D3.9

Coordination of Legal Aspects in USEMP – v2

v 2.0 / 2016-01-01

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

This document presents the results of the legal coordination and integration during M19-26 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), Giorgos Petkos (CERTH), Symeon Papadopoulos (Velti)
Reviewers	Giorgos Petkos (CERTH) Adrian Popescu (CEA)
Version	1.0
Status	Final
Dissemination level	PU: Public
Due date	2016-01-31
Delivery date	2016-01-27

Version	Changes
0.1	Initial Release, Katja de Vries with input from Noel Catterall (HWC), Giorgos Petkos (CERTH), and Symeon Papadopoulos (Velti) 20 December 2015
0.2	Revisions Mireille Hildebrandt (iCIS) 11 December 2015
0.3	Review Adrian Popescu (CEA) 12 December 2015
0.4	Review Giorgios Petkos (CERTH) 20 January 2016

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 (this deliverable: delivered in January 2016) and D3.13 (expected delivery 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). The first deliverable in this series (D3.4) reported on the period October 2013-April 2015, while this deliverable (D3.9) reports on a much shorter period (May 2015-December 2015). This explains why this deliverable is rather concise despite the fact that the reported period has been of intensive interdisciplinary collaboration and informational exchange.

Some of the results have been reported in other WPs, as described in section 4 of this deliverable. Moreover, many of the results of the legal coordination task in the reported period have already been described in other deliverables (for example, the "How, what, why"-section in DataBait, the explanatory animation about DataBait and the ongoing work with regard to anonymization/deletion of the DataBait data is described in D3.6). The input with regard to the functioning of the DataBait architecture and the implications in terms of IP rights is integrated in D3.7 and D3.8. The "disclaimer tab" about the "speculative" nature of the data derivatives in DataBait is reported in D3.7. We will not replicate all of this in this report but we refer the reader to the respective deliverables (D3.6-D3.8) and the relevant deliverables in the other WPs.

Some parts of the legal coordination work (e.g. adjustment of the DLA with regard to copyright, an additional tab in DataBait with information about how the user can use DataBait information to effectuate legal rights, and further reporting on the licenses on the databases used for training and testing of the DataBait data driven modules) are still work-in-progress and will be addressed in the next version of this deliverable (D3.13). Other parts will only be addressed in a later stage, when certain questions with regard to the architecture have been resolved. For example, we intended to report in this deliverable which data will be processed with regard to Twitter and/or Instagram (the second OSN next to Facebook to which DataBait might be applicable) but currently it is still unclear which functionalities DataBait would offer with regard to this second OSN, so our legal inventory of the processed data and recommendations with regard to the architecture will have to wait until we know more about what is technologically feasible. Also the update with regard to the Facebook data which are processed and some issues with regard to copyright and database rights can only be addressed after the result of the review by Facebook of the DataBait Facebook app is known.

As described in the DOW, the legal coordination and integration task (T3.6) 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 deliverable presents a 'snapshot' of this ongoing work of gathering the necessary insights in the DataBait architecture, generating legal requirements, transforming them into technical specifications during intense mutual collaboration between the technical,

social and the legal partners, and coordinating the integration of these specifications based in legal requirements in the DataBait architecture.

In section 2 of this report we present a table with all the software used in DataBait and the type of IP license under which they are used. In section 3 we present the data from the pre-pre-pilot which were made accessible (in accordance with the Data Licensing Agreement) to some of the USEMP partners for use on their own premises (in contrast to all other data, which are stored at the premises of HWC and can only be accessed on the HWC server).

In section 4 we present a list of the interdisciplinary informational exchanges with other tasks and other WPs which contributed to this legal coordination task (T3.6).

2. List of software components

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/
	image analytics library	BSD 2-Clause Licence (+ CEA Components)	http://caffe.berkeleyvision.org/
	Java-ML	GPLv2	http://java-ml.sourceforge.net/
	commons-configuration	ASLv2	https://commons.apache.org/proper/commons-configuration/
	org.apache.httpcomponents	ASLv2	https://hc.apache.org/
	commons-dbcp	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/

3.WP7 pre-pre pilot experimental datasets

This section describes the data that were shared between technical partners from the pre-pilot for evaluation/testing purposes:
Data should include some or all pre-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 that predicts personal attributes of users based on the users' likes, posted messages and detected visual concepts (for more details on this modules please see D6.4).

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

iMinds:
Data requested:
- Facebook Data.
- Survey responses

Data will be 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

4. Interdisciplinary information exchange with other WPs and tasks

WP2:

- Qualification of data
- Difference between legal and perceived sensitivity of personal data
- Correct use of 'privacy' (disclosure score instead of privacy)

The legal researchers and lawyers from the iCIS team were actively involved the preparation and writing of D2.3 on the content and data specifications in DataBait. This deliverable looked at the variety of data that is used during the development of DataBait and that is foreseen to be handled by the USEMP system once it is operational. The legal iCIS team interfaced with the other partners about the actual data which will be handled during system operation, the data from external datasets that are used to develop and evaluate the DataBait algorithms, data storage and the flow of data through the system. In order to ensure that DataBait is compliant with data protection law, qualification of all data was necessary (Which data are personal data? Which data are 'sensitive' in the sense of Art. 9 of the General Data Protection Regulation (GDPR)?). The deliverable also elaborated on the difference between legal sensitivity of personal data (Art. 9 GDPR) and perceived sensitivity of personal data (e.g. while 'income level' is not an Art. 9 type of data, some end-users of OSNs might perceive such information as being intimitate and sensitive). Another issue that was addressed is the correct use of the term 'privacy'. Because 'privacy' has a very specific meaning within the field of law, one of the main pieces of information provided to DataBait users was renamed "disclosure score" instead of the earlier "privacy score" – which covers the content of this score better. See also below, WP6. Last but not least, the deliverable links to the legal analysis conducted in WP3.

WP3:

- Interaction with T3.7 (D3.5) on socio-economic value of data → 'granular licensing' (is there a shared value for the industry and end-users based on transparency instead of opacity and 'less data')

The omnipresence of personalized ads and requests to consent with vague conditions when using internet services, are like tips of an iceberg, indicating the presence of an invisible economy of personal data. As shown in D3.5 this opaque situation is not only unpleasant for internet users, but also undesirable from an industry perspective. In D3.6 we studied the possibilities within the legal data protection and intellectual property frameworks to create a more transparent and fair business ecology with regard to the commercial value of OSN and browser data. The legal tool we explore in D3.6 is 'granular licensing': is it possible that data subjects, based on the specified purpose and within the confines of use limitation, can set defaults as the context for further processing as well as the type of data controllers with whom the data may be shared? The answer to this question isn't obvious – the purpose specification principle seems to be at odds with any 'generalities' about the purpose of the processing, so some legal 'acrobatics' is needed to create a solution that stays within the

boundaries set by this principle. What also has to be taken into account is whether commercial actors and users are actually helped by the solution of granular licensing. We explore these complex questions, at the intersection of the study in the socio-economic value of data, needs within the ecology of data driven business and legal empowerment of data subjects, in more detail in the final year of the USEMP project.

WP5:

- IP issues regarding with regard to the (various components of) the data driven DataBait modules

The main objective of WP5 is to propose and investigate the development of multimedia mining tools that can derive additional data from texts and pictures. The legal iCIS team had in-depth discussions with several of the technical partners (notably CEA, CERTH and Velti) about the exact functioning of these 'profiling' tools in order to assess if any intellectual property rights (IPRs) are infringed by DataBait profiling tools. This relates to software components used (see above, chapter 2), as well as the exact steps taken in the profiling process (see D3.7: Are IPR protected data copied during the process? Are their exceptions allowing the making of such copies? Which elements of the profiling process could be protected by IPRs? How do algorithms employed by DataBait relate to IPR protected algorithms? Etc.). The preliminary conclusions in D3.7 are that the likelihood that DataBait infringes on IPRs is not very high. Nevertheless there are some areas (database rights, trade secrets) where the possibility of an infringement could not fully be excluded and that need to be monitored during the remainder of the project. This will, again, require cloe collaboration with the technical partners.

WP6:

- Correct use of the term 'privacy'.

WP6 works towards the development of a scoring framework that quantifies the exposure of personal information of a user. Initially, this was framework was referred to as the "privacy scoring framework". Nevertheless, concerns were raised about the legal meaning of the term "privacy". In particular, "privacy" in the legal sense in EU law refers to Article 8 of the European Convention of Human Rights1 (right to respect for private life), which has very specific meaning (mainly a right that prevents power imbalances between state and citizen or citizen and other actors). Yet, the "privacy scoring framework" also contains information that does not fit this definition. For instance, the framework could represent the fact that a person's OSN data indicates that he/she likes swimming, pets or some specific brand; this does not fit the definition of privacy in the legal (human right) sense. Nevertheless, the inclusion of such information in the scoring framework informs the user about the types of information that he/she discloses through his/her digital trails. This can be interesting in terms of transparency and data protection, but data protection is not equivalent to privacy. Therefore, it was decided to use the more accurate term "disclosure scoring framework" in lieu of "privacy scoring framework". At the same time, it was decided to refer to the second main module produced within WP6 as the "disclosure settings framework" instead of the earlier "privacy settings framework".

WP8:

- Contributing to questionnaires (with 'legal' questions) for user research
- Leading role in collaborative work on explanatory DataBait animation

WP8 deals with the user evaluation of DataBait within the pilot studies. iCIS has actively given feedback on the drafting of the pilot studies as a whole and has generated questions, relevant from a legal point of view, to be posed to user in the qualitative interviews during the pilot. One of the main goals of these questions is that they explore how users are supported in the exercise of their informational rights through DataBait and how this support can be further ameliorated. A legal notion like 'profile transparency', and its contribution to citizen empowerment, is thus put to the test from a multidisciplinary perspective. Moreover, during the second year of the USEMP project iCIS took the lead in the very collaborative creation of an explanatory video on DataBait was produced (https://www.youtube.com/watch?v=dJinztt5PrA). Creating the content of the video was a truly interdisciplinary enterprise: the video aimed to appeal to users, support the informational rights of the data subject following from 10 and 11 DPD 95/46 towards the consortium as the DataBait datacontroller, and, obviously, also had to be a technically correct representation of the functioning of DataBait

5. Concluding remarks and planned further research

This report has presented the results of the legal coordination and the integration during the the period May 2015-December 2015.

As stated in the introduction of this deliverable, there is still quite some work-in-progress which will be reported in D3.13 (at the end of the project). This includes:

- adjustment of the DLA with regard to copyright
- an additional tab in DataBait with information about how the user can use DataBait information to effectuate legal rights
- further reporting on the licenses of the databases used for training and testing of the DataBait data driven modules
- inventory of which data will be processed with regard to Twitter and/or Instagram (the second OSN next to Facebook to which DataBait might be applicable)
- update with regard to the Facebook data which are processed, once the result of the review by Facebook of the DataBait Facebook API is known.
- update with regard to some issues relating to copyright and database rights, once the result of the review by Facebook of the DataBait Facebook API is known.
- update on the question if DataBait infringes on any copyright on graphic user interfaces once we have studied the latest version of the DataBait visualization in more detail
- additional research on lifecycle management [anonymisation & pseudonymisation]
- further exploration of how our research on a granular permission system for personal data (as reported in D3.6) and for copyright protected user generated content (as reported in D3.8) could impact the DataBait architecture [compliance check]
- further exploration of how our research into the empowering possibilities of image rights and personality rights can impact on the DataBait architecture
- checking on further IP issues [compliance check Facebook ToS, IP rights of USEMP]
- check what DataBait revision would be necessary in view of the role played by the concept of counter-performance other than money in art. 3(1) and (4), 13 (2) sub b and c, art. 15(2) sub b, 16(4) a and b in the upcoming Directive on Digital Content.

.