D3.13

Coordination of Legal Aspects in USEMP – v3

v 0.3 / 2016-10-25

Coordination by Katja de Vries and Mireille Hildebrandt (iCIS-RU). Contributions by Noel Catterall (HWC), Theodoros Michalareas (Velti) and Giorgos Petkos (CERTH).

This document presents the results of the legal coordination and integration during M27-M36 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.

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 and Mireille Hildebrandt (iCIS)
Reviewers	Giorgos Petkos (CERTH), Andreas Drakos (Velti)
Version	0.3
Status	Final
Dissemination level	PU: Public
Due date	2016-09-30
Delivery date	2016-10-25

Version	Changes
0.1	Initial Release, Katja de Vries with input from CEA and CERTH (relating to section 6) 25 September 2016
0.2	Internal review 1 (Drakos, Velti)
0.3	Internal review 2 (Petkos, CERTH)

Table of Contents

1.Introduction

The hands-on integration of legal conditions into the USEMP architecture and the DataBait tool is reported is three deliverables: D3.4 (delivered in May 2015), D3.9 (delivered in January 2016) and D3.13 (this deliverable, delivered in October 2016). These three deliverables are reports 'that provide a description of the harmonised legal constraints applicable to USEMP data, algorithms and platform' (DoW, p. 55). This deliverable (D3.13) reports on the period January 2016-September 2016. There is some overlap with the content of earlier versions of this deliverable. To be precise: section 5 (software components) and section 7 (data used in the pre-pilot) are repetitions of what was discussed in D3.9. This is because some issues (e.g. checking the use of software or databases in the creation of DataBait does not infringe any copyrights) have kept returning over the whole period of the USEMP project. We have continuously tried to update and expand our analyses. However, in some cases (section 5 and 7) we have concluded that the information was still up to date.

The deliverable is rather concise despite the fact that the reported period has been of intensive interdisciplinary collaboration and informational exchange. However, many results from this collaboration and exchange are already incorporated in the other WP3 research deliverables (D3.10-12) and in the deliverables of other WPs.

2. Updated DLA and PDPA

In this section we present the adjusted, final[1] versions of the Personal Data Processing Agreement (PDPA) and the Data Licensing Agreement (DLA). The DLA binds the user and the USEMP consortium and the PDPA – which binds the USEMP consortium partners internally. The DLA is part of the PDPA: by signing the latter each consortium partner is bound "*to comply with and perform in accordance with the USEMP Data Licensing Agreement (DLA, as attached to this contract) when processing the personal data of DataBait users*" (clause A of the PDPA).

In comparison with the earlier versions of the PDPA and the DLA it (1) includes a separate clause with regard to copyright (clause C of the DLA), (2) is updated to the latest version of DataBait (e.g., all mention of 'monetization' and the 'FIRE infrastructure' are removed, (3) has a slightly improved wording. The latter two changes have been a very interdisciplinary and collaborative effort in which all partners – both technical and social – have actively contributed.

[1] This final project version of the DLA and PDPA has not been actually signed by the legal departments of all partners or by the users. Given the relatively minor changes and the fact that the project was near to its end when these final updates were finalized, it seemed better to keep these versions as a basis for any potential post-USEMP project version of DataBait. See more on that in the next section ("3. Post USEMP version of the DLA and PDPA"). While, theoretically, it would have been optimal to have a complete and final version of the DLA and PDPA ready before any user testing , in practice this is not feasible. The DataBait architecture was constructed and adjusted during the whole duration of the project; moreover the DLA and PDPA are based on legal research that has been ongoing for the whole duration of the USEMP project.

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 application and services, and being joint data controllers,

Hereby agree:

(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 data of DataBait users, who are defined as the USEMP end-users who have signed the DLA 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.

(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 a specified button that can be accessed through the web interface of the DataBait application, 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 changes. 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 that can be accessed through the web interface of the DataBait application.

(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.

(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 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 application 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, political opinions, religious beliefs, relationship status, living situation as well as potential lifestyle preferences, including those that may interest specific types of brands and enterprises. When developing the DataBait application, 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 application 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 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 application 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 additional Facebook data of DataBait users, stored at HWC, for the visualisation of user's demographics and other statistical information. 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.

(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				

USEMP Data License Agreement

The parties:

(1) [……You……], 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], provider of the USEMP platform and services, joint data controllers, from hereon called '**USEMP consortium partners**'

Hereby agree:

(A) After registration, You may use **DataBait-Facebook app**. The DataBait-Facebook app will be used by the USEMP consortium partners to collect data that You share on Facebook. After registering to the DataBait You can, if You choose to, also install the **DataBait web browser plug-in**. The DataBait web browser plug-in will be used by the USEMP consortium partners to collect Your browsing data and allows You to control the trackers on the pages You visit. Together the app and the plug-in form the '**DataBait application**'. The data collected by the DataBait application can be data You posted (volunteered data), or online behavioural data reflecting what You did on Facebook and – if You install the DataBait browser plug-in – what You did on the internet (observed data).

B) You license the use of Your volunteered, behavioural 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 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).

C) The data we gather through the DataBait-application may contain creations (such as images, photos, text, video) protected under copyright or neighbouring rights. These protected creations will be copied and stored on our servers and adapted for the purposes of our research (in particular for inferring additional information from Your data, for the scientific purpose described in clause B, by using automated data analytic tools). We will not commercialise Your creations or distribute these to third parties. We will not communicate Your creations to the public: Your creations will only be made available to You (when You access the Databait application via the web interface) and to our research teams (for the purposes of our research).

You agree with this use of Your creations covering the worldwide territory, for the duration of our research project. You will not receive any remuneration but You will have access to our research results.

D) This license agreement confirms Your explicit consent to store the DataBait application on Your devices.

(E) 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 through the DataBait web interface, thus providing You with an online presence awareness tool.

(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 DataBait application.

(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, relationship status, living situation and those concerning health or sex life.

(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.

(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 all personal data will be either deleted, anonymized or processed for related scientific research. In the latter case the relevant USEMP consortium partner will ask You for Your consent.

(J) The USEMP consortium partners will not provide Your personal data to any third party.

(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:

3.Post-USEMP version of the DLA and PDPA

In this section we look at the post-USEMP version of the DLA and PDPA. The plans for DataBait after the end of the USEMP project are described in D9.7. As described in that deliverable, the possibility for users to make use of DataBait after the end of the project depends on whether we can secure funding to keep DataBait running and/or convince other reliable parties (research institutes or NGOs) to become a long-term host of the DataBait web application. The purpose of DataBait in the post-project period might change slightly. For example, if a civil rights organization or NGO was to take over, the purpose would no longer be scientific, though it still would be non-profit. The latter is important to underline: we do not intend to exploit DataBait commercially. We explored the possibility of (semi-) commercial, yet user-empowering exploitation of DataBait in D3.5 and D3.10 (the idea of "granular licensing"). As explained in D9.7, there are ample opportunities to commercially exploit some of the constitutive data technologies ('modules') of DataBait developed in USEMP and it is very well possible that some of these can contribute to personal data management (PDM) solutions which enhance profile transparency and user control over one's digital trail; however, if we look at the DataBait tool we have not found any commercially viable business model that would:

(a) not be contrary to one of the main premises of the USEMP project, that is, to provide user empowerment with regard to the potential commercial exploitation of their data based on machine learning analyses,

(b) would have a useful role given the EU legal framework for data protection and the current commercial ecology. Let us clarify why this is the case. One could imagine that certain users would like to be able to granulary license their data ("*you can use my health data for cancer research, but not for insurance purposes; nobody is allowed to use data about religion, etc.*") in exchange for profile transparency mediated through a third party transparency tool provider ("*you will keep my data safe in a personal data management-'safe' and I'll give instructions for sharing and non-sharing; for example, you can share my health data for cancer research, but then I should get to know exactly what happens to these data, etc*."). However, we show in D3.10 that such a solution is highly problematic in terms of the purpose specification and limitation principle (which prohibits any blank checks – even limited blank checks! - with regard to how data may be used). Moreover, in terms of a viable role within the business niche this would also be very problematic as it would put the independent position of DataBait at risk. the empowering strength of DataBait currently lies exactly in its position as an independent provider of transparency, not as a mediator between users and businesses/organisations looking for user data. It is also hard to imagine what the incentive for organisations and users would be to use a mediating

transparency provider and why this provider would want to take on the large responsibility for preserving the data secure and provide transparency). Nevertheless the idea of granular licensing should not be discarded completely. The business model might work in a specific context (for example in a medical environment, where the patient is given the power to set limits on to which medical professional has access to which data) but not in the current way DataBait is created (aiming to empower users who make use of the services of large OSNs or browsers).

Based on the above considerations we have come to the conclusion that it is important to ensure that DataBait is not exploited in a commercial way, or perverted to uses that are in conflict with its purposes and premises (e.g. if a business would like to use the DataBait technology to mine sensitive data and, instead of using them to inform and empower the user, would sell these data commercially). Obviously, it would make it even worse if this business was to 'lure' users into using their service using the name 'DataBait'. To prevent such scenarios from coming about and to preserve the integrity of DataBait we:

(1) filed a trademark application for the name "DataBait" with the European trademark office (as to prevent others from using DataBait for purposes that the consortium partners do not agree upon) – see section 4 for more information about the trademark application;

(2) are in the process of establishing an international non-profit organization (based in Brussels). All[2] the current consortium members will be members of this non-profit organization. The starting point for drafting the statutes of this non-profit organization is the current consortium agreement, but will need to be adjusted with regard to certain points such as admissibility conditions for new partners that would like to join the non-profit organization "DataBait". Other partners will be allowed to join "DataBait" at a later stage under the condition that all current partners agree and that their organization is in agreement with the statutes.

Any processing of user data by a post-USEMP version of DataBait will also be based on a PDPA and DLA. Given the fact that we intent to keep this post-USEMP version of DataBait very close to DataBait in its current version (same functionalities, same purpose of user empowerment, possibly also the same scientific purpose) the expected adjustments to the DLA and PDPA are not very large. We have gone through the current DLA and PDPA and anticipated the points where change will be needed. However, the exact adjustments will

[2] Or: almost all consortium members. This has been extensively discussed between the consortium and it will depend on the funding that will be collected. Not all consortium members have the resources to contribute of their own for the establishment of DataBait as a legal entity (an international non-profit organisation), a cost that was also not perceived in the USEMP Description of Work, even if they are interested in supporting the effort.

depend on the setting in which DataBait would be kept online and accessible for users. As described in D9.7 we have several routes to explore that might generate funding for DataBait (for example, the Open Society Foundation who at the point this document was prepared, have expressed interest in funding DataBait), but until we have something more definite we cannot give an adjusted post-project version of the DLA.

However, some pointers for this adjusted version are:

- References to the USEMP project in the DLA and PDPA will be replaced by references to the international non-profit organisation “DataBait VzW” and its statutes
- Clause (D) of the PDPA needs to be updated to the actual situation. Can all partners in "DataBait" still perform the necessary personal information assurance risk assessment ? How frequent ?
- Clause (E) of the PDPA (specifying the processing of personal data performed by each party) needs to be updated to the actual situation. The adjustments will probably include the removal of several processing actions (as compared to the current PDPA) because we are no longer a fully-fledged research project – the DataBait tool has already been created and is now in the stage that it only might need some further adjustments to improve its functioning.
- Clause (I) of the DLA (specifying the purpose of the data processing and the moment when the data will be deleted) will need to be adjusted to the new situation.

4. Trademark application for the verbal mark "DataBait"

As noted in the previous section, we have, in order to preserve the integrity of DataBait, filed a trademark application for the name "DataBait" with the European trademark office (https://euipo.europa.eu/ohimportal/en/apply-now) as to prevent others from using DataBait for purposes that the consortium partners do not agree upon.
We applied for the verbal mark "DataBait" for two classes and specified the applicable subcategories from those two classes:

- **Class 9 (IT GOODS)**. *Subcategories:* Privacy software, Privacy protection software, Computer software, Computer programs, Computer software programs, Image recognition software, Educational software, Science software
- **Class 42 (IT SERVICES)**. *Subcategories:* Data mining, IT services

Initially, the intention was to apply for the trademark on behalf of all partners so that all partners would co-own the trademark automatically. However, as this turned out to be legally and administratively rather complicated[3], theConsortium decised that Radboud University applies for the trademark and sign a contract with the other partners to keep the trademark on behalf of and for the account of all the consortium partners. This contract will also entail agreements about additional costs and decision-making about the trademark (e.g. regarding opposition if the brand is employed or registered by others), as well as an agreement on how to handle these things after termination of the consortium agreement.
Below (next page) is a draft agreement which we are currently adjusting to a real agreement. Key points from the agreement are:
(a) that Radboud University keeps the trademark on behalf of all the consortium partners,
(b) that all decisions with regard to the trademark will be taken by a majority of the partners,
(c) that the trademark will not be commercialized unless partners unanimously agree,
(d) that no decisions or actions will be taken that contravene the consortium agreement insofar as relevant.

[3] It required 6 proofs of power of attorney for Radboud University and 6 official signatures, and could have resulted in potential complications due to the fact that partners are in different member states.

DataBait Trademark Agreement[4]

Between:

1. **Stichting Katholieke Universiteit** (with its legal address at Geert Grooteplein-Noord 9, 6525 EZ Nijmegen) more particularly **Radboud University Nijmegen** with its offices at Comeniuslaan 4, 6525 HP Nijmegen part of which is **the Faculty of Science**, established at Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands (postal address: Postbus 9010, 6500 GL Nijmegen, the Netherlands) represented by its legal representative Prof. dr. L.M.C. Buydens, Dean of the Faculty of Science hereinafter referred to as "Radboud"

2. ___
3. ___
4. ___
5. ___
6. ___
7. ___

hereinafter, jointly or individually, referred to as "Consortium Partners" or "Consortium Partner" or "Consortium Partners or Consortium Partner"

WHEREAS:
The Consortium Partners have participated in the FP7 Project _____ funded by_____ as a consortium. And have signed a Consortium Agreement on _____.
The Consortium Partners wish to register for a EU Trademark called DataBait (hereinafter "Trademark").
The Consortium Partners wish to describe the issues concerning the Trademark in this agreement.

NOW, THEREFORE, IT IS HEREBY AGREED AS FOLLOWS:

1. Radboud has proceeded towards the trademark application and applies for the trademark on its own USEMP budget on behalf of all the Consortium Partners.
2. Radboud keeps the trademark on behalf of all the Consortium Partners.
3. All decisions with regard to the Trademark will be taken by a simple majority of the Consortium Partners.
4. The Trademark will not be commercialized unless all Consortium Partners unanimously agree.
5. No decisions or actions will be taken that contravene the Consortium Agreement insofar as relevant.
6. Additional costs regarding to the Trademark (e.g. regarding opposition if the brand is employed or registered by others) will be decided by simple majority.
7. The termination of the Consortium Agreement will not affect the obligations of this Trademark Agreement.
8. This Trademark Agreement will be terminated when the trademark expires or if the Consortium Partners otherwise decide.

AS WITNESS:

_____

[4] This agreement is a draft – it is not the final version

The Consortium Partners have caused this Trademark Agreement to be duly signed by the undersigned authorised representatives in separate signature pages.

[INSERT NAME OF PARTY]
Signature :
Name :
Title :
Date :

5.List of software components

In this section we look at the software components used in the creation of the DataBait architecture. We look at the specific licensing conditions, to ensure that no copyright infringement is made.

As far as the open licences are concerned, the USEMP consortium is aware that the use of the same creations in the context of a follow up project should be examined again. It is indeed not self-evident that possible requirements of a non-commercial purpose or research purpose are still met. Ideally, it should be verified for each licence mentioned in (i) whether the actual use of DataBait is covered (non-commercial, scientific purpose); (ii) whether potential future use is covered (non-commercial, but no longer done by scientific entities); and (iii) whether the licence comes with any obligations, such as “share alike” licence of creative commons (obligation to share the work under the same terms) and the attribution (does this obligation only apply when the used work is distributed or also when the cc-licensed works are used to extract meta-data but the works themselves are not re-distributed).

With regard to the publication of Databait code under open source licence after the end of the USEMP project[5], it is to be verified before making any code accessible whether the licences on the used third party code allows publishing under open source.

[5] A large part of the modules developed within WP6 are already available in open source. The package can be found here: https://github.com/MKLab-ITI/usemp-pscore. No third party code which does not allow open source publishing was used for the creation of these modules.

What software components (and under which license) were used in the creation of the DataBait data modules ('algorithms')?

Databait tools WP7 group of functionality	Software components	Distribution license	Reference
LimeSurvey Server		GPLv2	
Backend API Server list of 3rd party components	Facebook Capture engine	MIT Licence	
	Public internet facing proxy	BSD 2-Clause Licence	http://nginx.org/
	Caffe image analytics library	BSD 2-Clause Licence (+ CEA Components)	http://caffe.berkeleyvision.org/
	Java-ML	GPLv2	http://java-ml.sourceforge.net/
	Apache commons-configuration library	ASLv2	https://commons.apache.org/proper/commons-configuration/
	Apache http components library	ASLv2	https://hc.apache.org/
	Apache commons dbcp library	ASLv2	https://commons.apache.org/proper/commons-configuration/
	spring-framework	ASLv2	https://spring.io/
	hibernate-core	LGPL 2.1	http://hibernate.org/
	Jackson	ASLv2	https://github.com/FasterXML/jackson
	Aspectj	Eclipse Public License - v 1.0	https://eclipse.org/aspectj/
	mysql-connector-java	GPLv2	http://dev.mysql.com/downloads/connector/j/
	Restfb	MIT Licence	http://restfb.com/
	slf4j	MIT Licence	http://www.slf4j.org/
	Hsqldb	BSD	http://hsqldb.org/
	Kryo	BSD	https://code.google.com/p/kryo/
	MySQL Server	GPL license	https://dev.mysql.com/
	Tomcat Application Server	ASLv2	http://tomcat.apache.org/
Front-end server components	Django application server	BSD license	https://www.djangoproject.com/
	python-social-auth	BSD license	http://psa.matiasaguirre.net/
	Reportlab	BSD license	http://www.reportlab.com/

	Webgl Globe visualization	ASLv2	https://github.com/dataarts/webgl-globe/blob/master/LICENSE
	D3.js	BSD license	http://d3js.org/
	SimpleWeatherJS	MIT Licence	http://simpleweatherjs.com/
	Bootstrap	MIT Licence	http://getbootstrap.com/
Machine Learning/Graph DBs	Hadoop	ASLv2	https://hadoop.apache.org/
	Mahout	ASLv2	http://mahout.apache.org/
	Caley	ASLv2	https://github.com/google/cayley
	Helios.JS	GPL	https://github.com/entrendipity/helios.js
Browser plugin	Disconnect.ME	GPL	https://github.com/disconnectme/disconnect
Privacy scoring framework	pymongo	Apache License 2.0	http://api.mongodb.org/python/current/
	Flask	BSD license	http://flask.pocoo.org/

6. List of datasets used

In this section we look at the datasets used to train the DataBait algorithms. We look at the specific licensing conditions, to ensure that no copyright infringement is made.

In the column under the heading "*Extra info about use of the dataset*", we looked at questions such as:

(1) Is the entire dataset copied (downloaded)?

(2) For which purpose is it used (extraction of meta-data, find correlations between data, merging of several datasets owned)?

(3) By USEMP partner or others?

(4) How long is the copy kept?

(5) Is it deleted or is it saved and re-used?

(6) Are the data fed into the USEMP system to check if the USEMP system functions correctly? Are only the meta-data drawn from the datasets re-used?

In the column under the heading "*Extra info about the license*" , we looked at questions such as:

(1) Is the actual use of DataBait (non-commercial, scientific purpose) covered?

(2) Are potential future uses (non-commercial, no longer done by scientific entities) covered?

(3) Does the licence come with any obligations, such as the "share alike" licence of creative commons (obligation to share the work under the same terms) and the attribution (does this obligation only apply when the used work is distributed or also when the cc-licensed works are used to extract meta-data but the works themselves are not re-distributed)?

Module	Dataset	License	Extra info about use of the dataset	Extra info about the license
Logo recognition	Logo recognition dataset	Flickr terms of use - CC licenses	The dataset consists of CC licensed product images that were gathered from Flickr in order to improve logo detection. The images are used to build logo recognition models (D5.5) and are not directly exploited in DataBait. The original data will be retained for research purposes after the end of USEMP.	Images from this dataset may have one of several CC licenses used by Flickr. All the licenses permit redistributio n, some allowing commercial use, while other imposing redistributio n for noncommer cial purposes. The original data are not used within the system and only the learned models are used.
Multimodal concept detection	ImageNet	Non-commercial research	Data from ImageNet are used to train visual concept models, which are then exploited in D5.3 to enrich image textual descriptions. A local copy of ImageNet is kept for ongoing research purposes. Only the learned models and not the original images are used in the system.	ImageNet provides thumbnails and URLs of images, in a way similar to image search engines. Access to the original images is

				granted to researchers who wish to use the images for non-commercial research and/or educational purposes.
Text similarity	Wikipedia	Wikipedia licensing terms – CC license	Wikipedia articles are used to build a structured semantic space. The method described in D5.1 relies on co-occurrence statistics between the concepts, identified by Wikipedia page titles, and the actual content of the article. Wikipedia data are not directly used in DataBait. Only the resulting word representations are exploited for multilingual text similarity.	The CC license permits unconstrained data redistribution and use. The original data are not used in DatBait, only the word representations obtained from them.
Concept detection	Wikipedia	Wikipedia licensing terms – CC license	The same word vectors obtained from Wikipedia pages that were used for text similarity are also exploited for multimodal concept detection in D5.3.	The CC license permits unconstrained data redistribution and use. The original data are not used in DatBait, only the word representations obtained from them.
Opinion mining	Movie reviews dataset	Non-commercial	Movie reviews are used to learn opinion mining models (D5.4). The original data are not used in	Each source carries specific

			DataBait and most of them are not kept after the models are obtained. Review data may come from one of the following sources: MovieLens[6] collection, AlloCiné[7], MovieMeter[8] and Nyheter24[9].	permissions. *MovieLens* permits use and redistributio n for research purposes, *AlloCiné* prohibits commercial use, while *MovieLens* and *Nyheter24* prohibit redistributio n. The original data are not used in our system, only the learned models.
Location detection	Location estimation dataset YFCC100M	CC license CC license	The first of these dataset is a particular subset of the second. CERTH downloaded the text metadata, while CEA downloaded both the text metadata and images. These datasets were used to learn general text and visual location detection models. Only the learned models and not the original data are used in the system. The original data will be retained for research purposes beyond the end and scope of USEMP.	The CC license permits use both for research purposes during the project and for commercial ones, provided that appropriate attribution is made. Note that the

[6] http://grouplens.org/datasets/movielens/
[7] http://www.allocine.fr/
[8] http://www.moviemeter.nl/
[9] http://nyheter24.se/sok/filmset

				original data are not used within the system; only, the derived location detection models are used.
Large scale visual concept recognition	Relevance- and diversity-based reranking dataset	CC license	The dataset has been downloaded by CERTH and was used to evaluate a relevance and diversity-based reranking method (D5.5, Sec. 4). It was downloaded by the link provided by the organizers of the MediaEval 2014 RDSI task[10]. The copy is kept and will be retained for future research. The data are not fed into DataBait.	All the photos are under CC licenses that allow redistributio n. There is no plan to use it for non-research purposes.
	ImageNet	Non-commercial research	Data from ImageNet are used to train visual concept models, which are then exploited in D5.2 to produce semantic image representations. Only the learned models and resulting image descriptors are used within DataBait and not the original images.	ImageNet provides thumbnails and URLs of images, in a way similar to image search engines. Access to the original images is granted to researchers who wish to use the images for non-commercial research and/or educational

[10] http://www.multimediaeval.org/mediaeval2014/diverseimages2014/

				purposes.
	YFCC100M	CC license	YFCC is used in the context of large scale visual concept learning only for internal evaluation purposes described.	All the images and videos in this collection are licensed under one of the Creative Commons copyright licenses (not necessarily the same) and can be used for benchmarking purposes.
Collection – based classifiers	My Personality	MyPersonality Terms of Use[11]	CERTH has registered as a collaborator and downloaded several of the databases available from the MyPersonality website[12]. The respective data has been used to conduct academic research that has been reported in D6.1, D6.4 and D6.5. No use of the data has been done within DataBait, nor will it be done in the future. The databases will be retained for future research.	The MyPersonality Terms of Use do not allow use for commercial purposes or in collaboration with commercial entities.
Network-based attribute detection	SNOW dataset	CC-BY	CERTH is the creator and owner of this dataset and has made it available under a CC-BY license on figshare.com. The dataset was used for evaluating the network-based attributed detection module (D6.1), which was ultimately not used within the system (due to restrictions in	The CC-BY license permits any use provided that attribution is made.

[11] http://mypersonality.org/wiki/doku.php?id=database_use_guidelines
[12] http://mypersonality.org/wiki/doku.php?id=start

			the use of the Facebook API for forming user networks).	
Disclosure settings assistance framework	PicAlert	No license information available	The dataset was obtained from a dedicated page where it is made available http://l3s.de/picalert/ but offering no information regarding its license. The dataset was used to test the performance of the private image classification module (D5.5), but is not used for the model that is deployed within DataBait. The dataset will be retained for conducting further experiments.	There is no information as to whether the dataset can be used for other purposes, beyond research (it is already used for research by numerous researchers) . There is no plan to use it for non-research purposes.
	YourAlert– this is an **internal** dataset	Collected and controlled by CERTH	Users voluntarily contributed about 40 images each along with private/public annotations (D5.5 Section 3.2.1). We used the images to extract visual features which along with annotations were used to build a private image classification model that is deployed within DataBait. Original images were processed by computer software (not viewed by any person) and deleted immediately after processing. The extracted features and the annotations have been made publicly available: https://github.com/MKLab-ITI/image-privacy The derivative metadata will be retained for conducting further experiments.	Since the controller of the dataset is CERTH, its use for both academic research and within DataBait is permitted.

	Kaggle social circle dataset	No license information available	The datasets have been downloaded by CERTH from their original location: http://snap.stanford.edu/data/index.html which does not mention any license information but only requests that research done using this data should cite the SNAP dataset webpage. The dataset was downloaded from CERTH and used to test the settings assistance approaches based on clustering ego-networks (D6.2 and D6.4). The dataset was only used for offline evaluation purposes and will not be used within DataBait. The dataset was deleted after its use.	There is no information as to whether the dataset can be used for other purposes, beyond research (it is already used for research by numerous researchers) . There is no plan to use it for non-research purposes.
URLs mapper	DMoz[13]	Creative Commons Attribution 3.0 Unported License.	This dataset has been used in order to develop the URLs mapper that associates URLs to specific privacy attributes. For more details please see D6.5	There is no plan to use this dataset for non-research purposes.

[13] http://www.dmoz.org/

7.WP7 pre-pre pilot and pre-pilot experimental datasets

This section describes the data that were shared between technical partners from the pre-pilot for model training and evaluation purposes:
Data should include some or all pre-pre pilot and pre-pilot users' Facebook collected data, where all personal identifiers have been anonymized. In particular, the facebook ids, usernames, phone numbers and email addresses of users have been anonymized by hashing.

Anonymized data has been stored at the HWC servers as an encrypted archive which can be downloaded by the partners listed below.

CERTH:
Data requested:
- User likes for all the users.
- Posts / status updates.
- Extracted visual concepts and logos.
- Survey responses.

Data has been used for:

- The development and validation of an inference module (the collection-based classifiers) that predicts personal attributes of users based on the collection of users' likes, posted messages and detected visual concepts (for more details on this modules please see D6.4 and D6.5).

VELTI:
Data requested:
- User likes for all the users.
- Survey responses

Data have been used for:

The development and validation of the DataBait inference (as a whole) as well as the audience influence module

iMinds:

Data requested:

- Facebook Data.
- Survey responses

Data have been used for:

- To investigate if there are contradictions between what people have claimed that is available online (survey) and what actually could be found.

Information Datasets

"id" - discarded and converted to a non-tracable guid
"metadata" - discarded
"type" - discarded
"name" - discarded
"firstName" - discarded
"middleName" - discarded
"lastName" - discarded
"link" - discarded
"bio": - discarded
"quotes": - discarded
"about": - discarded
"relationshipStatus": - unaltered
"religion": - unaltered
"website": - discarded
"birthday": - unaltered
"email": - discarded
"timezone": - unaltered
"verified": - unaltered
"gender": - unaltered
"political": - unaltered
"locale": - unaltered
"username": - discarded
"picture": - discarded

"hometown": - discarded
"location": - discarded
"significantOther": - discarded
"updatedTime": - unaltered
"thirdPartyId": - discarded
"currency": - unaltered
"tokenForBusiness": - discarded
"interestedIn":[ ], - unaltered
"meetingFor":[ ], - unaltered
"work":[ ], - discarded
"education":[ ], - unaltered
"sports":[ ], - unaltered
"favoriteTeams":[ ], - unaltered
"favoriteAthletes":[ ], - unaltered
"languages":[ ], - unaltered
"birthdayAsDate": - unaltered
"hometownName": - discarded
"likes":[ ], - unaltered
"surveyAnswers":[ ], - unaltered (identifying information treated as above i.e. email, address etc.) See D3.4 for data contained within the survey.
"images":[ ], - not in dataset
“posts” / “statuses” - unaltered

- USEMP survey data – anonymized from prepre-pilot and prepilot

Data have been used for: To improve the WP6 classifiers

8. Additions to the DataBait 'What, how, why?'-tab.

To give the user more context to the results generated by DataBait :

1. WP3 and WP6 have co-produced a tutorial on how to adjust privacy settings on Facebook. The part of the tutorial created in WP6 (which can be found in Annex 3 of D6.5) focuses in particular on controling 'social privacy' settings. The part of the tutorial created by us in WP3 (see D3.10) looks at institutional privacy : how does Facebook inform users about how they are profiled, why they see particular advertisements, how they can adjust this, etc. Moreover, we compare the information provided by Facebook (as a fulfilment of their obligation to provide profile transparency) with the independent, speculative information provided by a third-party transparency tool such as DataBait. We conclude that both sources of information complement each other nicely.

2. We suggested including references to other independent transparency tools :

- Transparency tools which function through input-output matching ('blackbox testing'): Xray (http://xray.cs.columbia.edu/)
    - AdFisher (https://www.cs.cmu.edu/~mtschant/ife/)
    - SunLight (https://columbia.github.io/sunlight/ or www.cs.columbia.edu/~djhsu/papers/sunlight.pdf)

  DataBait as a transparency tool is based on merely 'input' (in contrast to the input-output matching of other tools). This makes DataBait more 'speculative' than the aforementioned tools. See the 'disclaimer tab' for a closer discussion of the benefits and drawbacks of this.

- Services that allow you to block trackers:
    - Ghostery (https://www.ghostery.com/)

3. We suggested including references to some legal sites that help to get a grip on the legal EU data protection framework and the possibilities to take legal steps towards OSNs and browsers :

Commission site about data protection:

http://ec.europa.eu/justice/data-protection/

The new General Data Protection Regulation

http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC

Standard forms to request access to your Facebook data:

http://www.europe-v-facebook.org/EN/Get_your_Data_/get_your_data_.html

European Digital Rights: good site to get informed about the civil society work done in defense of rights and freedoms in the digital environment in Europe:

https://edri.org/

4. We suggested including a 'disclaimer'-tab about the 'speculative' nature of DataBait. See D3.10 for more details.

9.Deletion of all data and open source

As stated in the DLA all DataBait data will be deleted within three months of the end of the project (before 31 December 2016). All pre-prepilot data kept by other partners than HWC will also be deleted.

The question of data anonymization does not have to be explored because no data are kept. The only thing that is preserved are the DataBait predictive models (or 'algorithms') trained on user data. Given the state-of-the-art, nothing in these models can be tied to any identifiable persons; re-identification is not possible. These models will, as much as possible[14], be made available online in an open source format (the exact licensing conditions of which, in some cases[15], still need to be established). The DataBait website containing documentation and links to the DataBait open source software (aimed at the research community, other developers of transparency tools, policy makers and anyone interested in the DataBait project) will be put online shortly after the ending of the USEMP project. In contrast to the DataBait web application (that is, the user version of DataBait), this website does not depend on finding additional funding. The DataBait web application will not be available for users until further funding has been secured and a new legal form for DataBait is in place. See D9.7 (final exploitation report) for more details.

Any processing of new DataBait data will happen based on an adjusted DLA and PDPA. For the sake of preserving a certain sense of continuity all current users of DataBait are asked to leave their email addresses if they are interested in being contacted again and using DataBait is it comes online again in a post-project version. In this way, we will not lose contact with the user group. From a user perspective, interruption of the DataBait service might be discouraging, but from a pragmatic and legal perspective we do not see any other way to do this.

When new funding has been found to put the user version of DataBait online again, it is possible that the DataBait server will no longer be located at HWC (in the UK). The location of the post-USEMP project DataBait server will depend on (1) the exact format in which DataBait is resuscitated (e.g. if it is hosted by a research institute or NGO that takes a leading role), and (2) the implications of the Brexit for data protection in the UK. All the above are items not foreseen in the initial Description of Work of the USEMP project and are items that affect DataBait but over which the USEMP consortium has no (or limited) control. As the USEMP consortium, we have identified the issues and we have planned a number of actions to allow the new legal entity, once created, to be able to overcome them.

[14] For the visual mining models/algorithms created by CEA applications for patents are made. The partners at CEA will check with their research institution whether some parts of their work can nevertheless be disclosed in some way or that some exceptions can be made, but this will probably be a long procedure.

[15] See footnote 5. A large part of the modules developed within WP6 are already available in open source. The package can be found here: https://github.com/MKLab-ITI/usemp-pscore. No third party code which does not allow open source publishing was used for the creation of these modules.

10. Concluding remarks

This report has presented the results of the legal coordination and the integration during the the period January 2016-September 2016. It presents:

- The adjusted DLA and PDPA (including the adjustment of the DLA with regard to copyright);
- Reflection on the post-USEMP version of the DLA and PDPA;
- The steps taken to file a trademark application for the verbal mark “DataBait”;
- A list of software components used in the “DataBait” tool and the respective licenses on which their use is based;
- A list of data sets used in the creation of DataBait’s predictive models and algorithms. The list specifies the exact ways in which these data were used and what uses are allowed for by the respective licenses;
- A list of pre-pre-pilot (and pilot) data which were made available for access outside HWC to a subset of the USEMP partners
- A list of additions to the informative tab in DataBait
- Which steps are taken at the end of the USEMP project with regard to deleting all data
- Which steps are taken to disseminate the USEMP research and make the DataBait web application available online again after the USEMP project has ended.

In regards to the data processed, the list previously presented in D3.4 is still up to date, pending the approval from Facebook (at the moment of writing DataBait is under the Facebook review process). The same applies if DataBait is to be integrated with other OSNs (e.g. Twitter, Instagram), which during the project have been analyzed but from the technical perspective not integrated due to numerous technical issues. Possible conflicts between OSN IP rights and DataBait’s transparency tool are discussed in D3.11 while the possibilities for having a granular permission system are discussed in D3.10. The further exploration of how our research into the empowering possibilities of image rights and personality rights can impact on the DataBait architecture is sufficiently discussed in D3.12. Because we concluded that these empowering possibilities are currently too complex to integrate in the architecture of DataBait we do not discuss this in this deliverable. However, the results of our research in this respect will be published on the informative DataBait website which will be the main point of information for DataBait and USEMP after the end of the project (D9.6 and D9.7).

.