; maybe PROFILE-INPUT *AD*: these data *could* include protected grounds – depending on the content
user_activities	Access to a person's list of activities as listed on their Profile. This is a subset of the pages they have liked, where those pages represent particular interests.		Maybe	*DP*: PD; these data *could be* LSD – depending on the content; maybe PROFILE-INPUT *AD*: these data *could* include protected grounds – depending on the content
user_education_history	Access to a person's education history through the education field on the User object.	This contributes to C7 (User Profile* and Interests).	Maybe	*DP*: PD ; maybe PROFILE-INPUT
user_hometown	Access to a person's hometown location through the hometown field on the User object. This is set by the user on the Profile.	This contributes to C7 (User Profile* and Interests).	No	*DP*: PD
user_interests	Access to the list of interests in a person's Profile. This is a subset of the pages they have	This contributes to C7 (User Profile* and Interests).	Maybe	*DP*: PD ; these data *could be* LSD – depending on the content; maybe PROFILE-INPUT

	liked which represent particular interests[17].			*AD*: these data *could* include protected grounds – depending on the content
user_likes	Access to the list of things a person likes. Provides access to the list of all Facebook Pages and Open Graph objects that a person has liked.	C2 (Likes and Dislikes)	Yes See :C2/D1	PD ; these data *could be* LSD – depending on the content ; PROFILE-INPUT; the PROFILE-OUTPUT based on these data *could be* LSD – depending on the content of the inference made. *AD*: these data *could* include protected grounds – depending on the inferred content or the topic to which the likes refer.
user_location	Access to a person's current city through the location field on the User object. The current city is set by a person on their Profile.	This contributes to C7 (User Profile* and Interests).	Maybe	*DP*: PD ; maybe PROFILE-INPUT
user_photos	Access to the photos a person has uploaded or been tagged in. This is available through the	C3 (Photos) Contributes to C5 (Friends' activities upon user's OSN objects)	Yes See : C3/D5 C5/D7	*DP*: PD ; these data *could be* LSD – depending on the content ; PROFILE-INPUT ; the PROFILE-OUTPUT based

[17] The user_interests permission is deprecated. On Tuesday, June 23, 2015, this permission request will be silently ignored. Please see Facebook's changelog for more information.

	photos edge on the User object.			on these data *could be* LSD – depending on the content of the inference made. *AD*: the PROFILE-OUTPUT data *could* include protected grounds – depending on the inferred content
user_relationships	Access to a person's relationship status, significant other and family members as fields on the User object.	This contributes to C7 (User Profile* and Interests).	No	*DP*: PD ; LSD *AD*: sexual orientation is a protected ground in the field of employment
user_relationship_details	Access to a person's relationship interests as the interested_in field on the User object.	This contributes to C7 (User Profile* and Interests).	No	*DP*: PD ; LSD *AD*: sexual orientation is a protected ground in the field of employment
user_religion_politics	Access to a person's religious and political affiliations.	This contributes to C7 (User Profile* and Interests).	No	*DP*: PD ; LSD *AD*: religious beliefs is a protected ground in the field of employment
user_status	Access to a person's statuses. These are posts on Facebook which do not include links, videos or	C1 (Posts) Contributes to C5 (Friends' activities upon user's OSN objects)	Yes. See: C1/D2 C5/D7	*DP*: PD ; *could be* LSD – depending on the content of the status update; PROFILE-INPUT;

	photos.			the PROFILE-OUTPUT based on these data *could be* LSD – depending on the content of the inference made. *AD*: these data *could* include protected grounds – depending on the (inferred) content
user_tagged_places	Access to the Places a person has been tagged at in photos, videos, statuses and links.			*DP*: PD
user_videos	Access to the videos a person has uploaded or been tagged in.	Contributes to C5 (Friends' activities. upon user's OSN objects)	Yes. See: C5/D7	*DP*: PD ; *could be* LSD – depending on the content of the videos *AD*: these data *could* include protected grounds – depending on the content of the videos
User_groups	A list of groups that a user is a member of	This contributes to C7 (User Profile* and Interests).	Maybe	*DP*: PD ; *could be* LSD – depending on the topic of the group *AD*: these data *could* include protected grounds – depending on the topic of the group
User_work_history	The user's work history	This contributes to C7 (User Profile* and Interests).	No	*DP*: PD

Metadata (which come along with e.g. user_status', 'user_posts' and 'user_tagged_places')				
Location related data e.g. : "place": place of the user who posted the status update "name":name of the location of the user, e.g. a concert hall or the public library "street":street name "city": city name "state":name of state "country":country name "zip":zip code "latitude":latitude "longitude":longitude			No	*DP*: LD
"id": id of the user who posted the status update			No	*DP*: PD

Table A.2. PD collected with DataBait browser plugin

Table A.2.a. PD collected with DataBait browser plugin (browsing behaviors). [18]

#	Name	Description	Is this data used to infer anything?	Legal qualification in terms of EU data protection law and EU anti discrimination law[19]
A1	Site Unique Visits	web sites (URL) visited by the user	Maybe	*DP*: PD ; could be LSD depending on the (inferred) content or the topic of a site ; PROFILE-INPUT *AD*: these data *could* include protected grounds – depending on the (inferred) content or a topic from a site
A2	Site Visits	# of times a user visited a web site (URL)	No	*DP*: PD
A3	Time Spent Per Site	Time a user spent during one visit. (time opened the URL at his browser)	No	*DP*: PD
A4	Images	Images Uploaded/Accessed/Downloaded by the user.	No	*DP*: PD ; could be LSD depending on the content of the image *AD*: these data *could* include protected grounds – depending on the content of the image
A5	Videos	Videos Uploaded/Accessed/Downloaded by the user.	No	*DP*: PD ; could be LSD depending on the content of the video *AD*: these data *could* include protected grounds – depending on the content of the video
A6	Actions	Click on specific element at the web site.	No	*DP*: PD ; could be LSD depending on the content of the specific *AD*: these data *could* include protected grounds – depending

[18] This table is based on annex D of D7.1.

[19] The protected grounds according to EU data protection law are: sex, racial or ethnic origin, religion or belief, disability, age, sexual orientation and nationality. See chapter 6.2 of D3.1.

				on the content of the specific element
A7	Text	Text Uploaded/Accessed at the web site.	No	*DP*: PD ; could be LSD depending on the content of the text *AD*: these data *could* include protected grounds – depending on the content of the text
A8	News	Page views of specific news elements.	No	*DP* : PD ; could be LSD depending on the content of the news element *AD*: these data *could* include protected grounds – depending on the content of the news element

Table A.2.b. PD collected with DataBait browser plugin (trackers)[20]

#	Name	Description	Is this data used to infer anything?	Legal qualification in terms of EU data protection law and EU anti discrimination law[21]
B1	# of Trackers for Site URL	The number of tracking services when a LIO user visits URL	No	*DP*: PD
B2	Tracker	The ID of the tracking services when a LIO user visits a URL	Maybe	*DP*: PD
B3	Tracker email	A Tracker of users email (e.g., google-mail)	No	*DP*: PD

[20] *This table is based on annex D of D7.1.*

[21] The protected grounds according to EU data protection law are: sex, racial or ethnic origin, religion or belief, disability, age, sexual orientation and nationality. See chapter 6.2 of D3.1.

Table A.3. Original list of PD collected with DataBait OSN app[22]

#	Name	Description	Is this data used to infer anything?
C1	Posts Feed [23]	An individual entry in a profile's feed. The profile could be a user, page, app, or group.	Yes
C2	Likes and Unlikes [24]	The Facebook Pages that this person has 'liked'.	Yes
C3	Photos Or Photos Uploaded [25]	Represents an individual photo on Facebook.	Yes
C4	Friends-list or Friends [26]**Erreur ! Signet non défini.**	A person's 'friend lists' - these are groupings of friends such as "Acquaintances" or "Close Friends", or any others that may have been created.	Yes
C5	Friends' activities upon user's OSN objects	Represents an action of a friend in one of a user's objects on Facebook.	Yes
C6	News [27]	The person's news feed.	No (this data is not simply collected due to Facebook API restrictions but is partly reconstructed from a set of other data). See table A1 for more details.
C7	User Profile and Interests[28]	A user represents a person on Facebook. The /{user-id} node returns a single user.	No (this data is not simply collected due to Facebook API restrictions but is partly reconstructed from a set of other data). See

[22] This table is based on annex D of D7.1. This table is useful as it shows the codes (#) assigned to the various data types. In table A.1 we refer to these codes from table A.3. Otherwise the table is largely superfluous because the exact data collected from the OSN are described in more detail in table A.1 above. The data listed in this table are the data which the consortium initially intended to collect. Collecting these data is not always possible due to restrictions in the Facebook API (e.g.,. Facebook does not allow access to C4, the complete friends-list of the user, but only shows those friends which also use the DataBait tool). The data which are actually collected in the DataBait tool are presented in table A.1.

[23] https://developers.facebook.com/docs/graph-api/reference/v2.0/user/feed/

[24] https://developers.facebook.com/docs/graph-api/reference/v2.0/user/likes

[25] https://developers.facebook.com/docs/graph-api/reference/v2.0/photo/

[26] https://developers.facebook.com/docs/graph-api/reference/v2.0/user/friendlists

[27] https://developers.facebook.com/docs/graph-api/reference/v2.0/user/home/

[28] https://developers.facebook.com/docs/graph-api/reference/v2.0/profile

			table A1 for more details.

Table A.4. PD from tables 6.1.1-3 (used for training USEMP Tool Algorithms)[29]

#	Name	Description	Type	How likely is it that this data will be used to infer data derivatives?
D1	Likes (metric: C2)	Facebook pages that the user has liked.	List of URLs (or Facebook page ids)	Certain/very likely
D2	Shared Pages (metric: C1)	Pages/Links that the user has shared.	List of URLs.	Certain/very likely
D3	Site Unique Visits (metric: A1)	URLs that the user visits in their browser.	List of URLs	Maybe/unlikely
D4	Trackers (metric: B2)	Tracker URLs/ids associated with the visited websites.	List of URLs/ids	Maybe/unlikely
D5	FB Images (metric: C3)	Images that the user has uploaded and where the user is tagged.	List of URLs (or byte arrays)	Certain/very likely
D6	User network (metric: C4)	List of friends + connections between them (probably useful for value estimation)	List of Facebook profile ids, list of connections between them	Certain/very likely
D7	User friends reactions (metric: C5)	The reactions of friends in a user's posts.	List of likes, shares, comments on user's posts, comments.	Certain/very likely

[29] This table is adopted from annex D of D7.1.

Table A.5. Derived data: inferred from a subset via [A] the OSN app and [B] the browser plugin[30]

#	'Privacy dimensions (i.e., categories into which the derived data are organised: see D6.1)	Derived attributes	Which data from Annex D, D7.1 are used to establish or infer this? (more than one answer is of course possible)	Which method is used if the data are inferred?	Which data are used to train (and/or test) the classifier (model) if data are inferred?	Legal qualification in terms of EU data protection (DP) law and EU anti-discrimination (AD) law
A	Demographics	1. Age	To be established later in the USEMP project	Personal attribute behavioural detection. Other methods to be established later in the USEMP project	(1) MyPersonality dataset (2) USEMP pre-pilot dataset)	DP: PD; PROFILE-OUTPUT AD: Protected ground in the field of employment
		2. Gender	Likes	Personal attribute behavioural detection. Other methods to be established later in the USEMP	(1) MyPersonality dataset (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT AD: Protected ground in the field of (1) employment, (2) access to goods and services

[30] This table is based on deliverable D6.1. It is not certain that all these data will be inferred.

				project	(3) ImageNet	
		3. Nationality	To be established later in the USEMP project	Multimodal concept detection. Personal attribute behavioural detection. Other methods to be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (2) ImageNet	DP: PD; PROFILE-OUTPUT AD: Protected ground but many exceptions (i.e. particular areas where differentiation based on nationality is allowed)
		4. Racial origin	To be established later in the USEMP project	To be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD, LSD; PROFILE-OUTPUT AD: Protected ground in the field of (1) employment, (2) access to goods and services, (3) education, (4) social advantages, (5) social protection
		5. Ethnicity	To be established later in the USEMP project	To be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD, LSD; PROFILE-OUTPUT AD: Protected ground in the field of (1) employment, (2) access to goods and services, (3) education, (4) social advantages, (5) social protection
		6. Literacy level	To be established later in the	To be established later in the USEMP	(1) USEMP pre-pilot/system operation dataset (data from survey,	DP: PD; PROFILE-OUTPUT

			USEMP project	project	OSN and browsing behaviour data)	
		7. Employment status	To be established later in the USEMP project	To be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT
		8. Income level	To be established later in the USEMP project	To be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT
		9. Family status .	Likes	To be established later in the USEMP project.	(1) MyPersonality dataset (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD, could be LSD if it reveals information about one's sex life (or according to the pGPDR: sexual orientation or gender identity); PROFILE-OUTPUT AD: if the data reveals sexual orientation- this is a protected ground in the field of employment law
B	Psychological Traits	1. Emotional stability	To be established later in the USEMP project	Personal attribute behavioural detection. Other methods to be established later in the USEMP	(1) MyPersonality dataset (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT; possibly LSD (if characterized as health data)

				project		
		2. Agreeableness	To be established later in the USEMP project	Personal attribute behavioural detection. Other methods to be established later in the USEMP project	(1) MyPersonality dataset (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT; possibly LSD (if characterized as health data)
		3. Extraversion	To be established later in the USEMP project	Personal attribute behavioural detection. Other methods to be established later in the USEMP project	(1) MyPersonality dataset (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT; possibly LSD (if characterized as health data)
		4. Conscientiousness	To be established later in the USEMP project	Personal attribute behavioural detection. Other methods to be established later in the USEMP project	(1) MyPersonality dataset (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT; possibly LSD (if characterized as health data)

		5. Openness	To be established later in the USEMP project	Personal attribute behavioural detection. Other methods to be established later in the USEMP project	(1) MyPersonality dataset (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT; possibly LSD (if characterized as health data)
C	Sexual Profile	1. Sexual preference	Likes	Personal attribute behavioural detection. Other methods to be established later in the USEMP project	(1) MyPersonality dataset (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT; LSD AD: if the data reveals sexual orientation- this is a protected ground in the field of employment law
D	Political Attitudes	1. Parties (Part of list for Belgium: CD&V; Groen!; N-VA; Open VLD /Part of list for Sweden: Centerpartiet; Vansterpartiet; Folkpartiet liberalerna)	To be established later in the USEMP project	Concept detection, opinion mining and textual similarity. Other methods to be established later in the USEMP project	(1) Wikipedia (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (3) SentiWordNet	DP: PD; PROFILE-OUTPUT; LSD

		2. Political ideology (Communist; Socialist; Green; Liberal; Christian democratic; Conservative; Right-wing extremist)	Likes	Concept detection, personal attribute behavioural detection, opinion mining and textual similarity. Other methods to be established later in the USEMP project	(1) MyPersonality dataset (2) Wikipedia (3) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (4) SentiWordNet	DP: PD; PROFILE-OUTPUT; LSD
E	Religious Beliefs	Supported Religion (Atheist, Agnostic, Christian, Muslim, Hinduist, Buddhist, Other, etc.)	Likes	Concept detection, personal attribute behavioural detection, opinion mining and textual similarity. Other methods to be established later in the USEMP project	(1) MyPersonality dataset (2) Wikipedia (3) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (4) SentiWordNet	DP: PD; PROFILE-OUTPUT; LSD AD: religious belief is a protected ground in the field of employment law
F	Health Factors & Condition	1. Smoking	To be established later in the USEMP project	Large scale visual concept recognition. Other methods to be established later in the USEMP	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (2) Yahoo Flickr Creative	DP: PD; PROFILE-OUTPUT; possibly LSD (if characterized as health data)

				project	Commons 100 Million (3) ImageNet	
		2. Drinking (alcohol)	To be established later in the USEMP project	Large scale visual concept recognition. Other methods to be established later in the USEMP project	(1) Yahoo Flickr Creative Commons 100 Million (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (3) ImageNet	DP: PD; PROFILE-OUTPUT; possibly LSD (if characterized as health data)
		3. Drug use	To be established later in the USEMP project	To be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT; possibly LSD (if characterized as health data)
		4. Chronic diseases	To be established later in the USEMP project	To be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT; possibly LSD (if characterized as health data)
		5. Disabilities	To be established later in the USEMP project	To be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT; possibly LSD (if characterized as health data) AD: Disability is a protected ground in the field of employment law

		6. Other health factors (e.g.: Exercise (yes / no); Late night shifts (yes / no); Staying up late)	To be established later in the USEMP project	To be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT; possibly LSD (if characterized as health data)
G	Location	1. Home	To be established later in the USEMP project	Location detection and concept detection. Other methods to be established later in the USEMP project	(1) Location estimation data set (2)Wikipedia (3) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (4) Yahoo Flickr Creative Commons 100 Million	DP: PD; PROFILE-OUTPUT
		2. Work	To be established later in the	Location detection and concept	(1) Location estimation data set	DP: PD; PROFILE-OUTPUT

			USEMP project	detection. Other methods to be established later in the USEMP project	(2)Wikipedia (3) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (4) Yahoo Flickr Creative Commons 100 Million	
		3. Favourite places	To be established later in the USEMP project	Location detection and concept detection. Other methods to be established later in the USEMP project	(1) Location estimation data set (2)Wikipedia (3) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (4) Yahoo Flickr Creative Commons 100 Million	DP: PD; PROFILE-OUTPUT
		4. Visited places	To be established later in the USEMP project	Location detection and concept detection. Other methods to be established later in the USEMP project	(1) Location estimation data set (2)Wikipedia (3) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (4) Yahoo Flickr Creative Commons 100 Million	DP: PD; PROFILE-OUTPUT

H	Consumer Profile	1. Brand attitude	To be established later in the USEMP project	Concept detection, opinion mining, logo detection, multimodal concept detection	(1) Wikipedia (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (3) SentiWordNet (4)FlickrLogos-32 (5) ImageNet	DP: PD; PROFILE-OUTPUT
		2. Hobbies	To be established later in the USEMP project	To be established later in the USEMP project.	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behavior data)	DP: PD; PROFILE-OUTPUT; possibly LSD if the hobby reveals one's racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, or information with regard to one's health or sex life.
		3. Devices	To be established later in the USEMP project	Concept detection and multimodal concept detection. Other methods to be established later in the USEMP project	(1) Wikipedia (2) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (3) ImageNet	DP: PD; PROFILE-OUTPUT
I	n.a.	Detection of faces in images (number	To be established	Large scale visual concept	(1) USEMP pre-pilot/system operation	DP: PD ; PROFILE-OUTPUT

		and location)	later in the USEMP project	recognition. Other methods to be established later in the USEMP project	dataset (data from survey, OSN and browsing behaviour data) (2)YFCC100M	
J	n.a.	Detection of opinion (positive/negative/ neutral) from textual posts and status updates	To be established later in the USEMP project	Opinion mining. Other methods to be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (2) SentiWordNet	DP: PD; PROFILE-OUTPUT
K	n.a.	Disclosure score (How sensitive, uncontrollable and visible are your data?)	This score is based on the pPrivacy derived data described above (A-H). It is second-order derived data (based on first-order derived data)	To be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data) (2) Relevance- and diversity-based reranking dataset	DP: PD; PROFILE-OUTPUT
L	n.a.	Personal data value score (what kind of audience do you have on your OSN and to whom could reaching such an audience be	To be established later in the USEMP project. This score is second-order	To be established later in the USEMP project	(1) USEMP pre-pilot/system operation dataset (data from survey, OSN and browsing behaviour data)	DP: PD; PROFILE-OUTPUT

		valuable?)	derived data (based on first-order derived data)			

Table A.6. Potential PDs in training and testing sets, used to train and test classifiers

Dataset	Source	Purpose : in which USEMP data driven module (see description in D2.3) is the dataset used and for which purpose?	Inferred attributes	Does the dataset contain personal data ?
MyPersonality	http://mypersonality.org/wiki/doku.php	***Used in***: the 'Personal attribute behavioral detection' and, quite probably, also in the 'Topic-based attribute detection' modules. ***Purpose***: Integration as	A1, A2, A9, B, C, D.2, E	Not likely. Anonymized. However, details need to be further explored[31]

[31] This research will be of a general nature since these datasets are massive collections of millions of data, which cannot all be individually investigated. The research engagement will thus most likely be limited to terms and conditions and sampling. On this basis some generic tendencies can be identified and some general questions can be raised with regard to the issues concerning external databases, such as (a) what type of data is made available, (b) what kind of access is provided (can parts of the database be downloaded, or must researchers operate on the database at the servers of the providers?), (c) under what conditions is access provided or downloading allowed (notably in terms of authentication, logging of operations, prohibition to share with others), (d) what information do providers of access to such databases give about the anonymisation of the data, (e) which contractual or other obligations are stipulated for researchers, (f) what if the data has been anonymised in a way that does not count as such under EU law?

		training set in the aforementioned modules.		
Zerr's image privacy dataset	http://l3s.de/picalert/#ustudydata	***Used in***: the 'Disclosure settings assistance framework'. ***Purpose***: Integration as training set in the aforementioned module. This module assists the user to define his privacy settings. It is used to assist classification of images as private or public. The user is warned when he / she is about to post an image that is classified as private.	None. As mentioned it is not used to infer any profile attributes	Needs to be further explored
Location estimation dataset	http://www.multimediaeval.org/mediaeval2014/placing2014/ Dataset accessible only by competition participants	***Used in***: the 'Location detection' module. ***Purpose***: Integration as training set in the aforementioned module.	G	Needs to be further explored
Kaggle community detection dataset	https://www.kaggle.com/c/learning-social-circles	***Used in***: the 'Disclosure settings assistance framework'. ***Purpose***: Integration as training set in the aforementioned	None.	No. Fully anonymized.

		module. It is used in order to help group the friends of a user in circles.		
Relevance- and Diversity-based Reranking dataset	http://www.multimediaeval.org/mediaeval2014/diverseimages2014/	***Used in***: the 'Face recognition', 'Logo recognition',' Multimodal concept detection', 'Large scale visual concept recognition', 'Disclosure scoring framework', and 'Disclosure settings assistance framework' modules. ***Purpose***: Benchmarking of method used for the relevance and reranking module that is used as part of the aforementioned modules.	None.	Needs to be further explored
Wikipedia	https://dumps.wikimedia.org/	***Used in***: the 'Text similarity' module. ***Purpose***: Creation of a training set that represents different privacy-related dimensions	D.1, D.2, E.1, G.1, G.2, G.3, G.4, H.1, H.3, H.4	Wikipedia does not contain any personal data.
SentiWordNet	http://sentiwordnet.isti.cnr.it/	***Used in***: the 'Opinion mining' module. ***Purpose***: Integration as training set in the aforementioned module	D.1, D.2, E.1, H.1	No personal data. This is just a list of keywords and scores about their sentiment.
ImageNet	http://image-	***Used in***: the 'Large scale	F.1, F.2, H.1, H.3	Needs to be further explored

	net.org/	visual concept recognition' module. ***Purpose***: Training set for aforementioned module		
FlickrLogos-32	http://www.multimedia-computing.de/flickrlogos	***Used in***: the 'Logo recognition' module. ***Purpose***: Training set for the aforementioned module	H.1	Highly unlikely to contain personal data.
Yahoo Flickr Creative Commons 100 Million	http://webscope.sandbox.yahoo.com/catalog.php?datatype=i&did=67	***Used in***: the 'Location detection and Face recognition' module. ***Purpose***: Training set for the aforementioned modules	A.2, A.3, F.1, F2, G.1, G2, G.3, G.4	Needs to be further explored
Pre-pilot / system operation dataset	-	***Used in***: all modules . ***Purpose***: Using questionnaire data, OSN data and browsing behavior data as a training set. To be obtained and investigated at a later stage.	Most likely all	Needs to be further explored
SNOW -twitter data set, derived through public API	http://ceur-ws.org/Vol-1150/overview.pdf;	***Used in***: the 'Network based attribute detection' module. ***Purpose***: Used by the network based attribute detection module (see: section 6 of D6.1.).	Needs to be further explored	Needs to be further explored

Annex B: IP-protected content with preliminary legal qualifications

Table B.1. Data potentially subject to IP-rights

Name of data	Description	Possible IPR rights
Facebook data potentially subject to IP rights (subset from the full list in table A.1):		
user_about_me	Access to a person's personal description (the 'About Me' section on their Profile) through the bio property on the User object.	Maybe, user copyright (on content stories)
User_posts	Access to a person's posts on the User object	Yes, user copyright
user_photos	Access to the photos a person has uploaded or been tagged in. This is available through the photos edge on the User object.	Yes, user copyright / 3rd party copyright
user_status	Access to a person's statuses. These are posts on Facebook which don't include links, videos or photos.	Yes, user copyright
user_videos	Access to the videos a person has uploaded or been tagged in.	Yes, user copyright / 3rd party copyright
Browser data potentially subject to IP rights (subset from the full list in table A.1):		
A4 - Images accessed through browser	Images Uploaded/Accessed/Downloaded by the user.	Yes, 3rd party copyright
A5 - Videos accessed through browser	Videos Uploaded/Accessed/Downloaded by the user.	Yes, 3rd party copyright
A7 - Text accessed through browser	Text Uploaded/Accessed at the web site.	Yes, 3rd party copyright
A8 – News accessed through browser	Page views of specific news elements.	Maybe, 3rd party copyright

Table B.2. Data from external data sets potentially subject to IP-rights

Dataset	Source	Purpose	IP-protected content in the dataset?	Is the dataset protected by sui generis data base rights or copyrights on the database?
MyPersonality	http://mypersonality.org/wiki/doku.php The list of data derived from the Facebook profiles can be found [here]().	Integration as training set in the behavioral detection module and quite probably also in the topic based attribute detection module.	Yes, copyright protection – the dataset contains data of more than 4 million individual Facebook profiles. (of which a significant part is related to privacy attributes). It should be checked which copyright protected material listed in table B.1 (posts, photos, statuses, and videos) is included but a preliminary inspection seems to indicate the dataset does contain some copyright protected material (posts and photos). Copyright is held by the authors of these images.	Maybe, 3rd party copyright or sui generis data base rights
PicAlert dataset (aka “Zerr’s image privacy dataset”)	[http://l3s.de/picalert/#ustudydata]()	Integration as training set in a module that assists the user to define his privacy settings. It is used to assist classification of images as private or public. The user is warned when he / she is about to post an image that is classified as private.	Yes , copyright protection on images – the dataset contains publicly available images uploaded to Flickr (which have been classified as either “private” or “public” and are annotated with a title, description and tags). Copyright is held by the authors of these images.	Maybe, 3rd party copyright or sui generis data base rights
Location	http://www.	Integration as	Yes, copyright protection on	Maybe, 3rd

estimation dataset	multimediaeval.org/mediaeval2014/placing2014/ Dataset Accessible only by competition participants	training set in the location recognition module.	images – 5 million geotagged photos and 25,000 geotagged videos that are used for training, and 500,000 photos and 10,000 videos that are used for testing. All photos and videos are taken from the YFCC100M dataset (see below) and are available under the Creative Commons license. Copyright is held by the authors of these images.	party copyright or sui generis data base rights
Kaggle community detection dataset	https://www.kaggle.com/c/learning-social-circles The list of variables derived from the Facebook profiles can be found here (accessible when signing-up for Kaggle).	Integration as training set in the privacy settings assistance module. It is used in order to help group the friends of a user in circles.	No – the dataset contains data of individual Facebook profiles and their relations with other user profiles. These data (birthday, classes attended, degree attended, school attended, year school was completed, family name, gender, location, political views, religious views, work position, employer, etc.) do not contain copyright protected content.	Maybe, 3rd party copyright or sui generis data base rights
Relevance- and Diversity-based Reranking dataset	http://www.multimediaeval.org/mediaeval2014/diverseimages2014/	Benchmarking of method used for the relevance and reranking module that is used as part of the 'Face recognition', 'Logo recognition',' Multimodal concept detection', 'Large scale visual concept recognition', 'Disclosure scoring framework', and	Yes, copyright protection on images – the dataset contains Flickr pictures of 300 'places of interest' (POIs) each of which is described through a list of 50 ranked images (which are relevant and diverse representations of the particular POI). Copyright is held by the authors of these images.	Maybe, 3rd party copyright or sui generis data base rights

		'Disclosure settings assistance framework' modules.		
Wikipedia	https://dumps.wikimedia.org/	Creation of a training set that represents different privacy-related dimensions	Yes, copyright protection of images and text on Wikipedia. The authors of Wikipedia entries give permission for the reproduction and modification of the text as long as it complies with Wikipedia's licensing terms.	Maybe, 3rd party copyright or sui generis data base rights
SentiWordNet	http://sentiwordnet.isti.cnr.it/	Integration as training set in the opinion mining module	No, probably no copyright protected content – the dataset is a lexical resource built to support opinion mining and sentiment analysis tasks. It assigns three sentiment scores (positivity, negativity, objectivity) to synsets (concepts).	Maybe, 3rd party copyright or sui generis data base rights
ImageNet	http://image-net.org/	Training set for the visual concept recognition module	Yes, copyright protection of images – this dataset is built by populating a significant part of another dataset, called WordNet (containing over 100,000 words, mostly nouns), with images illustrating the words contained in WordNet. As of March 2015, the dataset contains over 14 million images depicting nearly 22,000 synsets (concepts). ImageNet does not own the copyright of the images. ImageNet only provides thumbnails and URLs of images, in a way similar to what image search engines do. Copyright is held by the authors of these images.	Maybe, 3rd party copyright or sui generis data base rights
FlickrLogos-32 (aka 'Logo recognition datasets')	http://www.multimedia-computing.de/flickrlogos	Training set for the logo recognition module	Yes, copyright protection of images. The dataset contains 8,240 Flickr images illustrating 32 logos. Copyright is held by the author of this content.	Maybe, 3rd party copyright or sui generis data base rights

Yahoo Flickr Creative Commons 100 Million ('YFCC100M dataset')	http://webscope.sandbox.yahoo.com/catalog.php?datatype=i&did=67	Training set for the location recognition and face recognition modules	Yes, copyright protection of images. The dataset contains 100 million Flickr images and videos (published under a creative commons license), with associated metadata (such as: identifier, owner name, camera, title, tags, geographic coordinates), that were shared between 2004 and 2014. Copyright is held by the author of this content.	Maybe, 3rd party copyright or sui generis data base rights
Pre-pilot / system operation dataset	-	Questionnaire data, OSN data, browsing behavior data. To be obtained and investigated at a later stage.	Maybe copyright protection – depending on whether these will contain material listed in table B.1 (OSN posts, photos, statuses, and videos; and/or images, videos, text and news items accessed through browser). Copyright is held by the authors of this content. The OSN will hold a transferrable license on some of this content.	Maybe, 3rd party copyright or sui generis data base rights
SNOW - Twitter data set, derived through public API	http://ceur-ws.org/Vol-1150/overview.pdf;	Used in section 6 of D6.1.	Yes, probably copyright protection on textual posts– dataset consisting of Twitter posts, Twitter lists that these users belong to, the network of interactions around a user and known privacy attributes of the user's network in order to predict the value of the privacy attribute for the user. The posts are probably copyright protected. Copyright is held by the author of the tweet.	Maybe, 3rd party copyright or sui generis data base rights
Data driven Modules (i.e. *not* a dataset but the USEMP module and the machine learning algorithm contained in			n.a.	No 3rd party copyrights and patents on software. Some of the algorithms were adapted from previous

it).				work of USEMP partners and some are entirely developed in the context of USEMP.

Annex C: list of abbreviations

LSD	Legal sensitive data (as defined in Art. 8 of Data Protection Directive 95/46 EC): □ racial or ethnic origin, political opinions; religious or philosophical beliefs; trade-union membership; data concerning health or sex life, or those relating to offences, criminal convictions or security measures
OSN	Online Social Network
AD	EU anti-discrimination law
PD	Personal Data
DP	EU data protection law
IP	Intellectual property
PROFILE-OUTPUT	Data which result from profiling
PROFILE-INPUT	Data used as input for profiling
DLA	Data Licensing Agreement
PDPA	Personal Data processing Agreement (of which the DLA is a part)
pGPDR	The proposed General Data Protection Regulation. This new EU law with regard to data protection will replace DPD 95/46.
DPD 95/46	Data Protection Directive 95/46/EC. This is the current main EU law with regard to data protection.