D3.6

Fundamental Rights Protection by Design for Online Social Networks - v2: Update of Deliverable 3.1

v 1.0 / 2015-05-19

Katja de Vries, Niels van Dijk, Sari Depreeuw and Mireille Hildebrandt (iCIS-RU).

Building on D3.1 this document presents an overview of the possibilities for fundamental rights protection by design (FRPbD), particularly Data Protection by Design (DPbD), in the context of behavioural tracking and personalized advertising based on the digital trail created by the use of Online Social Networks (OSNs) and browsers.

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)
Reviewers	Theodoros Michalareas (Velti) Ali Padyab (LTU)
Version	0.6
Status	Final
Dissemination level	PU: Public
Due date	2014-09-30
Delivery date	2015-10-07
Revised delivery	2016-02-15

Version	Changes
0.1	Initial Release, Katja de Vries and Niels van Dijk (iCIS)
0.2	Review Ali Padyab (LTU)
0.3	Review Theodoros Michalareas (Velti)

0.4	
0.5	Review Sari Depreeuw (iCIS)
0.6	Review Mireille Hildebrandt (iCIS) Minor adjustments by Katja de Vries and Mireille Hildebrandt following the comments of the Commission.

Table of Contents

Summary

Building on D3.1 this document presents an overview of the possibilities for fundamental rights protection by design (FRPbD), and more specifically data protection by design (DPbD), in the context of behavioural tracking and personalized advertising based on the digital trail created by the use of Online Social Networks (OSNs) and browsers. We focus on end users' rights derived from data protection law. Other relevant fundamental rights protection of internet users can be derived from Art. 8 ECHR (respect for private life) and EU anti-discrimination law. We discuss the latter in relation to the protection for the processing sensitive data in data protection law. Respect for private life is discussed in D3.8 as part of the analysis of portrait rights of internet users with regard to profiling practices.

D3.1 combined the legal analysis of these fundamental rights with a critical reflection on the architectural design of the DataBait tool created by the USEMP consortium. This resulted in a set of practical specifications for the the design of this tool (both the part giving insight and control about what the user discloses in her digital trail and who is tracking it, and the part providing awareness about the value of her digital trail and her Facebook friends and/or Twitter followers), based on legal design requirements which drive, frame and complement the technical and social requirements. The main contribution of D3.1 was the development of the so-called Data Licensing Agreement (DLA) that enables OSN users to license the processing of their personal data in compliance with current EU Data Protection Law in exchange for the use of a profile transparency tool (the DataBait tool). This license was developed for a consortium that processes these data for a purely scientific purpose (the USEMP consortium).

In this deliverable (D3.6) the DLA is further elaborated into a *modular* version that allows for more *granular* licensing of personal data processing.

What do we mean with *modularity*? The DLA signed between the USEMP consortium and each user of *DataBait* (i.e., the profile transparency tool developed in the USEMP project) is specifically tailored for this particular tool. However, by presenting a *modular* version of the DLA we show that the basic structure of the DLA is easily generalizable to similar transparency tools provided by other providers: only a few modifications are needed. We explore how the DLA should be modulated if the provider of a transparency tool was not a scientific consortium, like USEMP, but (a) an OSN, (b) a third-party commercial provider (i.e., a business offering the profile transparency tool in exchange for a fee or some other remuneration), or (c) an NGO (e.g. a civil society or consumer organization) or private nonprofit organization with a public goal (e.g. a charitable organization). Thus, the modular contract presented in this deliverable indicates which parts of the current DLA need to be adjusted if the tool was to be provided by a different type of provider, if the architecture of the transparency tool would differ from DataBait and if the tool targeted different OSNs/browsers.

Granular licensing means that, based on the specified purpose and within the confines of use limitation, data subjects can set defaults as the context for further processing as well as the type of data controllers with whom the data may be shared. This means that sharing and re-use are made into a topic of formalized 'negotiation' between data subject and controller, supported through a fixed set of standardized licensing options which might be presented in a layered format: a legal-technical format, a common sense format, and a

machine readable. With regard to the content and scope of the granular licenses, which could be included in the DLA, we draw an analogy with the distinction between functional and non-functional cookies for cookie consent (Art. 5(3) e-Privacy Directive. We also provide examples of various more fine-grained distinctions for granular licenses which would help a user to choose the purposes for which her data can be used. We elaborate on the format of the granular licenses by drawing an analogy with the Creative Commons licensing system in copyright to see how pre-set licensing settings and a layered structure of pictograms and legal text could be used to provide the user with transparency about the data processing she agrees to.

In addition to the modular and granular DLA this deliverable proposes four other sets of legal requirements applicable to profiling in OSN settings and tools which aims to provide transparency about it. The first set of requirements regards the empowerment and support of internet users in effectuating their informational rights (notably supporting “profile transparency”) towards OSNs. The second set regards the anonymization/pseudonimization of profiling data. The third set concerns requirements for the handling of sensitive data (Art. 8 DPD) and the creation of awareness with regard to profiling with possibly illegitimate discriminatory applications. The last set of requirements regards how to ensure profile transparency about the functioning of a profile transparency tool it self.

1.Introduction

1.1. The limits of purpose limitation?

The European legal framework for data protection enables the processing of personal data within the EU whilst making sure that this happens in a way which is lawful, fair and transparent. The *lawfulness* of the processing in ensured through the requirement that the personal data of European citizens are processed *based on a legal ground*, for example because the data subject, i.e. the person to whom the data relate, has given explicit consent for the processing or because the processing in necessary for the performance of a contract. Processing data just because it's practical or interesting is not a lawful ground. *Fairness and transparency* of the processing is guaranteed through the requirements that any processing has to be justified *by a specified, explicit and legitimate purpose* and that further processing for other incompatible purposes is not permitted (the so-called purpose specification and limitation principle), that the data processing is *in accordance with the proportionality principle* (that the processing is adequate, relevant and not excessive in relation to the purpose and that the data are accurate and up-to-date), that processing happens in a *secure* way (security has to be preserved though appropriate state-of-the-art organizational and technical measures) and that the data subject can exercise her *informational rights* (i.e. be informed about the identity of the controller, the purpose of the processing, and her right of access and rectification).

Some of these requirements of EU data protection law, particularly purpose limitation, proportionality and the effectuation of informational rights, are sometimes difficult to reconcile with the ecology of business practices and models clustered around the use of OSNs and browsers. Despite the existing data protection legislation, for many European internet users it is far from evident what happens to their digital trail, who has access or knowledge of their data, how data analytic software is used to derive profiles from their raw personal data, how profiles are applied to them, which actors are involved in the economy with regard to their personal data, and how this economy functions. Pop-up windows asking for consent for the use of cookies, long and complicated terms of service which hardly anyone reads before clicking 'I agree', and personalized ads following internet users around the internet, signal to internet users that their personal data are probably processed as part of a data economy, yet the precise nature of this economy is shrouded in mystery. 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 this deliverable we argue that everyone would benefit from a transparent handling of data where there is mutual agreement with regard to the context for further processing as well as the type of data controllers with whom the data may be shared, and where it is clear to *all* actors how and for what purposes a piece of personal data can be processed. Neither internet users, nor industry benefit from a situation where industry operate in a grey legal zone. Not only does this entail for the industry that they are unsure if their data processing actions are within the boundaries of law, but the opaque situation also is more likely to result

in personal data of unknown quality and origin. Ideally, a user would be aware of the fact that she has a profile which categorizes her as someone who, for example, enjoys good wines and loves books, and could adjust her information (e.g., '*No, I'm not interested in books about gardening and I don't like Chardonnay, but I do like Merlot wines and Italian twentieth century literature*'), and specify in which ways this information can be used (e.g., *'I do like to receive offers for Bordeaux wines, but I don't want the data about my drinking habits to be available to health insurers'*). While such a transparent and voluntarist system could decrease the quantity of data available to certain parts of industry (e.g., because many users might choose to prohibit the use of their data for price differentiations for services such as insurances), it would increase the reliability of the data and provide legal certainty about the purposes for which these data can be processed.

In the current debates around the proposed new European Data protection legislation[1] (the so-called *General Data Protection Regulation*[2] [pGDPR] which will succeed the current *Data Protection Directive 95/46/EC*[3] [DPD 95/46]) civil rights and industry concerns have become strongly polarized. Parts of industry have been pleading to make re-use of personal data easier by relinquishing the purpose limitation principle. Adjusting

[1] The proposed *General Data Protection Regulation*[1] [GDPR], the successor to DPD 95/46) http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf Unless stated otherwise we refer to the text of the pGDPR as proposed by the European Parliament: http://www.janalbrecht.eu/fileadmin/material/Dokumente/DPR-Regulation-inofficial-consolidated-LIBE.pdf The version originally proposed by the Commission can be found here: http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2012/0011/COM_COM%282012%290011_EN.pdf . A comparison between the versions proposed by the Commission, the Parliament, the Council and the recommendation of the European Data protection Supervisor can be found here: https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-07-27_GDPR_Recommendations_Annex_EN.pdf . <all aforementioned documents were accessed 7 September 2015>

[2] The proposed *General Data Protection Regulation* (GDPR) is currently being created in the so-called *ordinary legislative procedure* (formally known as the *codecision procedure*) of the EU, which is basically a bicameral legislative procedure: it gives the same weight to the European Parliament and the Council of the European Union (consisting of ministers from the 28 EU Member State governments). The GPDR was first proposed on 25 January 2012 by the European Commission (that is, the executive branch of the EU and the only EU institution empowered to initiate legislation) and now has to be jointly adopted by the European Parliament and the Council. The text proposed by the Commission [*Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, Brussels, 25 January 2012, COM(2012) 11 final] has been subjected to a first reading by the European Parliament and has been amended the on 12 March 2014 [*European Parliament legislative resolution on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, Strasbourg, 12 March 2014, COM(2012)0011 – C7-0025/2012 – 2012/0011(COD), online available at : <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2014-0212+0+DOC+XML+V0//EN>]. The amended text was examined by the Council of the European Union. In June 2015 the Council proposed a whole series of amendments in a document entitled the 'General Approach'. This document forms the basis for the so-called trilogue discussions (between Parliament and Council, facilitated and mediated by the Commission) which will take place in July-December 2015. If these discussions do not result in agreement over the proposed legislative text between Parliament and Council, it can go back and forth between Parliament and Council up to three times. A clear infographic clarifying the ordinary legislative procedure can be found here : <http://www.europarl.europa.eu/aboutparliament/en/0081f4b3c7/Law-making-procedures-in-detail.html> [last accessed 29 September 2014]. Looking at the current status of the proposed General Data Protection Regulation and the steps in the legislative procedures still to be taken, the GPDR will most likely enter into force by 2016.

[3] *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data*, Official Journal L 281, 23/11/1995, p. 31-50. The DPD 95/46/EC is currently the main legal instrument regarding general data protection, but is in the proces of being replaced by the proposed *General Data Protection Regulation*, which will probably enter into force in 2016.

legislation to meet this plea would however empty EU data protection from one of its core ideas, namely preventing personal data from circulating around for other purposes than the one for which they were initially collected. Such proposals have therefore, justly, been met with fierce criticism from the data protection community. The reaction of Working Party 29[4] to a proposal made by the Council, which would facilitate further processing for purposes incompatible with the original one, is exemplary:

> "The Working Party is very much concerned about the proposed provisions on further processing, especially in the context of Big Data. In fact, according to the Council, it will be possible for a data controller to further process data even if the purpose is incompatible with the original one as long as the controller has an overriding interest in this processing. […] The Working Party considers that this situation would render one of the fundamental principles of the data protection framework, the purpose limitation principle, meaningless and void. […] Such an approach, which conflates the notions of legal basis and further processing for compatible purpose, contradicts the EU data protection acquis and would be illegal under the current legal framework. It could furthermore have no other consequence but to undermine the whole new data protection framework and to dilute the level of protection for EU citizens in comparison to Directive 95/46/EC in force." (Article 29 Data Protection Working Party 29, 2015)

Data being re-used for purposes which no one foresaw at the time of collection, or even worse: data being transferred from one data controller to another as if they were simple goods that can be used for whatever purpose their 'owner' deems right, is in absolute opposition to the purpose specification and limitation principle (see section 5.1) which belongs to the core of EU data protection law. Moreover, the EU data protection regime prescribes that personal data can only be processed lawfully when two conditions are fulfilled: firstly, there has to be a *legal ground* (such as consent or a contract, Art. 7 DPD 95/46), which provides a legitimizing reason for the processing, and secondly, the data should only be processed for the explicitely specified purpose for which they were collected and no further processing for incompatible purposes should occur (*purpose specification and limitation*). This means that purpose specification and limitation cannot be simply consented or contracted away; for this would conflate two separate conditions (legal ground and justification through purpose specification) for lawful data processing into one. As recently underlined by the European Data Protection Supervisor (2015), it would be a black day for fundamental rights protection of EU citizens if the pGDPR would relinquish[5] the double requirement that processing should both be lawful (that is, based on one of the legal grounds enumerated in Art. 7 DPD 95/46) and justified by an explicit and specified purpose (Art 6(1) DPD 95/46). This double requirement means that even if the data subject explicitly consents

[4] WP29 is the independent European advisory body with regard to data protection installed by Art. 29 of the current Data Protection Directive 95/46,

[5] "All data processing must be *both* lawful *and* justified. The requirements for all data processing to be limited to specific purposes and on a legal basis are cumulative, not alternatives. We recommend avoiding any conflation and thereby weakening of these principles." (European Data Protection Supervisor, 2015, p. 5)

to re-use of data for a purpose which is incompatible with the one for which the data was originally collected, this would infringe on the purpose limitation principle and thus be unlawful under the current legal regime. While it might be fruitful if the new legislation would allow for an exception in the case of explicit, informed and freely given consent by the data subject ('fair re-use' for an incompatible purpose; (Koning, 2014)), so that data do not need to be collected anew when the data subject agrees[6] on the further processing, we agree with the EDPS that it is essential that the double requirement should be preserved.

If the current data protection regime is so strict, how is this compatible with the situation where users are tracked and targeted with personalized ads despite the fact that they move from one website to another? We will now discuss three legal 'loopholes' which can sometimes be used in an intermingled way: a vague and broad description of the processing purposes at the stage of getting user consent (in the general terms and conditions) undermining the rationale of purpose specification, a loose interpretation of what counts as a "compatible purpose" for the processing of personal data, and the possibility for the data controller (who has stated this as a processing purpose) to offer a multiplicity of actors the possibility to 'address' an end-user based on her profile (targeted ads), without disclosing or transferring any actual data to these actors.

Before discussing the details of each of these three loopholes, we first need to take a closer look at the exact formulation of the principle of purpose limitation in Art. 6(1)(b) DPD 95/46:

> "personal data must be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible […]."

The first sentence of Art. 6(1)(b) consists of two parts. The first half ("*personal data must be collected for specified, explicit and legitimate purposes…*") expresses the "principle of purpose specification". Each of the three elements (the *specificity* of the purpose, its *explicit* communication to the data subject and its *legitimacy*) are important and non-negotiable requirements (see below, section 5.1 of this deliverable, for an extensive discussion). The second half of the first sentence of Art. 6(1)(b) articulates the "compatibility clause" *("…and not further processed in a way incompatible with those purposes*"). Taken together the purpose specification principle and the compatibility clause constitute the principle of purpose limitation. The principle of purpose limitation expresses the basic idea that the processing of personal data should stay within the boundaries of the initial purpose defined by the data controller (the person or body who determines the purposes and means of the processing of personal data) so that the data subject (the person identifiably related to the personal data) knows what to expect and who to address with queries or complaints. This

[6] The risk here is that a data subject could be nudged ("Please consent to these new terms of service – don't bother reading them, just click 'yes'") or forced ("Please consent to these new terms of service – otherwise we'll deny you access to this service") to agree with further processing. This could be avoided by prohibiting contractual clauses that deny continuation of service in case the data subject does not consent, and by requiring the consent to be explicit and informed.

creates foreseeability based on legitimate expectations and prevents data processing getting out of bounds[7].

The first 'legal loophole' employed by the industry is the open texture of purpose specification, following from the fact that the term 'specified and explicit purpose' has not been tested in a court of law yet. Users of internet services, such as browsers or OSNs, often are asked to consent to the processing of their data for a long list of vaguely defined purposes. This renders void the prescription that purposes should be specified and explicit. When the compatibility clause is combined with an initial specification of purposes which is broad and vague, it might almost seem that everything is possible. Thus, while "collecting personal data for a specific commercial transaction with a customer, and later on deciding to export the data to another firm for the purposes of direct marketing is unlawful" (European Commission, 2012), this transfer would be lawful if the transfer were included in the initial purpose specification. How far can this provision be stretched? Working Party 29 (WP29) writes in its Opinion on purpose limitation:

> "Vague or general purposes such as 'improving users' experience', 'marketing', 'IT-security' or 'future research' will - without more detail - usually not meet the criteria of being 'specific'. However, the degree of detail in which a purpose should be specified depends on the particular context in which the data are collected and the personal data involved." (Article 29 Data Protection Working Party 29, 2013, p. 52)

The boundary between 'too vague' and 'specific enough' is not always easy to draw. Is a long list of specific purposes (a company might decide to 'play it safe' and include as many purposes as possible) in accordance with the principle of purpose specification? It seems doubtful that such an overly inclusive list would be considered legitimate and fair in the sense of Arts. 6(a) and (b) DPD 95/46 (see also below, section 5.1.2) if tested in Court; however – much will depend on the particular context. Let's explore a more concrete example: is it legitimate for a company to state that the purpose of the processing includes transferring data to "corporate affiliates and affiliates' services" and selling your personal data in case of a bankruptcy or merger?[8] All major online sites such as Facebook, Twitter, Google and LinkedIn include very similar transfer clauses[9] in their terms of service. For example, Twitter states:

> "Business Transfers and Affiliates: In the event that Twitter is involved in a bankruptcy, merger, acquisition, reorganization or sale of assets, your information may be sold or transferred as part of that transaction. This Privacy Policy will apply to your information as transferred to the new entity. We may also disclose information

[7] "Purpose specification and the concept of compatible use contribute to transparency, legal certainty and predictability; they aim to protect the data subject by setting limits on how controllers are able to use their data and reinforce the fairness of the processing. The limitation should, for example, prevent the use of individuals' personal data in a way (or for further purposes) that they might find unexpected, inappropriate or otherwise objectionable. At the same time, the notion of compatible use also offers some degree of flexibility for data controllers". (Article 29 Data Protection Working Party 29, 2013, p. 11)

[8] https://support.twitter.com/articles/20172501

[9] http://www.nytimes.com/interactive/2015/06/28/technology/Firesale-Listy.html?; http://www.nytimes.com/2015/06/29/technology/when-a-company-goes-up-for-sale-in-many-cases-so-does-your-personal-data.html?

about you to our corporate affiliates in order to help provide, understand, and improve our Services and our affiliates' services, including the delivery of ads."[10]

Such a clause may hold up in a court of law to the extent that the purpose remains the same or is found to be compatible, and insofar the Privacy Policy is indeed applied. However, in practice such transfers of data as a kind of 'assets' from one company to another will render the foreseeability of how the next company will handle the data more difficult and this may be said to go against the purpose of purpose limitation. This could be resolved by requiring a notification to the relevant data subjects, combined with a right to withdraw one's data if provided on the legal grounds of consent or the legitimate interest of the data controller. Such a withdrawal could perhaps be based on art. 12 or 14 of the current DPD 95/46. Less protection, however, remains if the aforementioned proposal of the Council, allowing further processing for an incompatible purpose as the controller has an overriding interest[11] in the processing, would be incorporated in the proposed GDPR. We strongly oppose such a lenient approach as it renders effective protection illusionary. Instead, we argue for clear and precise rules on how to proceed when personal data are transferred as part of a package deal in the case of bankruptcy, mergers, acquisitions and reoganization or sale of assets.[12]

This brings us to the second legal 'loophole' , which consists in a loose interpretation of the "compatible purposes" clause (*"…and not further processed in a way incompatible with those purposes*"), extending the legal uncertainty of the purpose limitation principle following from vaguely formulated purposes even further.

If the principle of purpose specification were not complemented by the "compatibility clause", re-use of personal data (whether by the same of another data controller) would become very much constraint. Especially if the one determining the purpose changes, the purpose will easily change. Since the goal of purpose limitation is not to prohibit reuse and transfer of data but to safeguard the legitimate expectation of the data subject, the 'compatibility clause' should ensure that this expectation is not violated. A transparent relationship between data subject and data controller requires that the controller remains the only one who is in control over the data and that the processing is limited to the initial purposes.

So how is it possible that personal data seem to be transferred between different actors all the time? Why do data seem to circulate so easily on the internet? This is mainly[13]

[10] https://twitter.com/privacy
[11] It should be noted that 'overriding interest' would be an open-ended legal notion, requiring further interpretation in further case law.
[12] Speaking of 'clear and precise rules' refers to the legal requirement for justification of an infringement of the fundamental rights to privacy and data protection, as stipulated by the CJEU, following the case law of the ECHR. Cf. e.g. par. 54 in CJEU, 18 April 2014, C-C-293/12 and C-594/12 (*Digital Rights Ireland v Irelan*d).
[13] All transfers for incompatible purposes are prohibited *unless* required by legislation of the Member State to which the controller is subject; for example if the processing of the personal data is necessary for carrying out a task in the public interest. This means that a provider of an online service such as Facebook or Twitter can, and sometimes even has to, transfer data to authorities or public bodies (e.g. the police or social services) for incompatible purposes if this is required by national law. Art. 13 DPD 95/46 lists that exemptions to the purpose limitation of Art 6(1) can be made in national legislation if such exemption safeguards one of the following interests: (a) national security; (b) defence; (c) public security;(d) the

due to a broad interpretation of the aforementioned “compatible purposes” clause which states that further processing is allowed as long as the purpose is not *incompatible* with the original purpose. Further processing of data for historical, statistical or scientific purposes shall be considered as compatible with the original purpose. Here we end up in a matter of definition: for example, can market research be qualified as a scientific or statistical purpose?

> "What ‘compatible’ means, however, is not defined and is left open to interpretation on a case-by-case basis." (European Union Agency for Fundamental Rights, 2014, p. 69)

Clearly, industry prefers to give a broad interpretation that gives the data controller quite some freedom to process data for other purposes than the initial one (as long as these purposes are ‘compatible’). The “compatibility clause” also allows for transfers between private entities as long as the purpose of the transfer is compatible with the one for which the data were initially collected and processed. We may, however, expect that once data protection legislation is appropriately enforced, data controllers will have to be more specific about reuse, turning the loophole into a safeguard instead of a vulnerability. Such enforcement is to be expected once the pGDPR comes into force, but even at this moment there seems to be a momentum for holding data controllers responsible for their stewardship of personal data. This is clear from the case law of both European Courts and the case law within the Member States of the EU, notably when based on tort law, while various types of class actions have been highly successful in challenging the unfettered sovereignty that some data controllers exercize over personal data.[14]

(European Commission, 2012) The third ‘loophole’ is strictly speaking not a loophole – because it does not result in the circulation of personal data. Yet, it does result in the *semblance* of data circulating around the web, intransparency and some possibly negative effects similar to when data would actually circulate. To use an analogy: imagine that you tell your mailman not to tell anyone where you live and who you are. Still, you receive an abundance of letters which seem to be very specifically tailored to address you (e.g. a single mother with a low income and an interest in cars). You reproach your mailman that he has told everyone about but he cunningly replies: “*No, I did not – I just told everyone that if they wanted to sent mail to a single mother with a low income and an interest in cars, that I would gladly pass the message on*”. The third ‘loophole’, which this analogy illustrates, is that in targeted advertising it’s often not the actual data which are being sold but the *possibility to*

---

prevention, investigation, detection and prosecution of criminal offences, or of breaches of ethics for regulated professions;(e) an important economic or financial interest of a Member State or of the European Union, including monetary, budgetary and taxation matters; (f) a monitoring, inspection or regulatory function connected, even occasionally, with the exercise of official authority in cases referred to in (c), (d) and (e);(g) the protection of the data subject or of the rights and freedoms of others.

[14] See art. 73-79b pGDPR (version of the Council), including substantial fines, tort liability and the right to be represented by a non-profit association; e.g. CJEU 18 April 2014, C‑293/12 and C‑594/12 (Digital Rights Ireland) and CJEU 13 May 2014, C‑131/12, (Google Spain v Costeja Gonzalez), Court of Appeal of England and Wales, 27 March 2015 Google Inc v Vidal-Hall & Ors [2015] EWCA Civ 311 (confirming a privacy tort for the violation of data protection rights), available at <http://www.bailii.org/cgi-bin/markup.cgi?doc=/ew/cases/EWCA/Civ/2015/311.html> and the draft legislation pubished February 2015 by the German Government to enable class action for the violation of Data Protection rights (*Entwurf eines Gesetzes zur Verbesserung der zivilrechtlichen Durchsetzung von verbraucherschützenden Vorschriften des Datenschutzrechts*, available at <http://www.cr-online.de/36818.htm>. On the role of the ECHR in the protection of personal data in the private sector, e.g. F. Boehm, Information Sharing and Data Protection in the Area of Freedom Security and Justice (Springer 2012), 75-76.

target customers with a certain profile.[15] This means that the personal data of a user of an OSN service are not transferred but that ads are routed towards certain profiles without revealing the actual identity of the owners of these profiles. Thus, what is sold is targeted advertising space and not personal data. Calling this business model a 'loophole' might be considered controversial as it does not involve a loose or broad interpretation of a data protection requirement (as with the two other 'loopholes'). An OSN who has targeted advertising as one of the purposes of processing can stay the sole data controller and yet an end-user is addressed by ads of all kind of actors, giving the impression that her personal data are transferred. However what circulates is not the user data but a user 'profile'. We would argue that the user would benefit from more transparency regarding such business practices and that more attention could be devoted (which is partly the case in some versions of the pGDPR) to the regulation of possible adverse, discriminatory, effects of targeted ads and personalization of settings. While the business practice of companies acting through intermediaries, without access or control over the actual personal data, is clearly preferable over a practice where the actual data are transferred, this set-up might also make it more difficult for the data subject to know whom is liable for what and whom to address with complaints or questions.

1.2. Four research strands: closing some loopholes

Instead of operating in a grey legal area of loopholes, all actors would benefit from mutual agreement on, and transparency about what kinds of processing are permitted for a piece of personal data. Mutual agreement and transparency are interrelated: when all actors (data subjects and data controllers) have a clear understanding of how the controller intends to process and share the data and how the involved data should be qualified in terms of EU data protection law, this forms a point of departure for informed decision-making and mutual agreement. Mutual agreement, in turn, adds to mutual understanding and transparency. Thus, to enable such mutual agreement and transparency, we explore four research strands with regard to data protection within the USEMP project. Each of these four strands of research contribute to the Data Protection by Design (DPbD)[16] realized by the DataBait tool, which is created within the USEMP project. The outcome of each research strand are legal requirements, which are then translated in technical specifications for the design of the DataBait tool.

The first research strands looks at how *technological and organizational transparency* can be provided and supported for data processing in general, and for profiling in particular. This research strand looks both at *empowerment* through a profile transparency tool and how the tool itself should be *compliant* with data protection law. With regard to empowerment we look at how a profile transparency tool (such as the DataBait tool developed in the USEMP project) can strengthen and support the informational rights a data subject has towards a

[15] For example, a commercial company selling particular type of clothing might want to target pregnant women, who are older than 35 and live in Paris, Lyon and Marseille. The only information this company would get is the amount of women targeted and feedback (e.g. do women in Paris click the ad more often than the ones in Lyon?). The commercial company never gets the access to the personal data; it is the OSN who stays in control.
[16] See section 1.3 for a clarification of the notion *Data Protection by Design*.

data controller. The main achievement of the DataBait tool in terms of DPbD is to empower the data subject (supporting the exercise of informational rights towards OSNs, such as Facebook and Twitter, and browsers, such as Chrome and Firefox) by providing her insight in her own raw data (what information do I share? who is tracking me?) and by showing what can be derived from these data. However, next to the research into how *DataBait* can be made into a strong profile transparency tool supporting the exercise of rights following from the data protection framework, making DataBait itself *compliant* with data protection law also provides a way of experimenting with optimal forms to comply with the requirements of (profile) transparency and fairness of data processing. DataBait users sign a contract before using this profile transparency tool, which avoids any unnecessary 'legalese' and specifies the purposes and the process of the data processing in a very comprehensive way. Furthermore, *DataBait* offers a clear interface where one can request data deletion and withdraw consent for the processing of one's sensitive data. Additional information to further enhance the transparency and fairness of the processing is provided in the '*DataBait: how, what and why?*'-section, in the form of flow-charts, an explanatory animation, lists of collected data types, a button which allows users to download all their data, and practical contact details for the exercise of the right of access and rectification.

The second strand gives *a legal clarification* of which data should be considered *'sensitive'* in the sense of Art. 8 DPD 95/46, and which data can be considered *anonymous* (i.e., not personal data and therefore outside the scope of DPD 95/46) and shows how our clarification of these two contentious legal notions would translate into DPbD requirements.

The third strand contrasts *'contract'* (Art. 7b DPD 95/46) *as a legal ground* for data processing with 'consent' (Art. 7aDPD 95/46). Contract is the legal ground (Art. 7 DPD 95/46) for all data processing taking place within the USEMP project. This is innovative, because most data processing on the internet takes place based on the legal ground of consent. The DataBait contract, which we call a Data Licensing Agreement (DLA), shows how, compared to the one-sided 'take-it-or-leave-it-approach' of consent, the legal ground of 'contract' might enhance the power balance between the data subject and the controller by creating mutual duties and rights. In this way, it turns both into contracting parties, thus exceeding their roles of 'mere' data subject and data controller. For such a contract to offer better protection than consent, the content of the contract is – obviously – crucial. Not any contract will be beneficial or empowering for data subjects. The protection of the USEMP DLA depends on the transparency it provides on the (1) existence of profiling, (2) the underlying logic or backend of the system and (3) the indication of potential consequences of being targeted as a specific type of consumer, citizen, user, and person. Moreover, we explore the possibility of a *modular* DLA, which could be used by different types of providers of profiling transparency tools (modularity 1) and by different OSNs and browsers (modularity 2). With regard to the first modularity we show how the DLA should be modulated if the provider of a transparency tool was (a) an OSN, (b) a third-party commercial provider, (c) an NGO (e.g. a civil society or consumer organization) or aprivate nonprofit organization with a public goal (e.g. a charitable organization). With regard to the second modularity we explore how the DLA looks if the profiling transparency tool applies to (a) Facebook, (b) Twitter, (c) Both Facebook and Twitter, (d) a browser like Chrome or Firefox.

The fourth strand draws an analogy between USEMP's Data Licensing Agreement and Creative Commons licenses and shows how a contractual *granular licensing system* could be a form of DPbD with regard to the requirements of purpose specification and transparency with regard to the processing. By offering a set of default forms of permitted

uses of personal data from which the data subject can choose –a concise, pre-set, and transparent 'menu' of options- the purpose of the processing would no longer be a singular offer made by a data controller for the data subject to consent to or reject. We provide examples of thematic (medical research, commercial personalization, etc.), institutional (NGO, non-profit, commercial, etc.), integrity based (eco-label, trust certificate, etc.), action based (no derived data, no application of existing profiles, etc.), data type (e.g. 'sensitive' data), and 'degrees of separation'-based (amount of controllers, time limitation, etc.) granular licenses which would help a user to choose the purposes for which her data are used. A granular licensing system would thus allow a user to give an explicit and specific license for the processing purposes and usages permitted with her personal data. This license could be provided as part of a contract between the provider of an internet service and their users. However, we also explore the possibility of software solutions supporting this type of granular licensing.

1.3. DPbD – the making of a transparency tool

Each of the four aforementioned ways of enhancing mutual agreement and transparency are a form of *Data Protection by Design* (DPbD[17]). In this section we explain the meaning of that notion and in which way the *DataBait* tool should be considered as embodying various forms of DPbD.

The terms "Data Protection by Design" (DPbDesign) and "Data Protection by Default" (DPbDefault), which have a prominent place in the pGDPR (Art. 23), are not explicitly mentioned in the current DPD 95/46. However, Article 17 (*Security of processing*) of the DPD can be seen as a first step towards DPbDesign:

> "*Security of processing***.** Member States shall provide that the controller must implement appropriate technical and organizational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing. Having regard to the state of the art and the cost of their implementation, such measures shall ensure a level of security appropriate to the risks represented by the processing and the nature of the data to be protected." (Art. 17(1) DPD)

Next to a very similar requirement of taking "appropriate organizational and technical measures" in the context of the security of the processing (art. 30 of the proposed GDPR), the proposed GDPR also contains a general article on *Data Protection by Design and by Default* (Art. 23 GDPR) which does not merely relate to the *security* of the processing but aims to meet *all* requirements of the proposed GDPR:

> "Having regard to the state of the art and the cost of implementation, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and

[17] In this deliverable we use the abbreviation DPbD for Data Protection by *Design*, unless we want to specifically distinguish between Data Protection by *Design* and by *Default*. In that case we abbreviate these notions as DPbDesign and DPbDefault.

> organisational measures and procedures in such a way that the processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject." (Art 23(1) of the proposed GDPR)

Despite the seemingly extensive definition of *Data Protection by Design* in Art. 23(1) an exact understanding of this notion is still heavily debated. Article 23(2) obliges the data controller to implement mechanisms to ensure *Data Protection by Default*, which is a certain form of *Data Protection by Design* based on the idea "that privacy intrusive features of a certain product or service are initially limited to what is necessary for the simple use of it" (European Data Protection Supervisor 2012, 7 March, p. 29-30). However, as the European Data Protection Supervisor (EPDS) argued in its Opinion on the GDPR, Article 23(2) does not give "any clear substance" to "data protection by default":

> "The first sentence does not add much to the general principles of data processing in Article 5, and the data minimisation principle in Article 5(c) in particular, except from the confirmation that such principles should also be embedded in the design of relevant systems." (European Data Protection Supervisor 2012, 7 March, p. 29).

Obviously, data protection requirements are not the only legal requirements which could be used to shape the technological and organizational design of IT architectures. Theoretically one could strive to use all kinds of fundamental rights ("Fundamental Rights Protection by Design" or "Legal Protection by Design") to shape the technological and organizational design of IT architectures. The latter notion, Legal Protection by Design (LPbD), is a term first coined by Hildebrandt (2011) conveying the idea that legal norms can be articulated in architecture and which is especially concerned with the articulation of *fundamental rights* in *ICT architecture*. LPbD is based on the idea that "the legal requirements of fundamental rights such as privacy and data protection must be translated into computer system hardware, code, protocols and organizational standards to sustain the effectiveness of such right in a changing technological landscape."[18] (Hildebrandt 2013, p. 10)

When we try to imagine how the right to profile transparency could be transposed into the technological and organizational design of systems and practices which profile end-users, tools like the one developed in the USEMP project (the *DataBait* tool) could be the answer. When the proposed GDPR comes into force, and DPbD becomes an enforceable legal requirement; the DataBait tools can be a good example of how profile transparency could be built into otherwise opaque automated profiling systems. In this sense the DataBait tool can act as the technical "extension" or mouthpiece of data protection law.

The DataBait tool offers DPbD solutions to strengthen various legal requirements from EU Data Protection law: it supports the requirements[19] of **(a)** offering technological and

[18] LPbD should not be confused with techno-regulation, such as DRM. It is based on articulating legal constraints, not an attempt to automate compliance behind the back of the data subject. This entails (1) democratic participation and legislative legitimization of LPdB, (2) in-built contestability (3) contestability of its implications in a court of law. LPbD requires deliberation over the question of which hard and fast rules must be hardwired, firmwared or softwired into the architecture of the Internet and the upcoming cyberphysical infrastructures of the Internet of Things.

[19] The USEMP project is *required by law*, like any data controller, to process the data of EU citizens in *compliance* with EU data protection law. It should be noted that, of course, the USEMP project is not required by law to strengthen the informational rights of data subjects towards big OSNs and browsers. However, we use the notion

organizational **transparency** with regard to data processing performed by browsers and OSNs (user empowerment) and its own data processing (compliance), **(b)** processing personal data based on a **legal ground**, **(c)** enhancing a **level playing** field between data subject and data controller, **(d)** strengthening the principle of **purpose specification**, **(e)** processing **sensitive data** (Art. 8 DPD 95/46) in accordance with an extra strict and careful regime, **(f) anonymizing** or **deleting** personal data as soon as their processing is no longer necessary (data minimization). In order to create design solutions with regard to the two latter requirements, we clarify the meaning of the ambiguous notions 'sensitive data' and 'anonymization' (chapter 3). The requirement of processing data based on a legal ground is realized through the creation of a particular legal contract (chapter 4). This contract, the so-called Data Licensing Agreement, is also a design translation of the requirements of transparency and enhancement of power equality between data subject and controller. The requirement of profile transparency is further translated through a set of design solutions described in chapter 2 ('*Supporting profile transparency'*) and chapter 5 ('*The granular licensing of personal data'*). Next to being a form of DPbD with regard to the transparency requirement, granular licensing of personal data is also a design solution with regard to the requirements of strengthening the principle of purpose specification. Thus, *DataBait* is not only a transparency tool (which is the focus of this chapter), but also a 'strengthening of the purpose specification'-tool, a 'enhance a level playing field'-tool, etc. However, because the *main* achievement of the *DataBait* tool in terms of DPbD is to empower the data subject by providing her insight in her own raw data and by showing what can be derived from it (in other words, supporting profile transparency) and because almost each of the other requirements incorporated into the *DataBait* design in some way relates to the enhancement of transparency[20], we brand *DataBait* as a transparency tool with some additional design solutions to other legal requirements. In chapters 3, 4 and 5 we look at how these other legal requirements are translated into the *DataBait* design. However, in the following chapter we focus only on *DataBait* as a profile transparency tool.

---

of 'legal requirement' in a broad sense: namely as any requirement based in EU Data Protection law, even if it is not a requirement directed at us as data controller (compliance) but towards another data controller (empowerment of the user towards this other data controller).

[20] Two examples of how the other data protection requirements incorporated in the *DataBait* design relate to the enhancement of transparency:

(a) one of the reasons why the contract is an such an empowering legal ground is the fact that it provides transparency about the processing and,

(b) granular licensing does not only strengthen the purpose specification principle by offering an alternative to fuzzy and broad purposes but also by offering additional transparency about the purposes of the data processing.

These are merely two examples of the many that could be mentioned: as specified in section 1.2, *all* the DPbD solutions incorporated in the *DataBait* tool contribute in some way to transparency and/or mutual agreement between data subject and data controller.

2. Research strand 1: Supporting profile transparency

This first research strands looks at how *technological and organizational transparency* with regard to data processing in general, and in particular profiling, can be provided and supported 'by design', that is by using a profile transparency *tool*. We explore both how user *empowerment* could be supported through such a tool and how the tool itself should be made *compliant* with data protection law.

In section 2.1 ('*User empowerment and profile transparency*') we explore what 'profiling' means and how transparency with regard to this way of handling data can strengthen and support the informational rights a data subject has towards a data controller, and, consequently, result in increased user empowerment. Furthermore we explain how transparency with regard to profiling can be achieved 'by design' and that the DataBait tool developed in the USEMP project is a profile transparency tool which is a form of realizing Data Protection by Design (DPbD). The main achievement of the DataBait tool in terms of DPbD is to empower the data subject (supporting the exercise of informational rights towards OSNs, such as Facebook and Twitter[21], and browsers, such as Chrome and Firefox) by providing her insight in her own raw data (what information do I share? who is tracking me?) and by showing what information can be derived from it.

However, next to the research into how DataBait can be made into a strong profile transparency tool that supports the exercise of rights following from the data protection framework, making DataBait itself *compliant* with data protection law also provides a way of experimenting with optimal forms to comply with the requirements of (profile) transparency and fairness of data processing. In section 2.2 ('*DataBait: a profile transparency tool compliant with Data Protection law*') we provide a list of the various ways in which the DataBait tool aims to comply with data protection law in the best possible way.

2.1. User empowerment and profile transparency

The USEMP project aims to develop tools that enable users of social networks and browsers to control their digital trail and to understand how their data are used by the providers of social networks and browsers and by third parties piggy-backing on these systems. One of the underlying assumptions of the USEMP project (or, as it will be known to the end-user: the *DataBait tools*) is that *knowledge is power.* The idea is that if light can be shed on the world of tracking and personalized advertising, this will result not only in a better informed user, but also in a more *empowered* user. Currently, the world of tracking and targeting is invisible to the ordinary internet and social network user and its opacity makes it impossible to answer basic questions such as: *What information can be inferred from my data? What is the economic value of my data? What kinds of measures can be taken based*

[21] The focus of DataBait is currently on Facebook and, possibly, Twitter. However, DataBait could probably be adapted to other OSNs like Instagram, LinkedIn and Google+. Because each OSN has its own particularities the specific adaptations which would be needed will depend on the OSN.

on my volunteered, observed and inferred data? Who tracks me? Which commercial actors have access to my data? Providing relevant knowledge to the user, the USEMP project also assumes that this knowledge is all the more powerful when it is not presented in a generalized way, but rather as knowledge about the particular data trail of a concrete user. Instead of knowledge about an abstract average user, the USEMP (aka *DataBait*) tool aims to provide personalized insights with regard to each individual user of these tools.

2.1.1. Profiling and the right to profile transparency

'User empowerment' and 'personalized knowledge' about the commercial impact and technological possibilities based one's digital trail are not legal terms as such. In order to know how USEMP relates to the law, we have to translate these terms in legal terminology. Some legal notions, derived from the field of EU data protection law, bear a very obvious relation to the *knowledge is power*-assumption underlying USEMP: the obligation of the data controller to inform a data subject about certain aspects of data processing (Arts. 10 and 11 of the *Data Protection Directive 95/46/EC* [DPD 95/46]) and the data subject's right to access data (Art.12 DPD 95/46) and to object to the processing of them (Art. 14 DPD 95/46). Another legal notion which is relevant to USEMP is *profiling*, that is, a specific kind of data processing which can be described as:

> "'profiling' means any form of automated processing of personal data intended to evaluate certain personal aspects relating to a natural person or to analyse or predict in particular that natural person's performance at work, economic situation, location, health, personal preferences, reliability or behaviour." (Art. 4-3a of the proposed *General Data Protection Regulation* [pGDPR], the successor to DPD 95/46)

The knowledge that the USEMP tools aim to provide is largely about this specific form of data processing: the tools do not only inform the users about which trackers track which of their data ("simple" data processing), but also about which evaluative knowledge could be derived from these data ("profiling"). While the term *profiling* as such is not present in the DPD 95/46, the Directive does contain a specific provision of what can be called *the right to profile transparency*. This right *to obtain knowledge of the logic involved in any automatic processing which significantly affects the data subject* can be derived from Article 15(1) in conjunction with Article 12(a) of the DPD 95/46:

Article 15 DPD 95/46

Automated individual decisions

1. Member States shall grant the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.

2. Subject to the other Articles of this Directive, Member States shall provide that a person may be subjected to a decision of the kind referred to in paragraph 1 if that decision:

(a) is taken in the course of the entering into or performance of a contract, provided the request for the entering into or the performance of the contract, lodged by the data

subject, has been satisfied or that there are suitable measures to safeguard his legitimate interests, such as arrangements allowing him to put his point of view, or

(b) is authorized by a law which also lays down measures to safeguard the data subject's legitimate interests.

Art. 12(a) DPD 95/46

Right of access

Member States shall guarantee every data subject the right to obtain from the controller […] knowledge of the logic involved in any automatic processing of data concerning him at least in the case of the automated decisions referred to in Article 15(1)

With the fast technological development and commercial diffusion of profiling techniques, *profiling* has developed into a proper legal notion gaining lots of attention. For example, the Council of Europe (Council of Europe 2010) and Working Party 29 (Article 29 Data Protection Working Party 29 2013, 13 May) have devoted quite some attention to the legal definition of *profiling* and to how *the right to profile transparency* could be further developed, and the term is abundantly present in the pGPDR (see Annex 7.1).

2.1.2. The right to profile transparency in the proposed GDPR

While it is undeniably true that a basic version of *the right to profile transparency* is already present in the current DPD 95/46, it will be presented more explicitly and in a stronger and more elaborate way in future data protection legislation enshrined in the pGPDR. This becomes particularly clear in Arts. 14ga and 14gb (regarding the information which has to be provided to the data subject when profiling takes place) and Art. 20 (fully devoted to profiling and providing when it is allowed and when it is not) of the pGDPR. Thus when comparing Art. 14ga of the pGDPR to the provisions in the current DPR 95/46, it is clear that the pGPDR is much more specific about the kind of information that has to be provided: the data controller shall provide the data subject with information about the *existence* of profiling, of *measures based on* profiling, and the *envisaged effects* of profiling on the data subject.

It is interesting that the focus is here not just on the profiling as such but also on what the profiling actually *does* in practice. Objecting to profiling should not just be a theoretical possibility but a right that can actually be used ("*The data subject shall be informed about the right to object to profiling in a highly visible manner*", Art. 20(1) pGDPR). Moreover, contrary to the current DPD 95/46, which prohibits profiling which has significant or legal effects and is based *solely* on the automated processing of data[22], the proposed GPDR also prohibits such profiling if it solely *or predominantly* relies on automated processing (Art. 20(5) pGPDR).

[22] "*… the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him*", Art. 15(1) DPR 95/46.

Furthermore, such profiling "shall include human assessment, including an explanation of the decision reached after such an assessment" (Art. 20(5) pGPDR).

Another striking difference is that the original Commission version and the Parliament version of the pGDPR[23] (Art. 20(3) pGDPR) prohibit profiling which is *solely* based on sensitive data[24] (that is, data revealing race or ethnic origin, political opinions, religion or philosophical beliefs, sexual orientation or gender identity, trade-union membership and activities, and the processing of genetic or biometric data or data concerning health or sex life, administrative sanctions, judgments, criminal or suspected offences, convictions or related security measures). These sensitive data are not wholly excluded from being used as *input* in a profiling process, but they should always be combined with other, non-sensitive, data. The Council version of the pGDPR does not prohibit the use of sensitive data as input, but does concur with the Commission and Parliament version of the pGDPR that profiling with discriminatory *effects* with regard to race or ethnic origin, political opinions, religion or beliefs, trade union membership, sexual orientation or gender identity should be prohibited and demands of the controller that effective protection against possible discrimination resulting from profiling should be in place.

Thus, to summarize: in all versions of the pGDPR both the output (the *effects*) of profiling should be scrutinized in order to prevent discrimination based on a set of protected grounds, and in some versions the *input* of profiling is to be scrutinized as well. Next to the prohibition of discriminatory profiling, the pGDPR also prohibits profiling in the field of employment (Art. 82(1) pGDPR). Thus, overall the pGPDR will offer a better protection against unwarranted forms of profiling and gives the right to profile transparency more teeth.

However, the pGPDR also introduces some provisions which could make the protection against unwarranted profiling somewhat weaker. Under the present Directive, profiling "which leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject" is only allowed if there is a law authorizing such processing or if such profiling is necessary for entering or performing a contract lodged by the data subject (Art. 15(2a) DPD 95/46). In the pGDPR such profiling could also be allowed if it is based on the consent of the data subject. Another addition in the pGPDR which could weaken the protection against profiling is the presumption (Recital 58(a) pGDPR) that profiling based solely on the processing of pseudonymous data (i.e., personal data that cannot be attributed to a specific data subject without the use of additional information) will not significantly affect the interests, rights or freedoms of the data subject. However, this presumption is highly contested and it will be interesting to see what the Council will do with it in the upcoming step of the legislative process.

Notwithstanding the differences between the current and future *right to profile transparency*, the main *rationale* for both of them is very similar to the assumption of user

[23] See for a comparison between the pGDPR versions of the Commission, the Parliament and the Council: <https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-07-27_GDPR_Recommendations_Annex_EN.pdf > (accessed 7 September 2015)

[24] Establishing whether a piece of data qualifies as sensitive data can sometimes be tricky in practice. Do the daily amount of steps taken by a user, collected by a step counting app, qualify as health data? Does a picture of a user qualify as biometric data because such data can be extracted from it by certain analytic software? We discuss these issues in detail in chapter 3.

empowerment underlying USEMP: profile transparency aims at preventing a data subject from being confronted with a “Computer says no” in a situation that significantly affects his or her interests.[25]

2.1.3. Profile transparency: what *DataBait* can and cannot do

As soon as law is involved, the devil is in the details: profile transparency is not something that a data subject can *always* appeal to. The law is more subtle than the straightforward adage that a user, to whom a profile is applied, can always request *full transparency*. Law is a practice of nuance. Even if we follow the rather straightforward formulation in DPD 95/46, the question *if* the right to profile transparency applies and *how to comply* with it, requires that we look into a set of specifics such as, for example:

- *Can the* “*profiling*” *at stake indeed be qualified as the action described in Art. 15* (“automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc. ”)*?*
- Is the provision of insight into the profile in accordance with what is required by Art. 12 (“knowledge of the logic involved in any automatic processing of data”)?
- Does the automated processing of data which evaluates the data subject in some respect result in a decision that produces legal effects or significantly affects the data subject?
- Is this decision solely based on the automated processing of data or is it based on a combination of human and automated decision making?
- Is there a legal ground legitimizing the profiling?
- Does the right to profile transparency adversely affect trade secrets or intellectual property rights of other actors (Recital 41 DPD 95/46)?

Only by looking at both the legal details and those of the technological architecture, it becomes possible to answer the question if a particular tool, system or practice is compatible with the right to profile transparency. This is even more so with the elaborate version of the right to profile transparency in the proposed GPDR, where even more aspects have to be considered, such as, for example:

[25] It should be noted that the commercial profiling applications studied in the USEMP project seem to be mainly steered by the interest of nudging a consumer into a particular commercial transaction (*Computer says : “Please, do.. ”*) and do not primarily aim to take decisions which are contrary to the user’s will (*Computer says no*). However, the line between nudging positively (“*Please, do..* ”), nudging negatively (“*Please, don’t..* ”) and denial of service (“*No!*”) is often thin and fluid. For example, think of an insurance company nudging a certain type of users to become their customers with specific discounts (positive nudging through price differentiation). One could say that the flipside of this positive nudge is that this company gives a negative nudge to potential customers who do not fit the profile.

- Does the profiling process result in discrimination against individuals on the basis of race or ethnic origin, political opinions, religion or beliefs, trade union membership, sexual orientation or gender identity?
- Is the profiling based solely on the processing of pseudonymous data?
- Can the data controller nevertheless attribute these pseudonymous data to a specific data subject?
- Is the profiling used in the context of employment?

With regard to the profiling performed by OSNs, browsers or third parties, the *DataBait* tool cannot answer any of these questions in a conclusive way but it does support the user by providing her insight in her own raw data (what information do I share? who is tracking me?), by showing what additional information can be derived[26] through the use of profiling techniques, what value her profile can be, and by informing her of her EU data protection rights. In this way, users of the DataBait tool could become more critical of the information they post (because they understand what can be derived from it) and have a better understanding of their informational rights (i.e. which information they can receive about the processing of their personal data from an OSN or browser). Thus, even though the use of DataBait does not give conclusive answers about the kind of profiling a data subject is subjected to, on which data it is based and for which purposes this derived knowledge is used, it does sensitize the data subject through the provision of (a) *factual knowledge about her digital trail* (by showing who is tracking her and which raw data are available), (b) *insight into what current analytic software can extract from raw data* (by showing which knowledge can be derived from the raw data), (c) *a sense of the potential (commercial) value of the digital trail of the data subject and her 'audiences', that is, her Facebook friends or Twitter followers* (by showing for what purposes her data and audience could be used[27]) , and (d) *awareness of her informational rights based in EU data protection law* (by informing her about her rights in relation to the other insights provided to her). Given this knowledge, a data subject can take control either by exercising her informational rights towards those profiling her (that is, by raising questions and concerns which are pertinent in relation to the rights following from EU Data Protection law) or by readjusting the information she shares (e.g., by blocking trackers and removing certain pieces of information from her OSN profile).

2.1.4. "Meaningful information" about the logic of profiling

[26] It should be noted that this 'derived' or 'inferred' information is not necessarily correct – it is better described as an informed guess. For example, a person who uses many negative words in her Facebook updates ('sad', 'anxious', 'dark', 'hate', etc.) is not necessarily depressed, but could be profiled as a depressed person based on a textual analysis of her posts.

[27] For example, if the DataBait tool has inferred that a user has a lot of pictures and posts related to classical music and that these posts provoke lots of comments among the Facebook friends of this user, the DataBait tool might suggest that the 'audience' of this user, and her influence on it, could make her commercially interesting for companies producing recordings of classic music (the user could promote records to her audience and get something in return: freebies or a financial remuneration). See D6.2 for an explanation how the notion of 'audience' is modeled in the architecture of the DataBait tool.

Providing the *DataBait* user with feedback about her digital trail has also been a way of giving flesh to the legal requirement of "meaningful information about the logic of any automated processing" (Art. 14gb of the pGDPR). Here the legal work interfaces with the ongoing work in Task 6.3 ("*Visualisation of and Interaction with user empowerment data*") on good user interfaces that display information in such a way that is does not become too complex or overwhelming. From a legal perspective (see e.g. Wauters, Lievens et al. 2014, p. 292), it is important to study if such (simplified) visualizations still offer enough detail to qualify as "meaningful information about the logic of any automated processing" (Art. 14gb GDPR; see also Art. 12a of the DPD 95/46). There is a fine line (Asgharpour, Liu, & Camp, 2007; Camp, 2006) between (a) ease of user interface and intuitiveness of the representation, (b) too much simplification. As an aside it should be noted that this point is a good example of the fact that a legal compatibility assessment is not always a one-way street where a technological or organizational architecture is simply checked against a set of legal requirements. Because legal terms (e.g., "knowledge of the logic involved in any automatic processing of data", Art. 12 DPD 95/46) do not always have an exhaustive definition, the design solutions in the USEMP project might actually be an inspiration to the lawyer. Combining legal requirements of profile transparency and the technological and social requirements has been an interactive process which has resulted in four categories of information in the main screen of the *DataBait* graphic user interface (figure 1)*:* users are provided with insight in who tracks them ('*User Trackers'*), the raw and inferred data that their digital trail contains or might contain ('*My Privacy*'[28]), to what kind of actors their influence on their specific set of Facebook friends might be of interest ('*Audience Influence*'[29]), and the possible commercial value[30] of their profile ('*Monetization Insights*[31]*'*). Each of these informational options supports users in exercising their informational rights towards OSNs and browsers.

[28] The heading "*My Privacy*" was used in the preliminary version of August 2015 (used for pre-pilot experiments). However, in following versions this heading will be changed – probably to something like "*What do I disclose?*", because this conveys the content of the provided information better. After all, for lawyers thinking within the framework of European fundamental rights, 'privacy' is a notion whose meaning can be found in art. 8 of the European Convention of Human Rights (right to respect for private life) and that has a very specific meaning (mainly a right that prevents power imbalances between state and citizen or citizen and other actors). We want to avoid the false impression that a high DataBait disclosure score (which indicates that the user discloses sensitive content, over which she has little direct control and/or that this content is visible to a large audience; see D6.2) would imply an infringement of Art. 8 ECHR.

[29] See D6.2 for the definition of 'Audience Influence' as used in the DataBait architecture.

[30] Commercial value is defined in a broad sense (see D3.5 on the socio-economic value of personal data) because putting an actual price to personal data is not unambiguous, as there are many ways to model economic or monetary value. (See for research on the actual price of data trails: Olejnik, Minh-Dung et al. 2014).

[31] The heading "*Monetization insights*" was used in the preliminary version of August 2015 (used for pre-pilot experiments). However, in following versions this heading might be changed – probably to something like "*Insights in my commercial value"*, because money is not the only way in which commercial or economic value can be expressed. See D3.5 on the socio-economic value of personal data.

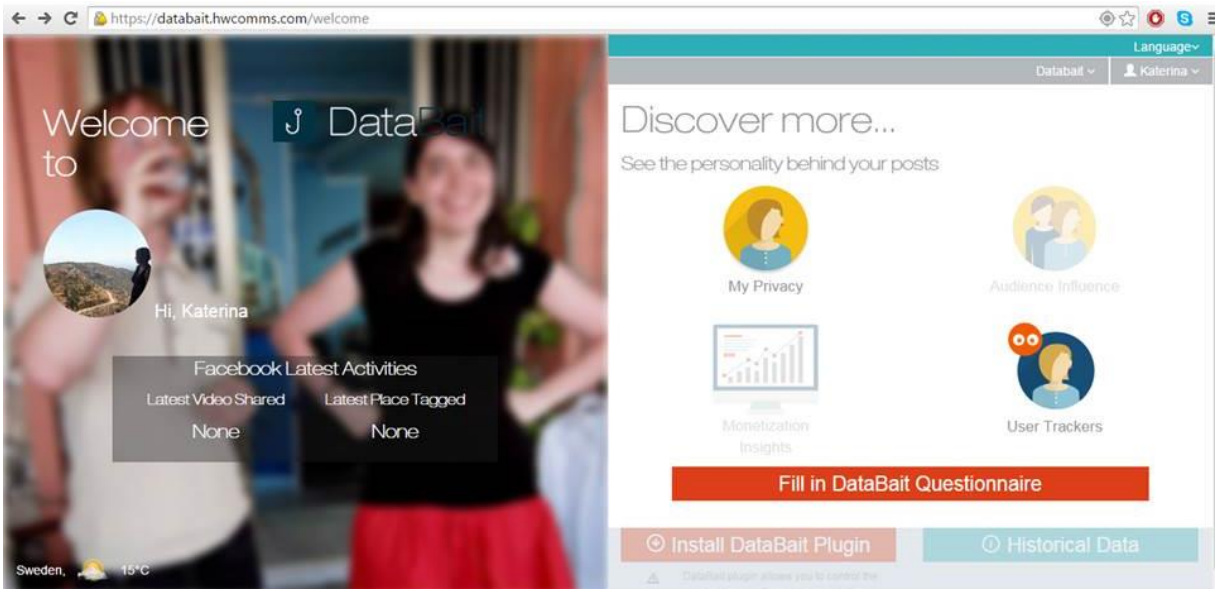


Figure 1. The main screen of the DataBait graphic user interface (preliminary version of August 2015 used for pre-pilot experiments).

2.2. DataBait: fair and lawful profile transparency

Creating the *DataBait* tool is not only an experiment in creating a profile transparency tool (that is, a form of *DPbD*) supporting the exercise of rights following from the data protection framework (that is, *user empowerment*). It is also a way of experimenting with optimal forms to *comply* with the requirements of (profile) transparency and fairness of data processing when creating a transparency tool. In this section we list the various ways in which the DataBait tool aims to comply with data protection law in the best possible way.

To begin with, *DataBait* users are expected to sign a contract before using this profile transparency tool. This contract, the so-called *Data Licensing Agreement*, is the legal ground for any data processing performed by the USEMP in relation to the data gathered by the *DataBait* tool. The contract supports transparency by avoiding any unnecessary 'legalese' and specifying the purposes and the process of the data processing in a very comprehensive way. In comparison to data processing based on user consent, using contract as a legal ground for the processing creates a more equal footing between the data subject and the data processor. Chapter 4 of this deliverable is devoted to explaining this way of DPbD through the contract.

Another way in which compliance with Data Protection law is realized in the current preliminary version of *DataBait* is the clear user interface where one can request data deletion and withdraw consent for the processing of one's sensitive data (see figure 2).

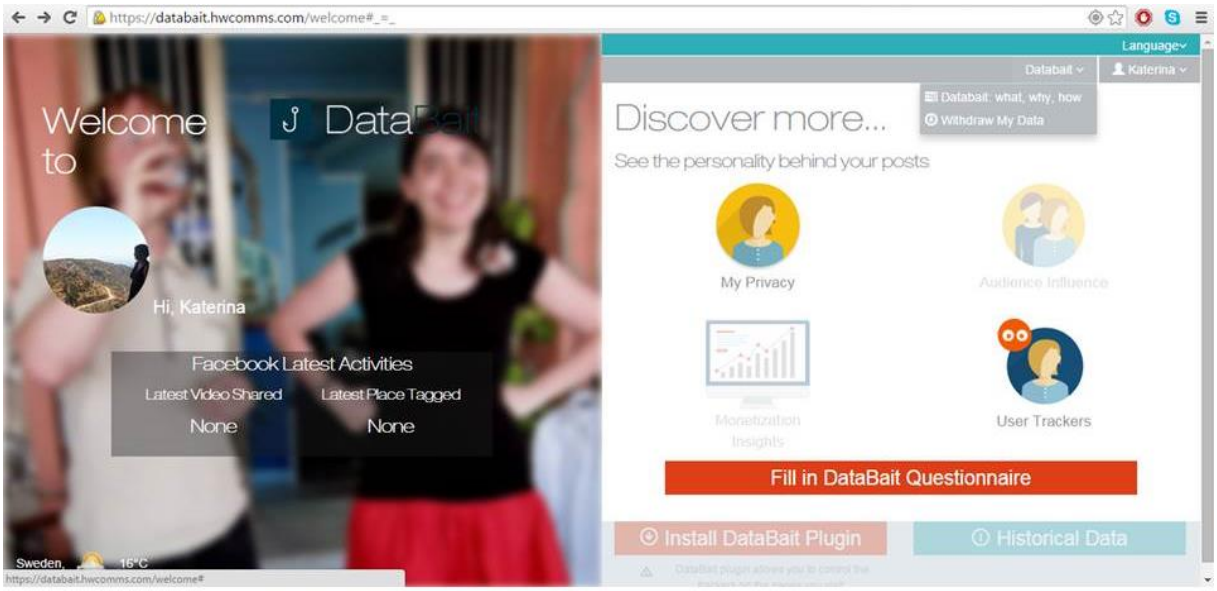


Figure 2. The 'Withdraw My Data'-button in the preliminary version of DataBait (August 2015) can be found in the upper right corner.

While all the information that the data controller is required to provide to the data subject is in principle included in the contract (the 'Data Licensing Agreement' or DLA) signed by each user of the *DataBait* tool, the USEMP consortium also provides additional information to further enhance the transparency and fairness of the processing in the '*DataBait: how, what and why?*'-section (see figure 3). When drafting this section we made a list of items that should be included in this section:

(1) The contract signed by each user (the 'Data Licensing Agreement', abbreviated as DLA) and the contract which all USEMP partners have signed amongst each other with regard to the processing of personal data (the 'Personal Data Processing Agreement', abbreviated as PDPA) which also incorporates the DLA and binds each USEMP partner.

(2) Some very practical info, including the identity of the data controllers in the USEMP project, an email address for each USEMP partner that processes personal data (to make further inquiries), information about the existence of the right to request rectification or erasure of the data concerning the data subject and of the right to object to the processing, and the right to lodge a complaint to the supervisory authority and the contact details of the supervisory authority.

(3) Text, graphs and/or an animation providing insight in the general logic of the processing and a precise description of operations performed. This includes references to the databases used for training and testing the USEMP analytic algorithms (used to derive additional information from the raw data of the user) and a list with precise descriptions of what categories/types of data are collected/processed by each USEMP partner and how they are processed (this list is also presented in clause E of the PDPA).

(4) Text, graphs and/or an animation clarifying the purpose for which the data are processed, for what estimated period, which recipients receive the data (no data sharing to third parties!), and what might be the consequences of such processing.

(5) Provision of which data are actually processed [corresponding to the data listed in D3.4 and D3.9].

These five requirements have been translated into four tabs in the '*DataBait: how, what and why?*'-section (see figures 3 till 6), entitled '*DataBait at a Glance*' (containing flow-charts, explanatory text, and an explanatory video-animation giving an overview of the processing actions performed on the data gathered by the *DataBait* tool, the purpose of the processing and the relevant responsible actors), '*Which of your personal data do we process?*' (containing lists of collected data types), '*Practical info'* (containing all practical contact details for the exercise of the right of access and rectification) and '*DataBait Contract: Terms of Service'* (containing the DLA and PDPA).

In later versions of *DataBait* we consider adding (a) a button which allows users to download all their data in the '*Which of your personal data do we process?*'-section, and (b) complementary information with regard to data protection rights EU citizens have with regard to, for example, OSNs and browsers processing their personal data in an additional tab entitled '*Your digital data protection rights*'. This section will also explainthat the legal regimes on data protection in EU is quite exceptional: as indicated in the *Schrems* case of the Court of Justice of the European Union, the protection of personal data against operations that would be unlawful in Europe but not in the US raises a number of issues about the safe harbor decision of the European Commission. This also relates to the fact that such data can be accessed by the US government without effective judicial redress for EU citizens. This implies that USEMP should take measures to prevent any of its partners from processing personal data in the US. An update of the DLA and the PDPA to confirm this will be made in the final year of the USEMP project. So far, all personal data processed by USEMP partners have been processed and stored within the EU."

The content and the format of the information in the '*DataBait: how, what and why?*'-section is not radically new in itself: it is mostly information which each data controller bound by EU law has to provide (explanation of the processing through text, graphs, and animations; lists of processed data types; the possibility to download your data; contact info; information about legal rights; terms of service and data policy). Yet, while many data controllers are bound by a double bind, namely the need to comply with data protection law as well as the need not to scare users away with too much insight in what is happening with user data, the USEMP project does not have such an ambiguous position. For example, when a user signs up for *DataBait*, the user cannot simply click 'consent' to the DLA contract as a whole, but has to click through every clause (twelve screens). A commercial data controller would probably refrain from choosing such a format, being afraid that the sign-up procedure will become too tedious and gives the user 'too much' insight in the data processing. However, the USEMP project is an excellent occasion to check and experiment with how users react to such format – do they indeed dislike it because it takes longer? Do they become better informed? Etc.-

Another aspect which makes USEMP stand out is the layered way in we provide information to the *DataBait* users. The explanatory text contains many hyperlinks where the interested user is able to find more detailed information. For example, the text explaining how the USEMP-DataBait algorithms were construed contains links to the databases used to train and test the algorithms. A particular form of the layered approach (which is used in Creative Commons licenses relating to copyright, namely with three layers, containing a human

readable or ‘common sense’ explanation, a legal explanation and a machine readable explanation) is further elaborated in chapter 5 with regard to so-called granular licenses. With respect to the human readable or ‘common sense’ layer we have to find a middle ground between oversimplification and buffer-overload by providing the user with too much technical details. As discussed in section 2.1, there is a fine line between making a complex situation understandable and oversimplification. Moreover, there is no ‘free lunch’ in communicating security or privacy risks in a heuristic way (Asgharpour et al., 2007; Camp, 2006): the "cost" of a heuristic presentation is that it will always contain a bias in some direction. Finding optimal ways of making *DataBait* compliant with EU Data Protection law is in this sense also an opportunity to find the best possible mental models in communicating the risks and implications related to data processing, and particularly profiling. The wish to use a layered approach and to choose optimal explanatory heuristics also has guided us in the way we have translated the legal requirement of transparency into the DLA contract (chapter 4) and the granular licenses (chapter 5).

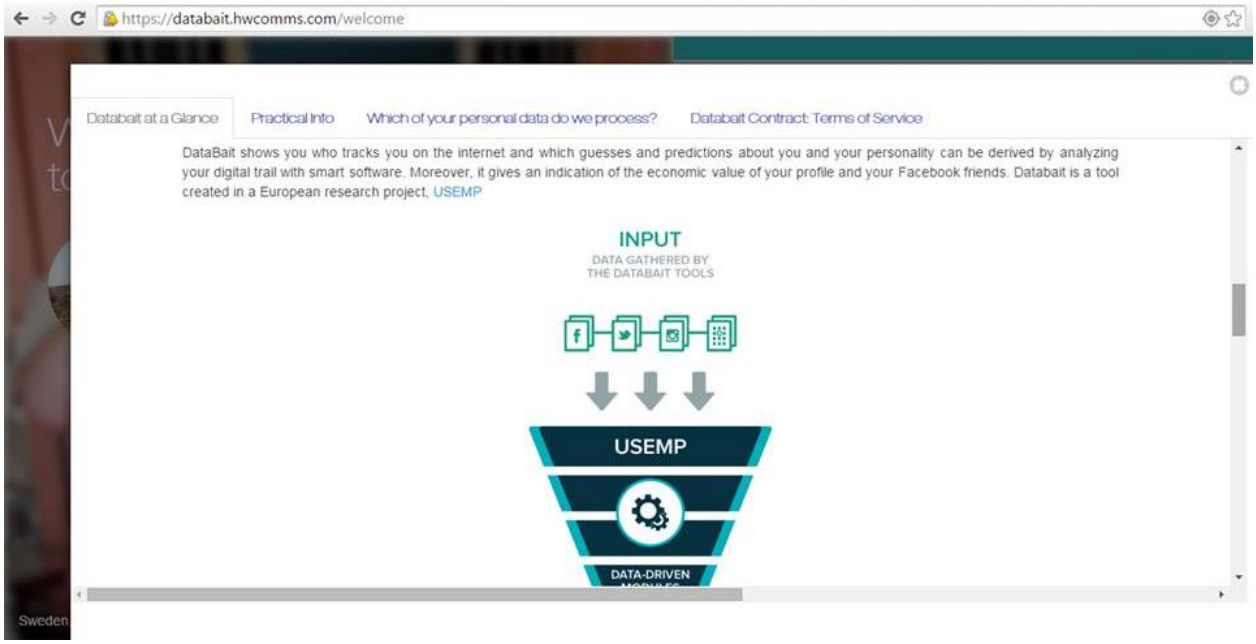


Figure 3. The ‘Databait: what, why, how’-section in the preliminary version of DataBait (August 2015) contains four subsections: the first one is ‘DataBait at a Glance’.

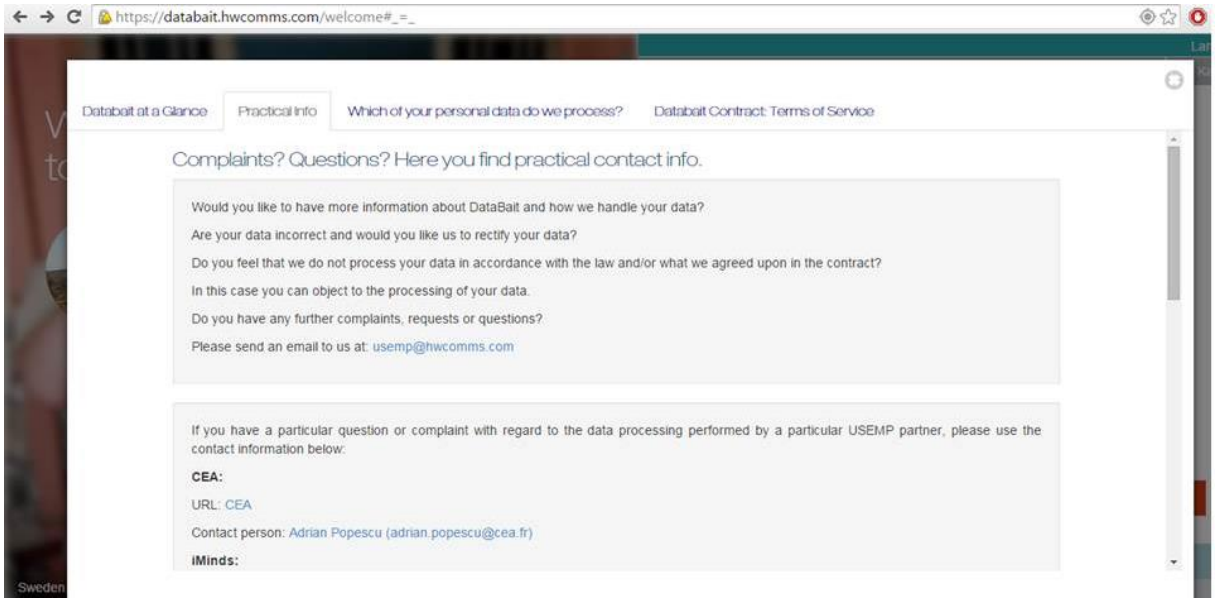


Figure 4. The 'Databait: what, why, how'-section in the preliminary version of DataBait (August 2015) contains four subsections: the second one is 'Practical info'.

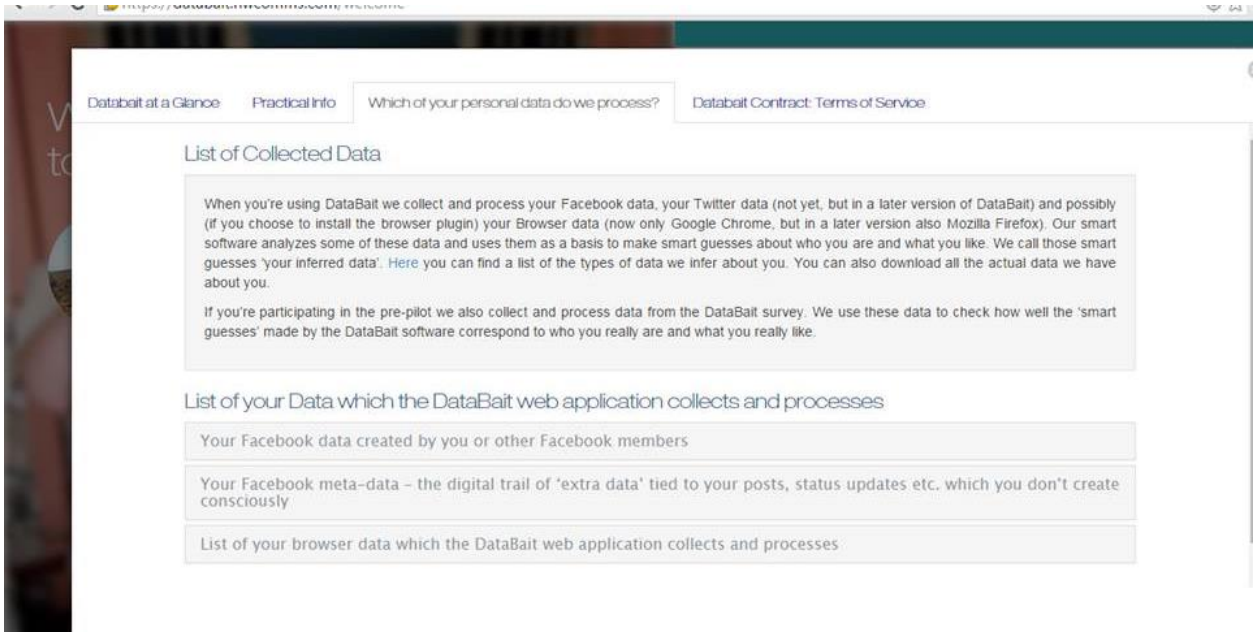


Figure 5. The 'Databait: what, why, how'-section in the preliminary version of DataBait (August 2015) contains four subsections: the third one is 'Which of your personal data do we process?'

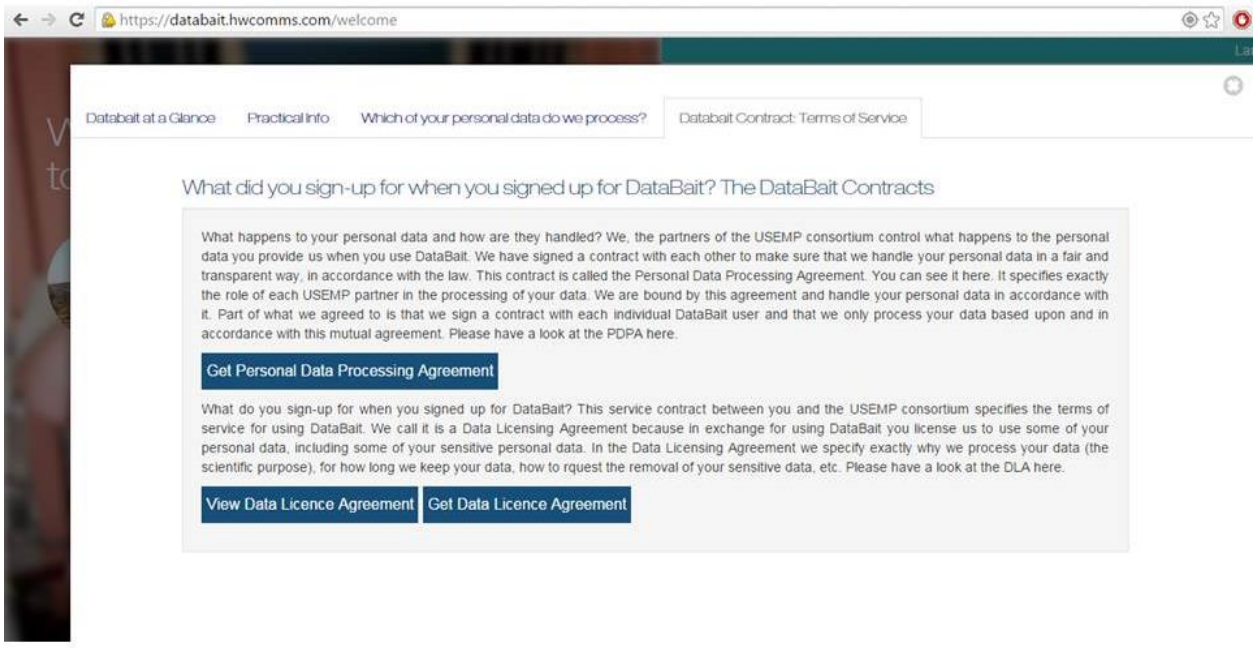


Figure 6. The 'Databait: what, why, how'-section in the preliminary version of DataBait (August 2015) contains four subsections: the last one is 'DataBait Contract: Terms of Service'.

3. Research strand 2: 'Sensitive personal data' and 'anonymisation'

The second strand of research in this deliverable gives *a legal clarification* of which data should be considered *'sensitive'* in the sense of Art. 8 DPD 95/46[32] (or Art. 9 pGPDR[33]), and which data can be considered *anonymous* (i.e., not personal data and therefore outside the scope of DPD 95/46) and shows how our clarification of these two contentious legal notions would translate into DPbD requirements.

What makes both notions ambiguous in their practical application is that they both contain an element of *possibility* or, almost Aristotelian, *potentiality*. In the same sense as Aristotle would say that an acorn should be understood as a potential oak and the boy as a potential man –they just have to realize their potential nature – some data could be classified as sensitive or personal, not because this is what they are now, but because of their potential

[32] Art. 8 (1) of DPD 95/46: "Member States shall prohibit the processing of 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."
[33] Art. 9 (1) of the pGDPR: "The processing of personal data, revealing: race or ethnic origin, political opinions, religion or beliefs, trade-union membership, and the processing of genetic data or data concerning health or sex life or criminal convictions or related security measures shall be prohibited."

of becoming such. Thus, in order to decide whether a piece of data should be qualified as *sensitive* personal data (and thus be treated with additional care, compared to 'ordinary' personal data) or whether a piece of data should be qualified as *anonymous* (and thus not be handled in accordance with data protection requirements which only relate to personal data) it is not enough to look at them at their face value. In order to assess whether a piece of data is personal or not, one has to assess whether there is a potential for a piece of anonymous data to be de-anonymized (turning it into personal data).

3.1. Anonymous data?

In its Opinion on anonymization techniques (Article 29 Data Protection Working Party, 2014) Working Party 29 describes randomization and generalization as the anonymisation techniques which are applied most widely. The Opinion assesses the robustness of each technique based on three criteria:

(i) Is it still possible to single out an individual ? Singling out is the possibility to isolate some or all records which identify an individual in the dataset.

(ii) Is it still possible to link together multiple records related to an individual ? Data are considered 'linkable' when it is possible to link, at least, two records concerning the same data subject or a group of data subjects (either in the same database or in two different databases).

(iii) Can information be inferred concerning an individual? Inference is the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attributes.

Thus in assessing whether data are anonymous, it is not enough that data are not directly related to an identified or identifiable person, because the possibility of singling out, re-linking and inferring information should be taken in to account. We follow the opinion of Working Party 29 (Article 29 Data Protection Working Party, 2014), and conclude that only data which, taking into account *all the means likely reasonably to be used*[34], cannot be de-anonymized can be qualified as anonymous. The standard of '*all the means likely reasonably to be used'* is a very high one (Article 29 Data Protection Working Party, 2014), meaning that data which have any reasonable potential of being de-anonymized should not be considered anonymous but as pseudonymous[35] personal data, that is, as data with a "potential identifiability" (p. 8) and as such fall within the scope of Data Protection law.

> "Data controllers often assume that removing or replacing one or more attributes is enough to make the dataset anonymous. Many examples have shown that this is not the case; simply altering the ID does not prevent someone from identifying a data subject if quasi-identifiers remain in the dataset, or if the values of other attributes are still capable of identifying an individual. In many cases it can be as easy to identify an

[34]Recital 23 of DPD 95/46 states: "..to determine whether a person is identifiable, account should be taken of *all the means likely reasonably to be used either by the controller or by any other person* to identify the said person" (*italics ours*) See similarly Recital 26 of the pGDPR.

[35] "'pseudonymous data' means 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) pGPDR.

individual in a pseudonymised dataset as with the original data. Extra steps should be taken in order to consider the dataset as anonymised, including removing and generalising attributes or deleting the original data or at least bringing them to a highly aggregated level" (Article 29 Data Protection Working Party, 2014, p. 21)

Full anonymization will often be not easy, because the technological possibilities for de-anonymization abound and are constantly increasing. The Working Party stresses that no anonymization technique is devoid of shortcomings per se. Thus, the rule of thumb in assessing whether data are anonymous is straightforward: *any* reasonable[36] potential for de-anonymization disqualifies data from being labelled as anonymous and brings them within the scope of data protection law[37]. The difficulties in the qualifying whether data is anonymous do not lie in the qualification rule as such (*every* potential for de-anonymization disqualifies the qualification of anonymity), but in the almost impossible task of having an overview of all the technological possibilities for de-anonymization and of all other existing data-sets which could be combined in such ways that de-anonymization becomes possible[38].

The analysis of Working Party 29 leads to the conclusion that many of the widely used techniques actually result in pseudonymization[39] instead of anonymization because some re-identification might still be possible. Real anonymization is only possible if "the prerequisites (context) and the objective(s) of the anonymisation process [are] clearly set out in order to achieve the targeted anonymisation […]. The optimal solution should be decided on a case-by-case basis, possibly by using a combination of different techniques, […].

Finally, data controllers should consider that an anonymised dataset can still present residual risks to data subjects. Indeed, on the one hand, anonymisation and re-identification are active fields of research and new discoveries are regularly published, and on the other hand even anonymised data, like statistics, may be used to enrich existing profiles of individuals, thus creating new data protection issues. Thus, anonymisation should not be regarded as a one-off exercise and the attending risks should be reassessed regularly by data controllers."

Currently the USEMP project data does not envision anonymizing the data at the end of the project, but simply deleting them (see details in table 1) at the end of the project. By deleting the data, difficult questions about the potential for reidentification can be avoided. However, if it turns out that the value of the data gathered in this project makes it important to preserve the data in anonymized form, all partners will collaborate to endure optimal anonymisation of the data. This is shown in table 1 hereunder. The deletion and/or anonymisation of the USEMP data will be performed in accordance with the implications

[36]In assessing what is to be a reasonably likely potentiality, "[i]mportance should be attached to contextual elements: account must be taken of "all" the means "likely reasonably" to be used for identification by the controller and third parties, paying special attention to what has lately become, in the current state of technology, "likely reasonably" (given the increase in computational power and tools available)."" (Article 29 Data Protection Working Party, 2014, p. 6) For an empirical example of how seemingly anonymized data have been de-anonymized, see for example work on Netflix: (Narayanan & Shmatikov, 2008).

[37] The extent to which pseudonymous person data should be protected with a 'lighter' regime of data protection than 'ordinary' (non-pseudonymous) personal data, is still a topic that is fiercely debated.

[38] See also: (Ohm, 2010).

[39] 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)

following from the *Rijkeboer* decision of the EU Court of Justice[40], that is, deletion or anonymization of data should not interfere[41] with the data subject's right to access of her data following from Art. 12(a) DPD 95/46. The implications of the *Rijkeboer* decision[42] for the anonymisation and deletion of the USEMP data will be further explored in the final version of this deliverable (D3.10).One important point to keep in mind is that when a consortium like USEMP shares any anonymised data with other (scientific) interested parties, these data cannot be considered anonymised if the consortium still has the identifying information. This means that any sharing of anonymized data will not be possible during the project (while USEMP still has identifying information) but only after the project has ended and the consortium itself has deleted all identifying information[43] and has only has preserved anonymised data.

Personal data processed in the USEMP project, ordered according to source:	**Premise**	**Deletion/anonymization/pseudonymization? (if, when)**
A. Personal data collected with the DataBait OSN app	*All personal data processed in the USEMP project are stored at HWC; next to that some small, pseudonymized subsets are stored at CERTH, VELTI and iMinds.*	*All personal data processed in the USEMP project will be either deleted at the end of the project or, if it turns out that there is a considerate research value in preserving them, properly anonymized. During the project no anonymization/pseudonymization techniques are applied, apart from the*
B. Personal data collected with the DataBait		

[40] CJEU, C-553/07, College van burgemeester en wethouders van Rotterdam v. M.E.E. Rijkeboer, 7 May 2009. Online available at: http://curia.europa.eu/juris/document/document.jsf;jsessionid=9ea7

[41] Working Party 29 writes in this regard: "It should also be emphasized that anonymisation has to be held in compliance with the legal constraints recalled by the European Court of Justice in its decision on case C-553/07 (College van burgemeester en wethouders van Rotterdam v M.E.E. Rijkeboer), relating to the need to retain the data in an identifiable format to enable, for instance, the exercise of access rights by data subjects. The ECJ ruled that "*Article 12(a) of the [95/46] Directive requires Member States to ensure a right of access to information on the recipients or categories of recipient of personal data and on the content of the data disclosed not only in respect of the present but also in respect of the past. It is for Member States to fix a time-limit for storage of that information and to provide for access to that information which constitutes a fair balance between, on the one hand, the interest of the data subject in protecting his privacy, in particular by way of his rights to object and to bring legal proceedings and, on the other, the burden which the obligation to store that information represents for the controller*." (Article 29 Data Protection Working Party, 2014, p. 8)

[42] See for a recent application of the *Rijkeboer* decision by the Dutch court of The Hague (1 September 2015, case number 200.162.134/01, online available at: http://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:GHDHA:2015:2332), in which telecom provider KPN was held liable for infringing on the right to access and had to pay damages for deleting a history of text messages of one of their customers, as these data constituted crucial evidence which would probably have allowed this customer to win a court case which he lost due to the lack of these data. From April 2010 onwards the customer had requested KPN access to these data on several occasions. The most recent request of the customer was dated 10 April 2014. However, KPN did not respond to any of these requests and deleted the data in May/June 2014. See for a commentary (in Dutch) of this case: http://dirkzwagerieit.nl/2015/10/02/kpn-draait-op-voor-schade-wegens-onrechtmatig-verwijderen-persoonsgegevens/

[43] This will be checked with our technical partners and colleagues.

browser plugin C. Personal data collected in the DataBait surveys in the pre-pilot. D. Personal data *inferred* from a subset of the data collected through the OSN app [A] and the browser plugin [B] E. Personal data in training and testing sets, used to train and test classifiers	*Additional clarification:* 1. All data are stored at HWC. 2. HWC will, however, provide CERTH[44], VELTI[45] and iMinds[46] with a set[47] of pseudonymized[48] data from the pre-pre-pilot for temporary usage at their own premises. 3. CERTH and VELTI have remote access to backend servers for integration which gives indirect access to imagery data stored on the system, as well as indirect access to the social media stores. 4. CEA has remote access to the Image Processing server, which also allows indirect access to social media data and imagery data.	*pseudonymization of the subsets provided to CERTH and VELTI.* *Additional clarification:* 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, 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 from the pre-pre-pilot for temporary usage at

[44] Pseudonymized data from the pre-prepilot requested by CERTH are:
- User likes for all the users. Likes should not be hashed.
- Posts / status updates
- Extracted visual concepts and logos.
- List of friends of each user
- Survey responses

Data will be used for the development of the likes-based inference module and validation of the results that it produces

[45] Pseudonymized data from the pre-prepilot requested by Velti are:
- User likes for all the users.
- Survey responses

[46] Pseudonymized data from the pre-prepilot requested by iMinds are:
- Facebook Data.
- Survey responses

Data will be used to investigate if there exists contradictions between what people have claimed that is available online (survey) and what actually could be found.

[47] This data set from the pre-prepilot contains all the survey data ("surveyAnswers"), apart from any identifying information such as email, address, etc.contained in the section “Contact Information”, and the following Facebook data: "relationshipStatus", "religion", "website", "birthday", "timezone", "verified", "gender", "political", "locale", "updatedTime", "currency", "interestedIn", "meetingFor", "education","sports", "favoriteTeams", "favoriteAthletes", "languages", "birthdayAsDate", and "likes". The data set does not contain any images. In order to pseudonymize the Facebook and survey data the “id” has been *discarded and converted to a non-tracable guid*. Moreover, the following Facebook data have been *discarded*: "metadata", "type", "name", "firstName", "middleName", "lastName", "link", "bio", "quotes", "about", "email", "username", "picture", "hometown", "location", "significantOther", "thirdPartyId", "tokenForBusiness","work", and "hometownName".

[48] Facebook id, username, phone number and email numbers are hashed.

		their own premises. 4. CERTH and VELTI have remote access to backend servers for integration which gives indirect access to imagery data stored on the system, as well as indirect access to the social media stores. 5. CEA has remote access to the Image Processing server, which also allows indirect access to social media data and imagery data.

Table 1. Where the DataBait data are kept and if/when are they deleted/anonymized/pseudonimized?

3.2. Explicit consent for the processing of sensitive personal data

Some personal data can be defined as 'sensitive' or 'special' and need to be handled with extra care. We follow the list of sensitive categories of data mentioned in Art. 9 (1) of the pGDPR (which gives a slightly more extensive list than Art.8(1) DPD 95/46):

> "The processing of personal data, revealing: race or ethnic origin, political opinions, religion or beliefs, trade-union membership, and the processing of genetic data or data concerning health or sex life or criminal convictions or related security measures shall be prohibited."

The rationale why the processing of data from this 'special' category deserves extra caution is explained by Working Party 29:

> "In its advice paper from 2011 to the European Commission the Working Party has explained the rationale behind this stricter legal regime. It stems from the presumption that misuse of these data in general, is likely to have more severe consequences for the individual's fundamental rights, such as the right to privacy and non-discrimination, than misuse of other, "less sensitive" types of personal data." (Article 29 Data Protection Working Party, 2015a, p. 1)

The strict regime for the handling of sensitive personal data entails that processing of such data is prohibited - *unless* an exception applies. The most important exception is *consent*, which has to be given specifically for the processing of this sensitive data. Moreover, the consent for the processing of sensitive needs to be *explicit* (Art. 8.2(a), Directive 95/46 and Art. 9.2(a) GDPR). This is a higher standard than consent for other (i.e, non-sensitive) personal data (which, under the current DPD 95/46, could sometimes also be "inferred" or "implicit", i.e. that the actions of the data subject imply consent). Working Party 29 (Opinion 15/2011 on the definition of consent) clarifies:

> "In legal terms "explicit consent" is understood as having the same meaning as express consent. It encompasses all situations where individuals are presented with a proposal to agree or disagree to a particular use or disclosure of their personal

> information and they respond actively to the question, orally or in writing. Usually, explicit or express consent is given in writing with a hand-written signature. For example, explicit consent will be given when data subjects sign a consent form that clearly outlines why a data controller wishes to collect and further process personal data.[…] The requirement for explicit consent means that consent that is inferred will not normally meet the requirement of Art 8(2). In this regard, it is worth recalling the Article 29 Working Party opinion on electronic health records stating that "*In contrast to the provisions of Article 7 of the Directive, consent in the case of sensitive personal data and therefore in an EHR must be* ***explicit****. Opt-out solutions will not meet the requirement of being 'explicit'.....*". (p. 25)

In transposing Directive 95/46 into national legislation, Member States have taken different approaches with regard to what is exactly required for legally valid consent (both in case the processing of sensitive and non-sensitive personal data)[49]. The USEMP consortium does not have the resources to check the national laws of all member states, nor does that seem necessary from a legal perspective: sticking to the strictest interpretation of the DPD 95/46 should suffice[50]. However, for the sake of completeness we take a quick look at Belgian and Swedish law in this regard: we pick these examples because those are Member States where two of the USEMP partners are based and where the first cohorts of DataBait users are recruited. The requirement of *explicitness* is repeated Swedish law.[51] In Belgian law[52] the

[49] For example, Spain has set set additional requirements for consent with regard to the processing of sensitive data. Article 7.2 of the Spanish Organic Law of Personal Data Protection (LOPD) 15/1999, 13th of December, requires consent for the processing of information relating to ideology, religion, beliefs and trade union membership to be "express" (i.e. explicit) and in writing. Article 7.3 requires that consent for processing of other sensitive personal data is "express" but does not require that it is in writing.
See <http://www.twobirds.com/en/news/articles/2006/use-of-consent-in-data-protection> (last accessed 12 February 2016) for a concise overview of how consent has been interpreted in some of the EU member states.

[50] Under current law (Arts. 4(1)(a) and 4(1)(c) Directive 95/46/EC) it is the *location of the establishment* of the data controller which determines applicable national law. The USEMP consortium is the joint controller of the processed DataBait data. Because USEMP is not a legal entity, there is not a single place of establishment. That means that the law of each Member State in which an USEMP partner is based (Belgium, the Netherlands, France, UK, Sweden and Greece) applies. Does this mean that the legal iCIS team has to check the national data protection laws of each of these member states? No. From a dogmatic point of view following the strictest interpretation of the DPD 95/46/EC should ensure DataBait is also compliant with national data protection laws. As clarified by the Court of Justice in Luxemburg in joined Cases C-468/10 and C-469/10, *Asociacion Nacional de Establecimientos Financieros de Credito (ASNEF)* and *Federacion de Comercio Electronico y Marketing Directo (FECEMD) v. Administracion del Estado*, 24 November 2011 (online available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62010CJ0468>), DPD 95/46 brings along a *full harmonization* which implies that Member States do not have the discretion to offer a lower protection or add additional requirements. This means that, if we follow the strictest interpretation of DPD 95/46, we don't have to check every national legislation of each of the aforementioned Member States.

[51] See the Swedish Data Protection Authority on consent ("samtycke") in the case of sensitive data (http://www.datainspektionen.se/lagar-och-regler/personuppgiftslagen/samtycke/) and Art. 13 (sensitive data) and 15 (consent for the processing of sensitive data) of the Swedish Data Protection law (https://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Personuppgiftslag-1998204_sfs-1998-204/), requiring *explicit* consent ("sitt *uttryckliga* samtycke")

[52] See the Belgian Data Protection Authority on consent ("toestemming") in the case of sensitive data (https://www.privacycommission.be/nl/gevoelige-gegevens) and Art. 6.2(a)(consent for the processing of sensitive data) of the Belgian Data Protection law (https://www.privacycommission.be/nl/node/3788), requiring *written* consent ("schriftelijke toestemming")

only requirement is that the consent for the processing of sensitive data has to be *written*[53]. Thus, the way DataBait requests consent for the processing of sensitive data (through a separate screen in the DLA) is in line with both Belgian and Swedish law.

Under the GPDR, that is, the new legal data protection regime, the requirement for consent for the processing of sensitive data is still that is has to be *explici*t (Art. 9.2(a) GDPR). In this respect things have stayed unchanged. However, what is relevant in the GDPR is that the standard for consent in general has been made stricter. Consent will have to be "unambiguous" (Art.4(8)):

> 'the data subject's consent' means any freely given, specific, informed and unambiguous indication of his or her wishes by which the data subject, either by a statement or by a clear affirmative action, signifies agreement to personal data relating to them being processed';

In order for consent to be considered as "freely" given, "utmost account shall be taken of the fact whether, among others, the performance of a contract, including the provision of a service, is made conditional on the consent to the processing of data that is not necessary for the performance of this contract" (Art. 7(4) GDPR). This additional test does not pose any problems for the consent of DataBait users: the processing of sensitive data (and informing users about it) is the core business of DataBait and thus clearly necessary for the performance of the contract between the DataBait user and the USEMP consortium.

A new requirement in the GPDR is that a written request for consent for the processing of any personal data must be presented in a manner which is "clearly distinguishable" from the rest of the written context in which it is presented (Art. 7(2) GDPR):

> "If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent must be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of the declaration which constitutes an infringement of this. Regulation that the data subject has given consent to shall not be binding."

We present the DLA in the form of separate screens (each article of the DLA is a separate screen – including the one on sensitive data) in order to fulfill the requirement of Art. 7(2) GDPR (i.e., to present the request for consent "in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language"). We do not 'hide' the request for consent in some general terms of service: the whole DLA concerns the fulfilment of data protection requirements and offering

[53] If "written" is interpretated is in line with EU legislation on electronic commerce and digital signatures, it is likely to include "handwritten" consent on paper as well as (some) electronic forms of consent. See WP 29, Opinion Opinion 15/2011 on the definition of consent (adopted on 13 July 2011), p. 26.

transparency about it. Whether our current presentation could be even further enhanced is something which we will discuss with the other partners. One could for example consider giving the screen with the consent for sensitive data another color. However, it would be also unnecessary to “overdo” it and scare users away from a tool that is precisely empowering in terms of data protection.

3.3. The grey zone of what qualifies as sensitive

There are data which are obvious cases of sensitive data. For example, a medical record in a hospital is an obvious instance of health data (pertaining to the health status of the data subject) and a municipality record stating that a person is Roma, Jew, etc. is an obvious instance of data revealing race or ethnic origin. There are also data which are seemingly innocuous. For example, think of a data subject who has uploaded a holiday picture on Facebook where she is standing in a bar with a sigarette and a glass of wine or a data subject who regularly uploads the data of her running app (stating where she runs,how fast, her heart rate, how many calories she has burned, etc.). These data are in some way related to health, but are they health data (and thus sensitive)? And does the fact that one’s skin color is visible in a picture make it racial data? n establishing whether personal data are sensitive or not, it is not enough to take raw data at their face value – also their *potential* sensitivity in case of inferences or combination with other data should be taken into account. With regard to establishing whether such data ‘from the grey zone’ is sensitive, the difficulty lays both in establishing the technical possibilities (*what can be extracted from the data? which software possibilities exist? what can be extracted from the data in combination with other data sets an which data sets are available for such combinations?*) and in the fuzziness of the qualification rule (*some* possibilities to extract sensitive information from a piece of data make it sensitive in the sense of Art. 8 DPD 95/46 and *some* don’t). In order to clarify which potentialities are relevant we follow the line set out by Working Party 29 with regard to health data (Article 29 Data Protection Working Party, 2015a, 2015b) and apply their analysis in an analogous way with regard to other types of sensitive data.

The bottom-line is that it would be unsustainable to consider every potentiality as a ground for qualifying data as sensitive. For example, photos and videos containing images of people can be a source of all kind of sensitive information: smart analytic software could extract racial (e.g. based on skin colour) and religious features (e.g. based on whether someone is wearing certain religious garments or jewelry), health or biometric data (e.g. based on gait, body shape, activity pattern, skin colour, behavioral analysis etc.). However, requiring that any photo or video containing a face or body of an identifiable person should be qualified as sensitive and thus be subjected to the extra strict regime of handling sensitive personal data would put an unreasonable burden on data controllers operating on the internet. Working Party 29 has given two possible situations when a piece of data which is not a clear-cut[54] case of health data (e.g. medical data) nevertheless qualifies as health data:

[54] “There remain some types of processing, where it is not obvious at first sight whether or not the processing of these data should qualify as the processing of health data. This is especially the case where the data are

1. The data are raw sensor data that can be used in itself or in combination with other data to draw a conclusion about the actual health status or health risk of a person: i.e., there has to be a demonstrable relationship between the raw data set and the capacity to determine a health aspect of a person,

2. Conclusions are drawn about a person's health status or health risk (irrespective of whether these conclusions are accurate or inaccurate, legitimate or illegitimate, or otherwise adequate or inadequate)

How to transpose this as a guideline in deciding whether any non-obviously sensitive data qualifies as sensitive? We think that following the lines set out here by WP29 for health data, two general guidelines could be proposed for the assessment of any data which are neither obviously non-sensitive nor a clear-cut instance of sensitive data. Firstly, that it does not matter whether an inference is correct: if it is likely that a company will use smart software which wrongly classifies all people with dark hair as “Asians”, this is nevertheless processing of sensitive (racial) data. Secondly, we propose that what is important in non-obvious cases is that there needs to be a realistic possibility and a significant chance that sensitive information will be extracted and used as such. We would like to propose that “intended use” (Article 29 Data Protection Working Party, 2015a, p. 4) of the data, that is, the concrete context in which data are likely to be used, could be a very useful criterion in deciding whether data qualifies as sensitive in non-obvious cases. Both “a demonstrable relationship” (Article 29 Data Protection Working Party, 2015a, p. 4) between raw data and sensitive information and the question whether it is likely that conclusions with regard to sensitive matters will be drawn from the data, show that the qualification of data from the ‘grey area’ (where it is not immediately obvious whether data should be qualified as sensitive or not) should not be an abstract exercise in theoretical possibilities and potentialities but look at concrete possibilities and intended uses (e.g. *What is the business model of the data controller processing these data? Does the data controller have access to other data bases that allow combinations of data that make it possible to infer information regarding health status?*) The notion of ‘intended use’ as we propose it should be distinguished from the specified purpose of the processing. The latter is a legal requirement for processing, while intended use is a factual criterion for establishing whether non-obvious cases qualify as sensitive data. Purpose specification and intended use can be related (e.g. when a data processor explicitely specifies that she will combine weight data with opinion data to assess the mental health of the data subject) but do not necessarily coincide (i.e. if a data controller does *not* state it as a processing purpose to establish the health status, religion, race, etc. from a certain type of data, this does not exclude the possibility that the concrete context can nevertheless make it likely that the data controller *could and would* extract this sensitive

processed for additional purposes and/or combined with other data or transferred to third parties. These types of data processing may create risks, including the risk of unfair treatment based on data about a person's assumed or actual health status. Clearly, these types of data processing deserve significant attention. If data are health data, but mistakenly treated as 'ordinary' personal data, there is a risk that the high level of protection deemed necessary by the European legislator is undermined. If seemingly innocuous raw data are tracked over a period of time, combined with other data, or transferred to other parties who have access to additional complementary datasets, it may well be that even the seemingly most innocuous data, combined with other data sources, and used for other purposes, will come within the definition of 'health data'.This risk specifically applies to further processing of such data for profiling and marketing purposes, given that the key business model of most apps is based on advertising.” (Article 29 Data Protection Working Party, 2015a, p. 3)

information).Let's take a closer look at how the Working Party arrives at its recommendation for qualifying health data through the lens of the notion of 'intended use'

The Working Party begins by observing that health data is a broad notion, covering anything from direct medical data, stating that a person has a certain disease, to more indirect data (e.g. data about buying certain medical devices, being a member of a support group or association like Weight Watchers or Alcoholics Anonymous, etc.), to 'light' medical information (e.g. wearing contact lenses or allergy information), to "data about a person's intellectual and emotional capacity (such as IQ), information about smoking and drinking habits" (Article 29 Data Protection Working Party, 2015a, p. 2), or data indicating a good (instead of ill) health. Also lifestyle data which might contribute to the establishment of disease risks can be health data.

> "According to the Working Party, health data therefore also include information about a person's obesity, high or low blood pressure, hereditary or genetic predisposition, excessive alcohol consumption, tobacco consumption or drug use or any other information where there is a scientifically proven or commonly perceived risk of disease in the future." (Article 29 Data Protection Working Party, 2015a, p. 2)

However, the Working Party also underlines that it assumes that :

> "…there is a category of personal data generated by lifestyle apps and devices that is, in general, not to be regarded as health data within the meaning of Article 8. This concerns data from which no conclusions can be reasonably drawn about the health status of a data subject. Not all raw data collected through an app (measurements) qualify as information (from which meaning can be derived) about the health of a person. For example, if an app would only count the number of steps during a single walk, without being able to combine those data with other data from and about the same data subject, and in the absence of specific medical context in which the app data are to be used, the collected data are not likely to have a significant impact on the privacy of the data subject and do not require the extra protection of the special category of health data. They are just raw (relatively low impact lifestyle) personal data (provided, the app does not process location data), not information from which knowledge about that persons health can be inferred." (Article 29 Data Protection Working Party, 2015a, p. 3)

We would thus add that whether such data are in the 'grey area' much depends on the *intended use* of the data. If an individual piece does not constitute health data as such (e.g. the amount of steps someone made during a day), it should be checked whether there is an intended plan to combine it with other data to use it in a health related way.

> "There has to be a demonstrable relationship between the raw data set and the capacity to determine a health aspect of a person, based on the raw data itself or on the data in combination with data from other sources. For example, if a diet app only counts the calories as calculated from input provided by the data subject, and the information about the specific foods eaten would not be stored, it would be unlikely that any meaningful conclusions can be drawn with regard to the health of that person (unless the daily intake of calories is excessive in absolute terms). But if data from a diet app, or heart rate monitor or sleep diary app are combined with information provided by the data subject (directly or indirectly, for example based on information collected from that person's social networking profile), conclusions (whether accurate

> or inaccurate) may be drawn about that person's health condition, such as medical risk or diabetics. In these cases it is likely that health data can be inferred from the combined data." (Article 29 Data Protection Working Party, 2015a, p. 4)

The Working Party states that data following from opinion or mood analysis conducted on texts posted in social media *could* constitute health data depending on the actual context and the purpose of the processing :

> " An example […] is analysis conducted on social media to detect whether people may suffer from a depression. Even though 'sad' messages sent by users, in general, do not have to be treated as health data by (generalist) social networks, the systematic analysis of such messages for the purpose of diagnosis/health risk prevention or medical research certainly qualifies as the processing of health data." (Article 29 Data Protection Working Party, 2015a, p. 3)

In summary, we think that in difficult, non-obvious, cases, in order to establish whether a piece of data should be qualified as sensitive data, it is important to look at the intended use of the data. It is not enough to look at a piece of data in isolation : if the intended use entails that the data will be combined with other data in such a way that what can be inferred from the data becomes more health related, this should be taken into account.We propose this rule of thumb : whether a *potential* to derive sensitive information from 'innocuous' (i.e. non-sensitive) raw personal data should result in the qualification of these data as 'sensitive' depends on their *intended use* (realistic possibility and significant chance that sensitive information will be extracted and used, given the concrete circumstances). Thus, a picture depicting faces of a particular skin color does not necessarily constitute sensitive data about race and ethnic origin – it depends on the intended use.

What does this imply for the derived data in the USEMP project and the raw data from which they are derived? The DataBait tool will inform the user about the possible information which can be extracted from her data. Thus, we will inform the user that data about race, health, religion etc. could be extracted from a certain picture, and that this could mean that the picture should be treated as sensitive data, but that this will depend on the actual intended use of the picture. The DataBait user will be given realistic examples of the intended uses for which the informaton could be used.

In table 2 we list the qualifications of the various derived data types. The DataBait user will be informed about which raw data could potentially be qualified as sensitive based on the potentially sensitive data which could be derived from it.

	'Privacy dimensions (i.e., categories into which the derived data are organised: see D6.1)	**Derived attributes**	**Legal qualification in terms of EU data protection (DP) law and EU anti-anti-discrimination (AD) law.** **SPD: sensitive personal data;** **PD: personal data**
	Demographics	1. Age	DP: PD AD: Protected ground in the field of employment
		2. Gender	DP: PD

			AD: Protected ground in the field of (1) employment, (2) access to goods and services
		3. Nationality	DP: PD AD: Protected ground but many exceptions (i.e. particular areas where differentiation based on nationality is allowed)
		4. Racial origin	DP: SPD 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	DP: SPD 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	DP: PD
		7. Employment status	DP: PD
		8. Income level	DP: PD
		9. Family status .	DP: PD, could be SPD if it reveals information about one's sex life (or according to the pGPDR: sexual orientation or gender identity)AD: if the data reveals sexual orientation- this is a protected ground in the field of employment law
	Psychological Traits	1. Emotional stability	DP: PD; possibly SPD (if characterized as health data)
		2. Agreeableness	DP: PD; possibly SPD (if characterized as health data)
		3. Extraversion	DP: PD; possibly SPD (if characterized as health data)
		4. Conscientiousness	DP: PD; possibly SPD (if characterized as health data)
		5. Openness	DP: PD; possibly SPD (if characterized as health data)
	Sexual Profile	1. Sexual preference	DP: SPD AD: if the data reveals sexual orientation- this is a protected ground in the field of employment law
	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)	DP: SPD

		2. Political ideology (Communist; Socialist; Green; Liberal; Christian democratic; Conservative; Right-wing extremist)	DP: SPD
	Religious Beliefs	Supported Religion (Atheist, Agnostic, Christian, Muslim, Hinduist, Buddhist, Other, etc.)	DP: SPD AD: religious belief is a protected ground in the field of employment law
	Health Factors & Condition	1. Smoking	DP: PD; possibly SPD (if characterized as health data)
		2. Drinking (alcohol)	DP: PD; possibly SPD (if characterized as health data)
		3. Drug use	DP: PD; possibly SPD (if characterized as health data)
		4. Chronic diseases	DP: PD; possibly SPD (if characterized as health data)
		5. Disabilities	DP: PD; possibly SPD (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)	DP: PD; possibly SPD (if characterized as health data)
	Location	1. Home	DP: PD
		2. Work	DP: PD
		3. Favourite places	DP: PD
		4. Visited places	DP: PD

	Consumer Profile	1. Brand attitude	DP: PD
		2. Hobbies	DP: PD; possibly SPD 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'shealth or sex life.
		3. Devices	DP: PD
	n.a.	Detection of faces in images (number and location)	DP: PD
	n.a.	Detection of opinion (positive/negative/ neutral) from textual posts and status updates	DP: PD
	n.a.	Disclosure score (How sensitive, uncontrollable and visible are your data?)	DP: PD
	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 valuable?)	DP: PD

Table 2. How should the data derived from the raw DataBait data be qualified in terms of Data Protection (DP) law and Anti-Discrimination (AD) law? In terms of DP law data can be either non-sensitive personal data (PD) or sensitive personal data (SPD)

3.4. 'Intended use': Sensitive data and anti-discrimination law.

Compared to the legal instruments of the EU, the rights derived from the European Convention on Human Rights (ECHR) often provide a broader but also a fuzzier protection. Not only because the primary goal of the ECHR is to protect the individual citizen against the State (and not against Facebook, Google or a databroker), but also because the route to the

Court in Strasbourg (a measure of last resort) is longer than the route with regard to EU legislation (or the national implementation thereof). National courts can raise preliminary questions with the Court of Justice of the European Union (CJEU) in Luxemburg about the interpretation of EU law. Nevertheless it is also precisely the broad formulation of ECHR rights which can sometimes provide protection where the more specific provisions of the EU fail to do so. This is particularly clear in the field of anti-discrimination law. As shown in figure 2, the anti-discriminatory law of the EU offers protection with regard to a very specific set of protected grounds (listed in Art. 13 of the *Treaty Establishing the European Community*[55] [TEC, 1997; entry into force in 1999]: sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation) and areas of life. The largest protection is offered with regard to race, and in the field of employment.

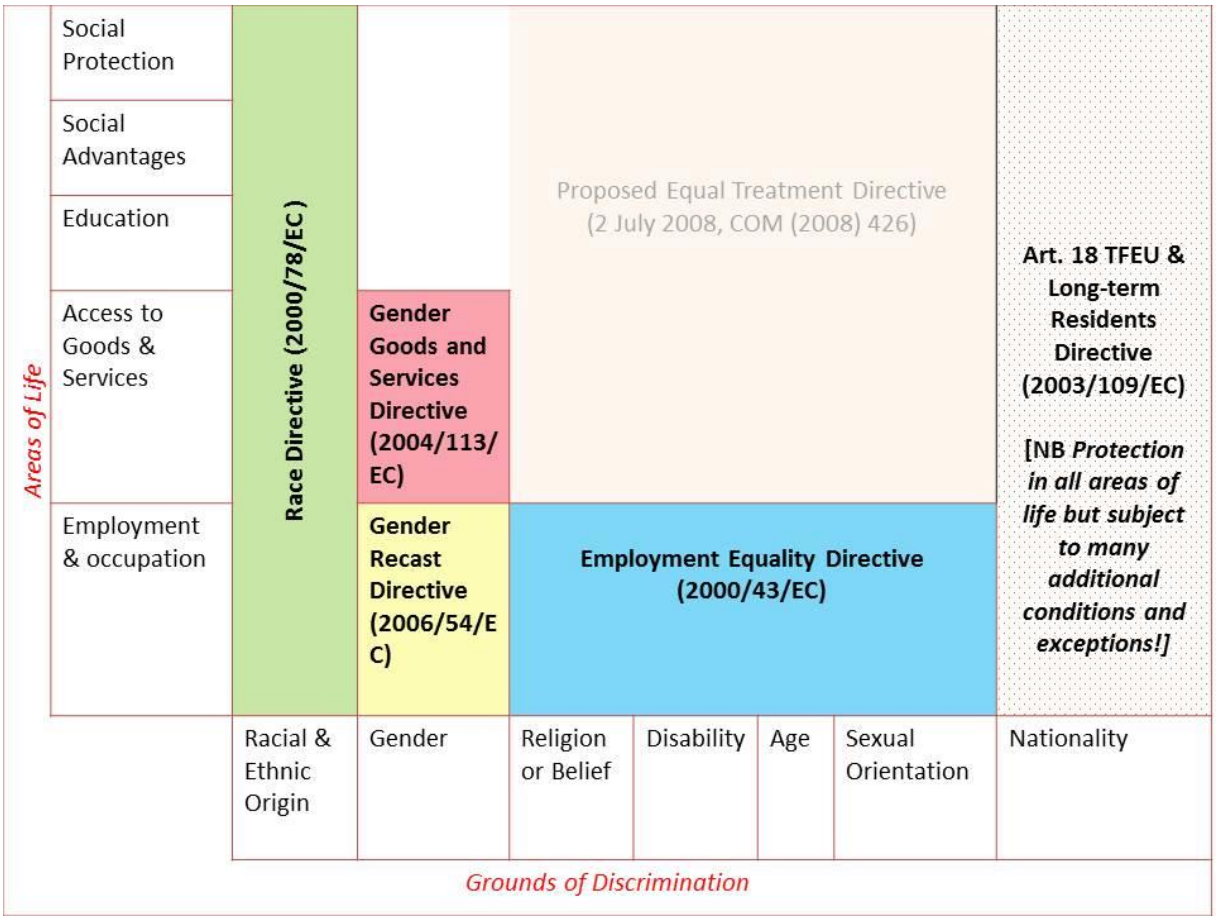


Figure 7: Protected grounds and areas of life in secondary EU anti-discrimination law

[55] Now replaced by Article 19 of the *Treaty on the Functioning of the Union* (TFEU, 2008). The content of Art. 19 TFEU and Art. 13 TEC is identical.

When comparing the anti-discriminatory provisions from EU data protection law with those from EU anti-discrimination law, there are some interesting overlaps as well as differences to be pointed out (see table 2).

Data Protection	Art. 9 (1) of the proposed General Data Protection Regulation (GDPR)	The processing of personal data, revealing: **race or ethnic origin, political opinions, religion or beliefs**, trade-union membership, and the processing of ***genetic data*** or data concerning *health or sex life* or criminal convictions or related security measures shall be prohibited.
	Art. 20 (3) of the proposed General Data Protection Regulation (GDPR)	Profiling that has the effect of discriminating against individuals on the basis of **race or ethnic origin, political opinions, religion or beliefs**, trade union membership, *sexual orientation or gender identity*, or that results in measures which have such effect, shall be prohibited. The controller shall implement effective protection against possible discrimination resulting from profiling. Profiling shall not be based solely on the special categories of personal data referred to in Article 9
Anti-Discrimination	Art. 21 Charter of fundamental rights of the European Union (CFREU)	(1) Any discrimination based on any ground such as: sex, **race**, colour, **ethnic** or social **origin, *genetic features***, language, **religion or belief, political or any other opinion**, membership of a national minority, property, birth, *disability*, age or *sexual orientation* shall be prohibited. (2) Within the scope of application of the Treaty […] any discrimination on grounds of nationality shall be prohibited.
	Art. 13 Treaty Establishing the European Community (TEC)	…take appropriate action to combat discrimination based on sex, **racial or ethnic origin, religion or belief**, *disability*, age or *sexual orientation*

Table 3: Discrepancies and overlaps between (1) the data categories classified as sensitive or discriminatory in EU data protection law and (2) the prohibited grounds in EU discrimination law. In this table the bold categories are the ones that overlap, the italic ones partly overlap, and the underlined ones are new additions to the list of sensitive data in Art. 9 (1) of the pGDPR (in comparison to the ones mentioned in Art 8(1) of the current DPD 95/46. This table is an updated and adjusted version of the table in: (Gellert, de Vries et al. 2012)

One way to explain these overlaps and differences is that data protection is more oriented on the *process* of data processing, while the anti-discrimination provisions look at discriminatory *effects*. Thus, data such as sex, age, and nationality (which is the kind of basic

information which one is required to provide frequently in OSNs) are not considered to be sensitive data from a data protection perspective, but as soon as one begins to take discriminatory measures based on them, for example in the area of employment, they become “toxic”. While the processing of *sensitive* data -for example, data which reveal racial origin or political opinions- requires additional safeguards in comparison to the processing of “ordinary” personal data (even when no actual discrimination results from it), data such as sex, age, and nationality are not considered to be sensitive *as such*.

However, when the sensitivity of data is to be judged in terms of their ‘intended use’ (see section 3.2), it becomes clear that the prohibition to process sensitive data (unless an exception such as explicit consent applies) comes closer in rationale to the ‘effect’-oriented provisions from anti-discrimination law. The provisions are increasingly merging into a partly overlapping continuum. Therefore we propose that the *DataBait* should inform users both of the protection with regard to sensitive data and the anti-discrimination provisions which might apply.

In designing the DataBait tool the notion of ‘intended use’ should be incorporated in the information provided to the user, when informing her that her data can be categorized as (potentially) belonging to the specific categories of data in EU anti-discrimination law (protected grounds) and EU data protection law (sensitive data and the protected grounds mentioned in Art. 20(3) GDPR). The question which needs to be posed is whether these categories of data are (likely to be) processed by commercial profilers. Additional question to be explored are how the user should be informed of the relevant legal provisions with regard to these particular kinds of data and whether users feel that the sensitive data and protected grounds deserve a higher level of protection than other data (e.g. income, log-in patterns, educational level, etc.)?

4. Research strand 3: A modular DLA

4.1. A DLA: legal ground and legitimate purpose

As discussed in section 1.1, lawful processing of personal data always has to be based on a legal ground legitimizing the processing and have a legitimate, specified and explicit purpose by which the allowed usages of the data are limited. Both conditions have to be fulfilled to make the processing lawful. With regard to the first requirement the DPD enumerates six legal grounds in Art. 7 DPD 95/46.

CRITERIA FOR MAKING DATA PROCESSING LEGITIMATE

Article 7

Member States shall provide that personal data may be processed only if:

(a) the data subject has unambiguously given his consent; or

(b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; or

(c) processing is necessary for compliance with a legal obligation to which the controller is subject; or

(d) processing is necessary in order to protect the vital interests of the data subject; or

(e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed; or

(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1).

The first, and most well-known, legal ground is “freely given and informed *consent*, the other five concern *necessity* in relation to (b) a contract, (c) a legal obligation, (d) the vital interests of the data subject, (e) the public interest or (f) the legitimate interests of the data controller (if these interests are not overruled by the fundamental rights of the data subject).” (Hildebrandt, 2014, p. 24)

The second requirement, which is a combination of purpose specification and use limitation, is described in Art. 6(b) DPD 95/46 (see also our discussion purpose specification in section 1.1):

"Personal data must be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes".

Both the requirement of a legal ground and purpose specification/use limitation need to be fulfilled, which means that:

"…one *cannot* consent purpose limitation away; a valid new legal ground does not imply that historical data can now be used for an incompatible purpose in relation to the one for which they were originally processed. Purpose binding thus ties whoever processes personal data to the explicit legitimate purpose as it was specified upfront, when the data were first collected. It chains that entity to its own stated – and necessarily legitimate - purpose." (Hildebrandt, 2014, p. 24)

The purpose of the processing is determined by the data controller.[56] Because the USEMP consortium has jointly determined what the purpose and means of the processing of personal data will be, and should thus be qualified as a joint data controller[57]. As the USEMP consortium does not possess legal personality, it was important to create an internal agreement between the partners (see section 4.4.) in which partners commit to implementing relevant data protection law when processing the personal data of USEMP end-users, while each partner exonerates the others from liability for data processing which is not under the actual control of these other partners.

The USEMP consortium has chosen to take the ground from Art. 7(b) as the legitimizing ground for all the processing: a data licensing agreement (DLA) between the USEMP consortium (the joint data controller) and the end-user of the USEMP tools is the legitimizing ground for the data processing in USEMP (see section 4.3 and Annex 7.2). It should be noted that this DLA is *not* merely a service license agreement (SLA) in which the part about data processing is only an appendix – i.e., a consent form attached to the main service agreement – but that this contract actually focuses on the *purpose* of data processing within the USEMP project and the mutual obligations between the USEMP consortium partners (the joint data controllers) and the end-user of the USEMP tools. These obligations are created in order to fulfil that purpose. Opting for a DLA, rather than the usual combination of a SLA combined with a privacy policy, user consent and lengthy terms and conditions, also aligns with the USEMP proposal to enable the licensing of the use of personal data by data subjects, as described in the USEMP Description of Work (DOW). Another way in which the DLA embodies the objective of user empowerment, is that it keeps matters as straightforward as possible and puts them in plain language: the DLA avoids any unnecessary "legalese". The DLA is implemented in the USEMP graphic user interface (GUI) and is part of the sign-up procedure. It is impossible to sign-up or use the DataBait tools

[56] "'**controller**' shall mean 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", Art. 2(d) DPD 95/46.
[57] "Joint controllers. Where several controllers jointly determines the purposes and means of the processing of personal data, the joint controllers shall determine their respective responsibilities for compliance with the obligations under this Regulation, in particular as regards the procedures and mechanisms for exercising the rights of the data subject, by means of an arrangement between them. The arrangement shall duly reflect the joint controllers' respective effective roles and relationships vis-à-vis data subjects, and the essence of the arrangement shall be made available for the data subject. In case of unclarity of the responsibility, the controllers shall be jointly and severally liable" (Art 24 of the pGDPR)

without first signing the DLA. Each article of the DLA is presented as a separate screen. All the text fits easily on one screen, making it unnecessary for the user to scroll down.

Building the DLA and the internal agreement into the DataBait GUI can also be considered as a way in which USEMP realizes Data Protection by Design (Art. 23 pGDPR) with regard to Art. 7 DPD 95/46 (legal ground) and the requirement of *data minimization* (Art. 6 DPD 95/46), which is an umbrella term for the requirement of *purpose specification* (that data must be collected for specified, explicit and legitimate purposes and that they must be adequate, relevant and not excessive in relation to the purposes for which they are collected), *use limitation* (that data should not be further processed in a way incompatible with those purposes), that data have to be *accurate and complete*, and that they are deleted or anonymised as soon as they are no longer needed for the purpose that led to their collection.

NB! An important counter argument with regard to the use of a DLA could be that such an agreement could be used to replace privacy policies, thus disabling data subjects from easily withdrawing their data. Such a DLA could integrate the usual complex and intransparent language, meant to lure data subjects into signing away the control over how their personal data is used. The USEMP DLA not only provides for, but also depends on the profile transparency that is a precondition for informed consent regarding the use of one's personal data. Our argument is not based on the idea that any type of DLA necessarily offers more protection than consent of processing based on the f-ground. On the contrary, the USEMP DLA is based on the fact that the b-ground allows to process only those personal data necessary for the performance of the DLA, meaning that a clear and enforceable description must be provided of the mutual obligations generated by the DLA. Further protection derives from paying keen attention to a number of private law and consumer law provisions that protect the weaker party against unreasonable clauses in contracts concluded between consumers and businesses. Finally, we argue that the USEMP DLA is highly relevant for all business models that are based on providing a so-called 'free service', by clarifying in clear and enforceable terms how the processing of the user's volunteered, observed and inferred data may affect the way she can be targeted by the service provider, third parties and government authorities.

4.2. A DLA as a tool to adjust power imbalances

One of the main legal concerns during the first year of the USEMP project has been to develop so-called a Data Licensing Agreement (DLA) for those participating as end-users of the USEMP platform. The aim in developing the DLA is threefold:

1. To provide a legitimate legal ground for the processing of personal data, notably also sensitive data and for downloading the USEMP DataBait tools;
2. To engage the end-user (data subject) by asking her to enter into an obligatory agreement with the USEMP partners (joint data controllers), clarifying mutual rights and obligations;

3. To present the end-user (data subject) with a clear, concise transparent agreement that is legible for lay people and covers all the relevant issues of compliance on the side of the USEMP service providers (joint controllers)

In this section 4.3 we therefore present the DLA and in section 4.4 the underlying personal data processing agreement that has been concluded between the USEMP Consortium Partners (as joint controllers), thus binding the partners to provide some form of profile transparency in exchange for a specified license to process the user's (data subject's) personal data. We will explain the relationship between the DLA and the consent requirement for processing sensitive data (art. 8 DPD) and between the DLA and the consent requirement for storing tracking mechanisms on the user's (subscriber's) device (art. 5.3 ePrivacy Directive).

The idea of employing a data licensing agreement is new and hopes to provide for a new way of addressing the power imbalances between users and providers of OSNs[58]. It is based on the fact that data subjects have a bundle of rights with regard to the processing of their personal data. This allows them to contract about such processing to the extent that processing is not e.g. mandatory for reasons of public security or necessary for the legitimate interest of the data controller.

4.3. The USEMP DLA

As indicated, the data licensing agreement (DLA) will be concluded between the USEMP Consortium Partners (as joint data controllers) and the end-users of the USEMP tools. It clearly defines the mutual legal obligations, taking the end-users seriously as participants in the research that is conducted. It is also the legitimizing ground for the data processing in USEMP ("contract" as decribed in Art. 7b of Data Protection Directive 95/46).

The DLA is implemented in the USEMP graphic user interface (GUI) and is part of the sign-up procedure. Each article of the DLA will be presented as a separate screen. The underlying Personal Data Processing Agreement (PDPA, see below) can be seen as an offer made by all each of the USEMP Consortium Partners to conclude the DLA; when the end-user clicks accepts this offer by clicking the button at the end, each USEMP Consortium Partner is bound by the DLA. We note that:

[58] The fact that the service of an OSN is rendered at no cost does not justify a weak position of the user in terms of consumer and data protection. Moreover, the notion of "service at no cost" must be nuanced. See e.g. Wauters e.a. 2014, p. 10: "*Since most SNS do not require an actual payment of a fee, we wonder if SNS can fall under the scope of the Consumer Rights Directive. […] However, it is often stated that personal data is the new currency of the Internet. A SNS offers its service to users and in exchange, they gather (explicitly through registration forms or 'secretly' via cookies) personal data of their users. Because of this personal data, they are able to offer personal advertisements in order to make a profit. Another indication may be found in the definition of information society services under the e-Commerce Directive (above), which includes service which are financed by advertising.*" Following Wauters it might be argued that based on the Consumer Rights Directive the license granted by the users to Facebook is too broad and not legally valid. The PDPA which is signed between the USEMP consortium and the users of the DataBait tool is a first step to a more balanced approach, and which can form the basis for a more granular licensing approach. This entails that a later, modular version of the DLA should include licensing of copyrighted material posted on the OSN.

- Articles A and F clearly define the obligations of DataBait users, while also specifying the consequences in terms of which data will be tracked.
- Article B clarifies that this agreement entails that the DataBait users license the usage of their personal data for a specified purpose.
- Article C provides for the consent required in art. 5(3) of the ePrivacy Directive.
- Articles D and E further specify the purpose for which the USEMP Consortium Partners will use and process the data, notably in terms of the OSN presence management tool and the monetization tool.
- Article G provides for the consent required for the processing of sensitive data (art. 8 DPD)
- Article H, I, J further specify the duty of care for the USEMP Consortium Partners when they process the personal data of the DataBait users, stipulating the ilfe cycle management of the involved personal data (collection, usage, deletion or full anonymisation). Art. H also emphasizes the reason for processing sensitive data.

In section 4.5 we present a modular version of the DLA, enabling data usage licensing via DataBait tools for profile transparency, with other service providers that may have a commercial interest in providing the tools. This entails that the purpose is extended or adapted.

Screen 1:

USEMP Data License Agreement

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]59, provider of the USEMP platform and services, joint data controllers, from hereon called 'USEMP consortium partners' 60.

Hereby agree:

section 4.5

Screen 2:

[59] The name of each partner will contain a hyperlink, such that users can click on it and check the organisational website of which is involved.

[60] This text will contain a hyperlink, stating:"*The USEMP consortium partners have entered a separate agreement between themselves, obliging themselves 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.*" When one clicks one "here" this will lead to the *USEMP Personal Data Processing Agreement* (see Appendix 3 of this Deliverable).

(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 article defines the obligation to install the DataBait tools, which is pertinent for participation in the USEMP research. It clarifies upfront that both volunteered (declared) data will be processed and observed (behavioural) data. In a later, modular version of the DLA, not necessarily focused on scientific research, the same article can be used.

Screen 3:

(B) You license the use of Your volunteered and observed personal data by the USEMP consortium partners, as gathered by the 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 article, first, makes clear that this is a *quid quo pro* agreement, creating legal obligations on the side of the user (data subject) in the form of licensing the use of the data that will be processed by the USEMP consortium, and on the side of the service provider (data controller) in the form of providing a form of profile transparency. Second, it determines the specific purpose of processing. As explained in section 4.4, in the modular version of the DLA, not necessarily focused on scientific research, part of this article *("…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")* will have to be adapted

Screen 4:

(C) This license agreement confirms Your explicit consent to store the DataBait tools on Your devices.

This article provides the consent required on the basis of art. 5.3 ePrivacy Directive for all and any tracking mechanisms to be stored on the user's (data subject's) device. That such tools contain tracking mechanisms is clarified in the previous articles A and B – the consent thus includes any cookies that are stored on the device, which are – in this case – necessary to fulfil the functionality of the service that is provided. This means that consent may not be required, since – according to the art. 29 WP consent is not required for functional cookies. To be on the safe side we have included this consent. We advise that this article is part of the modular versions of the DLA discussed in section 4.5.

Screen 5:

(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 article further explains the obligation on the side of the service provider and the purpose of processing, highlighting that the profile transparency provided is based on statistical inferences by others than OSN providers, meaning that the user is made aware of the fact that the USEMP Consortium partners are not reverse engineering software code of the OSN provider and cannot in any way provide certainty about how one may be targeted. This article also ensures that the transparency is provided in a user-friendly manner. This article is crucial in any DLA for DataBait tools. However, the term ‘scientific research’ might have to be adapted in some of the modular versions of the DLA (see section 4.4).

Screen 6:

(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.

As with the previous article, this article highlights that the monetary value is an estimation and in no way a claim as to the actual monetary value that may be generated with the DataBait tools. In fact the consortium has decided to refrain from providing an estimate of the monetary value, instead developing an estimate of the added value for the OSN provider or third parties with whom data may be shared. This article is not necessary in the modular version of the DLA, as it is very specifically related to the USEMP consortium and the particular functionality of the DataBait tool.

Screen 7:

(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 article clarifies that the user will participate in the research that enables to correlate their declared preferences or personality traits with the inferences drawn from behavioural data or the mining of multi-media content. This article may be part of the modular version of the DLA (section 4.5), depending on the particularities of the transparency tool and the services accompanying it. Any reference to USEMP will of course have to be removed in the modular version.

Screen 8:

(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.

Since consent is required for processing art. 8 DPD types of data, this article stipulates such consent. It highlights the intrusive nature of the processing of such data. It is part of the modular version of the DLA (section 4.5.).

Screen 9:

(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.

See commentary below the next Article.

Screen 10:

(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, anonymised or processed for related scientific research. In the latter case the relevant USEMP consortium partner will ask You for Your consent.

Articles H and I confirm the legal obligation for the USEMP partners (joint controllers) that the relevant data will be processed in accordance with the data minimisation principle, stipulating deletion or anonymisation as soon as possible (including a clear deadline) and security by design, while also explaining that to provide profile transparency the processing of both personal and sensitive personal data is necessary. These articles are part of the modular version of the DLA (section 4.5.) considering that this is a confirmation and reminder of the legal obligations of the service provider (data controller).

Screen 11:

(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 article is pivotal to ensure that in the context of USEMP data are not processed beyond the explicitly specified purpose, by the parties to the contract, simply prohibiting any transfer to third parties other than the FIRE infrastructure[61]. The article can be modulated depending on the specifics of the modular version of the DLA, for instance allowing to share data with specified third parties and/or specified types of third parties.

Screen 12:

(H) 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:

To prevent any confusion about the applicable national law, and to accommodate the natural person whose personal data are being processed we confirm that the national law of the end-user (data subject) of the USEMP platform is applicable. Under current EU Data Protection Law this seems the most apt, also for the modular version of the DLA (section 4.5). This may change under the proposed General Data Protection Regulation (pGDPR).

4.4. The USEMP PDPA

The USEMP DLA is included in an internal agreement between the USEMP Consortium Partners. This internal agreement, which we call the USEMP Personal Data Processing Agreement (PDPA), specifies which partner will do what kind of processing of personal data, and determines that and how the Consortium Partners are legally bound to treat the personal data they are processing. It also includes a clause which binds each partner to the DLA. We note:

- The USEMP partners act as joint data controllers because they have jointly determined the purpose of the processing of personal data within the USEMP project, namely scientific research as explicated in the DOW, the DLA and the PDPA.
- The DLA is part and parcel of this contract; the PDPA is an irrevocable offer to DataBait end-users to conclude the DLA contract. A link will be placed in the DLA to the DPDA contract.
- The PDPD contains strict obligations in terms of the appropriate security measures regarding the capture, storage and transmission of personal data, based on a risk assessment performed by each partner.
- The PDPA thus clarifies to the end-users of the USEMP tools which partner does what kind of processing of data and, finally exonerates partners from liability for data processing performed by other partners over which they have no actual control.
- The PDPA also addresses the user-friendly, layered and precise information to which end-users of the USEMP platform (data subjects) are entitled by stipulating that two buttons will be visible and operational on the platform's website: (1) to obtain more detailed information about the way USEMP Consortium Partners are bound to deliver on

[61] http://cordis.europa.eu/fp7/ict/fire/

the contract, by showing the PDPA contract and by adding a table which shows in even more detail what data are processed how and for what reasons in the design of the USEMP architecture; and (2) to obtain from the USEMP Consortium Partners the erasure of their sensitive data or the removal of the DataBait tools.

Below we reproduce the complete PDPA:

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:

(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 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.

(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.

(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 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 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.

(G) Belgium law will be applicable to this contract.

4.5. A modular DLA

In this section we present a *modular* version of the DLA, which could be used by different types of providers of profiling transparency tools (modularity 1) and with regard to different OSNs and browsers (modularity 2).

Modularity 1:

Who is the provider of the profile transparency tool?

(a) another scientific consortium (not USEMP)

(b) an OSN (either simulating inference mechanisms or providing insight in the real inference mechanisms)

(c) a third-party commercial provider

(d) an NGO (e.g. a civil society or consumer organization) or aprivate nonprofit organization with a public goal (e.g. a charitable organization).

Modularity 2:

Profile transparency with regard to which OSN/browser?

(a) Facebook

(b) Twitter

(c) Facebook and Twitter

(d) Another OSN (e.g. Instagram)

(e) Chrome

(f) Firefox

(g) Chrome and Firefox

(h) Another browser

With regard to the first modularity we show how the DLA should be modulated if the provider of a transparency tool was (a) another scientific consortium, (b) an OSN, (c) a third-party commercial provider, (d) a NGO (e.g. a civil society or consumer organization) or a private nonprofit organization with a public goal (e.g. a charitable organization). We think that it is most likely that a profile transparency tool would be offered by an independent third party such as a commercial provider, a NGO or a nonprofit organization. If the transparency tool would be provided by an OSN the question would be if the OSN would give insight in the actual inferences made (and how we would know that this information was reliable), or whether it would offer a 'simulation'-tool (showing possible inferences instead of the actual ones) like *DataBait*. While it does not seem very likely that an OSN would provide a transparency tool like *DataBait*, we do not want to exclude the possibility that an OSN could be the provider.

With regard to the second modularity we explore how the DLA would look when the profiling transparency tool would apply to (a) Facebook, (b) Twitter, (c) both Facebook and Twitter, (d) another OSN, (e) a browser like Chrome or Firefox, or (f) another browser. The second modularity ('*Profile transparency with regard to which OSN?*') is relatively easy to incorporate: because the DataBait tool created by the USEMP project relates to data gathered from Facebook, Twitter, and a browser like Chrome or Firefox, this is simply a matter of removing any superfluous wording. The first modularity might require more adjustments to the DLA, notably with regard to the purpose of the processing, which will have to be extended or adapted.

In deliverable D3.7 we further adjust the DLA by adding an extra Article to also cover the licensing of content protected by intellectual property (IP) rights (notably copyright), such as certain types of pictures, videos and status updates. In that deliverable we also show how this particular article will have to be adjusted in the modular version of the DLA, depending on whether the service provider is commercial, non-profit or scientific (in the latter two cases some exceptions might apply and reproduction could be possible in some cases without infringing on the copyright protection of the content). However, in this deliverable we leave IP licensing aside and purely focus on the licensing of personal data (see section 5.2 of this deliverable for an explanation of the difference between IP licensing of copyright protected content and 'ordinary', non-IP, licensing of personal data).

Screen 1:

Data License Agreement

The parties:

(1) [……], user of the _____ platform and services, from hereon called 'You' and

(2) [_____], the data controller(s), from hereon called '_____' .

Hereby agree:

Screen 2:

(A) You will install the following tools, apps, plug-ins and/or graphicuser interfaces:_____ The _____ app and the _____ web browser plug-in will provide access to Your Facebook/Twitter/other OSN profile and Your browsing behaviour on Your device(s). These tools will be used by '_____' [*name of the the service provider-data controller*] to collect data that You share on Facebook/Twitter/other OSN as well as data collected by the web browser. This data can be data You posted (volunteered data), or data captured by the _____ tools (observed data). The latter concerns online behavioural data (storing what You did on the Internet and on Facebook/Twitter/other OSN).

This article defines the obligation to install the DataBait tools, which is pertinent for using any profile transparency tool of this type. It can be used in all modular versions.

Screen 3:

(B) You license the use of Your volunteered and observed personal data by the '_____' [*name of the service provider-data controller*] , as gathered by the the _____ app and the _____web browser plug-in for the purpose of _____ and – within that context – to provide You through the _____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 article, first, makes clear that this is a *quid pro quo* agreement, creating legal obligations on the side of the user (data subject) in the form of licensing the use of the data that will be processed by the data controller, and on the side of the service provider (data controller) in the form of providing a form of profile transparency. This can *quid pro quo* form (performance on both sides) can be used in all modular versions. However the specific purpose of processing will depend on the type of service provider and the rationale/business plan behind the service.

Screen 4:

(C) This license agreement confirms Your explicit consent to store the_____ tools on Your devices.

This article provides the consent required on the basis of art. 5.3 ePrivacy Directive for all and any tracking mechanisms to be stored on the user's (data subject's) device. This article can be part of all the modular versions of the DLA.

Screen 5:

(D) The '_____' [name of the the service provider-data controller] will use analytic software/do research [*cross out what is not applicable*] to predict what kind of information Facebook/Twitter 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 "_____GUI".

This article further explains the obligation on the side of the service provider, and the purpose of processing, highlighting that the profile transparency which will be provided is based on statistical inferences by others than OSN providers, meaning that the user is made aware of the fact that the data controller are not reverse engineering software code of the OSN provider and cannot in any way provide certainty about how one may be targeted. This article also ensures that the transparency is provided in a user-friendly manner. This article is crucial in every modular DLA for a profile transparency tool. However, the precise content of this article will have to be adapted according to the particularities of the service provided. Not every service provider of a transparency tool will have (scientific or commercial) research purposes. This will also have to be adapted according to the particularities of the service and the provided tool.

Screen 6:

(E) You agree to participate in surveys and/or focus groups, to enable '_____' [*name of the service provider-data controller*] 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 provided tools.

This article may be part of the modular version of the DLA depending on the particularities of the transparency tool and the services accompanying it.

Screen 7:

(F) 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.

Since consent is required for processing art. 8 DPD types of data, this article stipulates such consent. It highlights the intrusive nature of the processing of such data. It is part of every modular version of the DLA.

Screen 8:

(G) '_____' [*name of the service provider-data controller*] will treat all Your personal data, especially Your sensitive data, with care and delete or anonymize them as soon as possible.

See commentary below the next Article.

Screen 9:

(H) '_____' [*name of the service provider-data controller*] will process Your personal data in a secure way and not keep them any longer than necessary for the purposes described in Article D. 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 ____ months of the ending of the project all personal data will be either deleted, anonymised or processed for related scientific research. In the latter case '_____' [*name of the service provider-data controller*] will ask You for Your consent.

Articles G and H confirm the legal obligation for the service provider-data controller that the relevant data will be processed in accordance with the data minimisation principle, stipulating deletion or anonymisation as soon as possible (including a clear deadline) and security by design, while also explaining that to provide profile transparency the processing of both personal and sensitive personal data is necessary. These articles are part of all the modular versions of the DLA considering that this is a confirmation and reminder of the legal obligations of the service provider (data controller).

Screen 10:

(I) '_____' [*name of the service provider-data controller*] will not provide Your personal data to any third party other than _____. The transfer of the data will happen in a secure way and only in as far as strictly necessary for the purposes decribed in Article D.

This article is pivotal to ensure that data are not processed beyond the explicitly specified purpose. The article can be modulated depending on the specifics of the modular

version of the DLA, for instance allowing to share data with specified third parties and/or specified types of third parties.

Screen 11:

(J) 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:

To prevent any confusion about the applicable national law, and to accommodate the natural person whose personal data are being processed each modular version of the DLA has to confirm that the national law of the user (data subject) of the provided profile transparency tool is applicable. Under current EU Data Protection Law this seems the most apt, also for the any of the modular versions of the DLA. This may change under the proposed General Data Protection Regulation (pGDPR). Moreover, if the service would also be provided to non-EU residents, this clause would have to be adjusted.

5. Research strand 4: Granular licensing

What if data subject and controller could reach a mutual agreement on the re-use and on the sharing of the personal data? In this chapter we explore how *granular* licensing of personal data use could stimulate transparency and put the data subject and controller on more equal footing. The licenses we propose are granular in a double sense. Firstly, the licenses are granular because they offer an alternative to the 'all-or-nothing' consent ('Either you agree with this set of data processing modalities or you cannot use this service') which is often required when using services on today's internet. Granular licensing could be a way of offering data subjects a break down in different options of choice. Secondly, the licenses are granular because they have a layered format: a layer specifying the legal intricacies, a layer which can be easily grasped by a lay person and a machine readable layer.

5.1. Purpose limitation in OSNs and browsers

Any activity when using a browser, like Google's Chrome or Mozilla's Firefox, usually generates "personal data", i.e. data which are tied to an IP address or a search profile and thus relate to "an identified or identifiable natural person" (Art. 2(a) DPD 95/46). Even when the user would hide her IP address and search without an active search profile (e.g. not being logged into Google while browsing with Chrome), the generated data could still be personal data if they could be related to the user by other means, such as an analysis of the content (e.g. users often tend to search for themselves on the internet and combined with thematics of other searches it would be possible to make a very solid guess about the person behind a certain search history in the same session) or by using device fingerprints (about the machine used to perform the searches).

When a user uses an OSN, such as Facebook or Twitter, most of the generated data will also be personal data, because they are tied to her profile and are thus related to "an identified or identifiable natural person".

All personal data gathered by browsers or OSNs with an establishment within the EU (e.g. Google and Facebook are not only based in the US but also have European headquarters in Ireland) or which use equipment located within EU territory (if one interprets this in a broad manner one could argue that a cookie placed on a device located in the EU should be understood as the use of equipment within the EU), have to be processed in accordance with EU data protection law (Art. 4 DPD 95/46). The scope of protection GDPR will most likely[62] be even wider and apply to any processing of personal data of EU residents, even if the company processing the data is not located on EU territory and does not have any equipment there.

As explained earlier (see above, section 1.1), the strength of the protection offered by the purpose limitation principle (Art. 6 DPD 95/46) is diminished by three legal 'loopholes': (1) a weak interpretation of the meaning of "a specified, explicit and legitimate purpose" allowing

[62] We write 'most likely' because the Commission, Parliament and Council do not seem to disagree very much with regard to this point.

for long lists of (sometimes vaguely defined) purposes in terms of services and data policies, (2) a broad interpretation of the 'compatible purpose'-clause, and (3) selling of targeted advertisement space instead of the actual personal data. So let's take a closer look at how OSNs and browsers use these loopholes concretely.

5.1.1. Vaguely defined purposes

When signing up for a OSN, such as Facebook or Twitter, or a browser, like Google's Chrome or Mozilla's Firefox browser, a user consents to a policy with regard to personal data processing[63]. In accordance with EU data protection law in this privacy policy the company providing the browser or OSN service explains which types of data are processed, for what purpose and with whom the data are shared.

Let's take a look at Facebook as an example of how the purpose and the actors with whom data can be shared allow for a very wide range of processing actions. Facebook states that the purpose of the processing is "to help us provide and support our services" and specifies that this includes (a) providing, improving and developing services, (b) communicating with Facebook users, (c) showing and measuring ads and services, (d) promoting safety and security. Facebook shares your data with your chosen audience (e.g. your Facebook friends or anyone who is logged into Facebook). Next to that Facebook shares and combines the user data with data gathered by other companies owned by Facebook[64] : Facebook Payments Inc., Atlas, Instagram, Mobile Technologies Inc., Onavo, Parse, Moves, Oculus, LiveRail, and WhatsApp. If the ownership or control of Facebook would change, the privacy policy states that personal data might be transferred to the new owner. Third-party apps, websites or other services that the user uses may also receive information about what the user posts or shares. The privacy policy also allows Facebook to exchange and share data with "third-party partners and customers" for example, when they "jointly offer services or from an advertiser about your experiences or interactions with them".The policy specifies that "third-party partners and customers" comprises "advertising, measurement and analytics services" as well as "vendors, service providers, and other partners who globally support our business, such as providing technical infrastructure services, analyzing how our services are used, measuring the effectiveness of ads and services, providing customer service, facilitating payments, or conducting academic research and surveys", "advertising, measurement and analytics services". With regard to sharing data with "advertising, measurement and analytics services",

Facebook proudly states that they don't transfer personal data unless the user gives permission to do so. This can be explained by the fact that Facebook does primarily sell targeted advertising space based on personal data (and anonymized/pseudonymized feedback-data about the effectiveness of the targeting), and not the data themselves:

> "We […] use all of the information we have about you to show you relevant ads. We do not share information that personally identifies you […] with advertising, measurement or analytics partners unless you give us permission. We may provide

[63] https://www.facebook.com/policy.php; https://twitter.com/privacy?lang=en ; http://www.google.com/policies/privacy/ ; https://www.mozilla.org/en-US/privacy/
[64] https://www.facebook.com/help/111814505650678

> these partners with information about the reach and effectiveness of their advertising without providing information that personally identifies you, or if we have aggregated the information so that it does not personally identify you. For example, we may tell an advertiser how its ads performed, or how many people viewed their ads or installed an app after seeing an ad, or provide non-personally identifying demographic information (such as 25 year old female, in Madrid, who likes software engineering) to these partners to help them understand their audience or customers"[65]

The privacy policy of Facebook thus allows the processing of user data for a wide array of purposes and makes it possible to share them with many actors. While there are some differences between the privacy policies of Facebook and another OSN like Twitter or Instagram – particularly in their wording and style of communicating with the user – they all formulate a rather broad purpose and a long list of possible recipients.

Twitter states that the main purpose of their data processing is :

"to provide our Services and to measure and improve them over time".

Currently the actual use of personal data by OSNs, like Facebook, Twitter or Instagram, and browsers, like Chrome and Firefox, differs. Browser vendors usually use the data they gather only for analytics and troubleshooting purpose and they don't target users for advertisements based on thse data. This may however change in the near future.

The privacy policy of Mozilla (Firefox), known for being rather good in complying with privacy and data protection requirements in comparison to other browsers like Chrome or Internet Explorer, reads :

> "When you give us information, we will use it in the ways for which you've given us permission. Generally, we use your information to help us provide and improve our products and services for you."

This formulation leaves not so much space to divert from the current practice of only using browsing data for performance improvement of the web browser itself. In contrast, Google's browser (Chrome) gives a very long description of the various processing purposes, from which we only reproduce some salient sections:

> "We use the information we collect from all of our services to provide, maintain, protect and improve them, to develop new ones, and to protect Google and our users. We also use this information to offer you tailored content – like giving you more relevant search results and ads. […] We use information collected from cookies and other technologies, like pixel tags, to improve your user experience and the overall quality of our services. One of the products we use to do this on our own services is Google Analytics. […] We may combine personal information from one service with information, including personal information, from other Google services "

[65] https://www.facebook.com/policy.php

Notwithstanding the differences between all of the aforementioned service providers, Twitter, Google and Firefox all have, like Facebook, a resale-clause, stating that data might be transferred to another owner in case the company is sold, merged or restructured. This nicely exemplifies that each of these companies to a bigger or lesser extent formulates a broader scope for their processing purposes than many end-users realize.

While we argue that the use of the aforementioned 'loopholes' should be critically looked at from the perspective of EU data protection (see also: Wauters, Lievens, & Valcke, 2014), we would like first to make an even more basic observation about the way in which the purpose of the processing tends to be formulated by data controllers in general: in an idiosyncratic, one-sided, take-it-or-leave, way. If an end-user wants to use a particular service, she will have to consent with the processing purpose set by the data controller (as long as the purpose is legitimate, specified and explicit). This observation might sound extremely naïve: after all, the role of the data controller is defined as the one determining "the purposes and means of the processing of personal data" (Art. 2(d) DPD 95/46) – so, why shouldn't the controller determine these purposes and means in a one-sided, idiosyncratic way? Yet, a data controller cannot simply establish whatever processing purpose she thinks is commercially interesting, as long as she specifies it in a specific and explicit way: the purpose also has to be *legitimate* (Art. 6(1)(b) DPD 95/46.

5.1.2. Legitimacy of the purpose-

What does it mean that the principle of purpose specification (Art. 6(1)(b) DPD 95/46) includes the requirement that the processing purpose should be legitimate? Let's imagine, for example, that a data controller (e.g. an OSN) has received the explicit and informed consent of the data subject (in accordance with Art. 7(a) DPD 95/46) after having informed the data subject in a very specific, detailed and explicit manner about the fact that her gathered data will be used for targeted advertisements based on race or that her behavioral data will be used for price differentiations for health insurances. Is this processing in accordance with the principle of purpose specification? Most likely not: after all, the first sentence of Art. 6(1)(b) DPD 95/46 requires that the processing purpose should not merely be specific and explicitely communicated, but also legitimate. The reason why we stress the requirement that personal data must be collected for legitimate purpose is that it is an often overlooked, or too narrowly understood, element of the purpose specification principle. The fact that processing is in accordance with at least one of the six criteria which make data processing legitimate (See Art.7 DPD 95/46, e.g. based consent or, following a balance of interest test, necessary for the purposes of the legitimate interests of the data controller) does not exhaust the requirement of the legitimacy of the purpose of Art. 6(1)(b):

> "[…] Article 6(1)(b) also requires that the purposes must be in accordance with all provisions of applicable data protection law, as well as other applicable laws such as employment law, contract law, consumer protection law, and so on. The requirement of legitimacy means that the purposes must be 'in accordance with the law' in the broadest sense. This includes all forms of written and common law, primary and secondary legislation, municipal decrees, judicial precedents, constitutional principles, fundamental rights, other legal principles, as well as jurisprudence, as such 'law' would be interpreted and taken into account by competent courts. Within the confines of law, other elements such as customs, codes of conduct, codes of ethics, contractual arrangements, and the general context and

> facts of the case, may also be considered when determining whether a particular purpose is legitimate. This will include the nature of the underlying relationship between the controller and the data subjects, whether it be commercial or otherwise. The legitimacy of a given purpose can also change over time, depending on scientific and technological developments, and changes in society and cultural attitudes. (Article 29 Data Protection Working Party 29, 2013, p. 19-20)

Following the Opinion of WP29 we argue that one should assume that the requirements that her processing should be fair and legitimate (Art. 6(a) DPD 95/46) and that the purpose should be legitimate (Art. 6(1)(b) DPD 95/46) also imply a proportionality test in the vein of Art. 8(2) ECHR (respect for private life). This means that the data controller is constrained in what she can request of the data subject to accept as a purpose of the processing, because, following Art.6(1)(b)DPD 95/46 the purpose has to be in accordance with *any* applicable laws, and which will include a balance of interests. Moreover, any processing which does not really contribute to the purpose of the processing could be considered as disproportionate and excessive, and consequently as illegitimate.

The legitimacy of the purpose is a criterion will become increasingly important for data controllers, as judges take a more active stance (see, for example, the much debated judgement on the "right to be forgotten"[66]) and as fines for non-compliance with data protection law will significantly increase under the new legislation (pGDPR)

5.1.3. Illegitimacy of any excessive processing beyond the purpose-

We believe that some of the problems with regard to the legal uncertainty surrounding purpose limitation could be solved that if data controllers could refine the purpose in mutual agreement with their end-users (not through "consent" to a standard, not individually negotiated contract but, for example, through a mutual agreement in which the end-user would have several options of licensing her data for particular options for sharing and targeting) and in a more standardized and transparent way (e.g. "You license us to use your data according to standardized license X"), this could also help in channeling the ways in which the aforementioned legal 'loopholes' are used to be in better accordance with the original rationale of the purpose limitation principle. This could also be beneficial for data controllers: data subjects who are co-involved in deciding on the purpose of the processing through a granual licensing system are likely to be more motivated to ensure that their data are accurate. An empowered data subject, who is a party to a mutal agreement with regard to how her data is used and who has active knowledge of the processing of her data, will feel a bigger involvement in the processing of her data than a data subject who has consented in a take-it-or-leave-it manner and is kept in the dark about the personal data economy taking place 'through the looking glass'. If data subjects were more actively engaged in caring for the quality of their data, data controllers would be able to base their marketing, targeting,

[66] *Google Spain v AEPD and Mario Costeja Gonzalez*, CJEU 14 May 2014, C-131/12.

personalization and risk assessment decisions on more accurate and up-to-date information. A granular licensing system would thus benefit both data subjects and controllers.

In order to clarify our proposal with regard to granular licensing of personal data, we will draw two analogies. The first analogy is with the way in which WP29 (Article 29 Data Protection Working Party, 2012) proposes to understand the requirement of the cook ie consent following Art. 5(3) from e-Privacy Directive 2002/58/EC. The second analogy is with the way *Creative Commons* licenses[67] for copyright protected works function.

5.2. Lessons from cookie consent.

Since the EU adopted new cookie legislation in the form of Art. 5(3) of the e-Privacy Directive (Directive 2009/136, amending Directive 2002/58/EC), making it mandatory for websites to request the prior consent before placing cookies on a users' computer or reading them back, internet users who want to access a website often have had to make the empty gesture of accepting cookies as a precondition for accessing the website. As WP29 has shown (Article 29 Data Protection Working Party, 2012) such a binary request for cookie consent ("either you accept all cookies, or access is denied to you") twarths the rationale of Art. 5(3) of the e-Privacy Directive, which builds on a distinction between functional cookies (for which no prior consent is required) and non-functional cookies (for which prior consent should be requested).

> "Member States shall ensure that the use of electronic communications networks to store information or to gain access to information stored in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned is provided with clear and comprehensive information in accordance with Directive 95/46/EC, inter alia about the purposes of the processing, and is offered the right to refuse such processing by the data controller. This shall not prevent any technical storage or access for the sole purpose of carrying out or facilitating the transmission of a communication over an electronic communications network, or as strictly necessary in order to provide an information society service explicitly requested by the subscriber or user." (Art. 5(3) of the e-Privacy Directive 2002/58/EC)

Art. 5(3) of the e-Privacy Directive does not intend to create an unnecessary extra hurdle for internet users, but sets out to give internet users a real choice in refusing cookies which are not necessary for the performance of the service requested by the internet user (such as tracking cookies for a service which does not have tracking as its core functionality). There are between two types of functional cookies:

1. technical cookies that have "the sole purpose of carrying out or facilitating the transmission of a communication over an electronic communications network

2. functional service cookies that are "strictly necessary in order to provide an information society service explicitly requested by the subscriber or user".

[67] https://creativecommons.org/licenses/?lang=eng

The distinction between functional and non-functional cookies could be used in an analogous way to make a distinguish between "functional" processing of data, i.e. necessary for the processing purpose of a service, and additional processing, i.e. beyond the strict processing purpose (e.g. any sharing of data with other parties or offering third parties the possibility to target users based on their profile, in as far as this does not strictly contribute to the service requested by the user).

A granular license clause with regard to personal data could offer the data subject an option to “negotiate” or “choose”, within the limits of the law and the broader boundaries of the purpose as defined by the data controller, how and for what purposes her data can be used. It should be underlined that the outer boundaries of this choice will always be delineated by the specific, explicit and legitimate purpose set out by the data controller. It would make no sense if an OSN, like Twitter or Facebook, would state a processing purpose (e.g. “*Delivering you a service that allows you to do X, Y, and Z, making profit based on selling targeted advertisements space and sharing your data with affiliate businesses A, B and C to improve our market position in comparison with competitors*”) and the data subject could go *beyond* these limits (e.g. “*I would like my data to be processed for the purpose of achieving peace in the Middle East*”). However, within the boundaries of am specific, explicit and legitimate purpose there might be room for “choice” or “negotiation”, which could (as explained in sections 5.3 and 5.4) benefit both the data subject (end-user) and the data controller (service provider).

When a data controller defines her purpose, there is, firstly,the processing purpose as strictly necessary for the performance of the service requested by the user. Here there is no room for negotiation or choice for the data subject; it is the data controller defining the use of the data. However, the controller can also define additional purposes, not strictly necessary for the performance of the service requested by the user, which involves *sharing* user data with third parties or allowing the data controller and/or third parties to *target* users with a certain profile. It is in these three latter categories (i.e., sharing data with third parties, targeting by data controller, targeting by third parties) that we believe there is room for ‘choice’ and ‘negotiation’; and, consequently for the granular licensing of personal data which we propose in this chapter.

Building on the analogy with the differentiation between functional and non-functional cookies, one of the choices offered to the data subject could be to exclude any processing which is not strictly necessary for the service, or to offer an even more fine-grained system in which the user can exclude some of the not strictly necessary processing actions while preserving others. The analogy with the cookie legislation evokes several difficult yet interesting questions. For example: How to decide what processing is strictly necessary for a service? An OSN might argue that the targeted advertisement space is part and parcel of its business model and thus strictly necessary. Or what about a data controller using browsing data, gathered from the browsing service offered by her, to ameliorate another service she offers (e.g. an email service)? Can a data subject exclude such *re-use* for the same purpose in a different context? What about *sharing* data for the same or a compatible purpose with a third party? For example, a browsing service sharing data with another company, in order to ameliorate the email service of this other data controller? Targeting – can that be excluded? Which methods can the data controller legitimately use to nudge the data subject to accept more than what is strictly necessary? Is the granular license which excludes all non-strictly processing something which each data controller should offer compulsary, or should we

conceive it as a voluntary option to offer the data subject such opt-out option? What effect would it have if Art. 7(4) of the pGDPR (EU parliament version) would come in to force?

> "Consent shall be purpose-limited and shall lose its validity when the purpose ceases to exist or as soon as the processing of personal data is no longer necessary for carrying out the purpose for which they were originally collected. *The execution of a contract or the provision of a service shall not be made conditional on the consent to the processing of data that is not necessary for the execution of the contract or the provision of the service pursuant to Article 6(1), point (b)."* (Art. 7(4) of the pGPDR, *italics ours*)

The second half of Art. 7(4) shows how the Parliament version of the proposed legislation prohibits that data subjects can be forced to consent in to data sharing which is not necessary for the service. If this Article is adopted in the final version of the GDPR, granular licensing will become an even more attractive option, because data controllers will not be able to force data subjects to share more than necessary. This situation would make the following question very poignant: Can a data controller ask for a financial remuneration for the use of her service if the data subject excludes all the processing which is not strictly necessary for the functioning requested service? In this way the choice is made less binary ("accept additional processing or we will deny you this service") – but is it desirable that each service would have a paid, more privacy-friendly verson? How can one avoid irritation in the end-user about the fact that she is burdened with one more choice - who likes a screen that is cluttered with questions about which kind of cookies, types of sharing data and targeting is permitted? In the final version of this deliverable (D3.10) we elaborate more extensively on such questions following from the analogy between cookie consent and granular licensing for personal data.

To get even more of a sense of how a system of standardized licensing options could be realized we turn in the next section to our second analogy, namely with Creative Commons licenses.

5.3. Lessons from Creative Commons licensing.

The word 'licensing' is strongly associated with intellectual property, such as licensing copyright protected content. However, the majority of personal data does not qualify as copyright protected content. The subject matter protected under copyright is not uniformly defined[68], but one can broadly say that in order to be a copyrightable, the subject matter should be "original" or the author's own "intellectual creation"[69] and reflect the author's personality[70]. More specifically, this is the case if the author was able to express her creative abilities in the production of the work by making free and creative choices[71]. A search string entered in a browser's search engine or an URL typed into the browser will thus definitely not be copyrightable; neither will a factual status update like "sunny weather in Amsterdam" be.

[68] Copyright is granted at the national level and is regulated in national laws but many harmonisation efforts have been made at the international and European levels.
[69] Judgment in *Infopaq International A/S v Danske Dagblades Forening*, C-5/08, ECLI:EU:C:2009:465, para. 37.
[70] Recital 17 in the preamble to Council Directive 93/98/EEC of 29 October 1993 harmonizing the term of protection of copyright and certain related rights , *O.J. L 290 , 24/11/1993 P. 0009 – 0013 ;*
[71] Judgment in *Eva-Maria Painer v Standard VerlagsGmbH and Others*, C-145/10, ECLI:EU:C:2011:798, para. 89.

However, some of the data posted on an OSN - pictures, videos and posts expressing an element of originality and authorship - will not only be personal data, but also qualify as copyright protected content.

In order to be able to freely use (and reproduce) all data generated on their website, an OSN like Facebook requires users not only to consent with their personal data policy, but also *license* them for all content protected by intellectual property law:

> "For content that is covered by intellectual property rights, like photos and videos (IP content), you specifically give us the following permission, subject to your privacy and application settings: you grant us a non-exclusive, transferable, sub-licensable, royalty-free, worldwide license to use any IP content that you post on or in connection with Facebook (IP License). This IP License ends when you delete your IP content or your account unless your content has been shared with others, and they have not deleted it."[72]

We discuss this license in more detail in D3.8 (with regard to OSN's solutions to user's claims of copyright protection). However, it is important to underline that the licensing system we propose would not only cover IP content ('IP licensing') but the licensing of any personal data. To 'license' basically just means "to give permission to"; it means "granting (a person) a licence or authoritative permission to hold a certain status or to do certain things"[73]. Consequently, the granular Personal Data Licensing (gPDL) approach, which is inspired by Creative Commons licenses (which are IP licenses, not licenses with regard to other information), does itself not have anything to do with IP licensing or copyright protection as such[74].

What we take from the Creative Commons licensing system and what we transfer to our granular licensing system of personal data is (a) the non-dogmatic, problem-oriented and pragmatic approach in creating licenses, (b) a comprehensive, standardized and transparent set of licenses specifying how a piece of data can be used, (c) using a layered format for the licences (a layer specifying the legal intricacies, a layer which can be easily grasped by a lay person and a machine readable layer). Let us elaborate on these three aspects.

1.

A first aspect we adopt in our granular personal data licensing (gPDL) system is the non-dogmatic, problem-oriented and pragmatic approach of the Creative Commons movement. Creative Commons (CC) licenses where not developed as from an ivory tower but in direct response to a concrete (societal) *problem*, namely how to stimulate the free circulation of content while preserving the system of copyright protection, and the need to come up with a solution to this problem. The six CC license types which currently exist have been created over time, in a bottom-up way. These CC licenses have resulted in a growing

[72] https://www.facebook.com/legal/terms
[73] See Oxford English Dictionary, http://www.oed.com/
[74] Moreover, it should be noted that CC licenses work worldwide, because copyright relies on old international treaties, which have harmonised important parts of copyright protection worldwide; this is translated in CC licences that are reviewed and adapted per national legal order. Because data protection is not as universally harmonized, the gPDL approach will be especially useful when the processing is limited to the EU. However, as a "good practice" it could also spread beyond the EU.

"pool of content that can be copied, distributed, edited, remixed, and built upon, all within the boundaries of copyright law"[75], thus stimulating the dissemination of knowledge and creativity. Thus, if we use with Creative Commons as an analogy, the first step towards developing a gPDL system has to be based in defining the concrete need or problem we are trying to address. However, an important *problem* which gPDL tries to solve is rather opposite in nature to CC: how to restrict the uncontrolled and opaque diffusion of personal data on the internet. In this sense the goal is to come to what we can tentatively call a "personal anti-commons" (PAC). In the light of the failure of the purpose limitation principle in this regard, one way to address this problem is to create a system where the way in which personal data are used is a matter of mutual agreement between data controller and data subject, and less of an idiosyncratic, one-sided offer to which the data subject can either consent or not. The lack of such a situation is the *need* which our gPDL addresses.

2.

A second aspect we adopt in our gPDL system is a comprehensive, standardized and transparent set of licenses specifying how a piece of personal data can be used. After all, it is not realistic to expect data subjects in their dealings with big internet companies to enter into individual negotiations. For example, an individual who is signing up to Facebook and who decides, upon critical investigation of the terms of service, that she is ok with Facebook combining her Facebook data with her Whatsapp data but that she would rather not have them combine them with her Instagram data (and in return, that she would be willing to abstain from some of the Facebook functionality if her wishes were to be granted) will not stand a chance. A company like Facebook does not gain anything from negotiating terms with users on an individual basis – it would merely create an enormous work load. Moreover, most individual users will lack motivation or knowledge to negotiate individual terms.

However, if a limited set of standardized and openly scrutinized personal data licenses existed, users (gaining empowerment and transparency about what happens to their data) and industry (gaining transparency about what is allowed with data and possibly more accurate data due through user engagement) might be interested. Industry could try to seduce the user to grant them a very permissive license (e.g. by offering additional services and functionalities), but at least the standardization of the licenses would offer the user more transparency about the 'cost' of this deal in terms of personal data usage and lead to better informed decisions. Even more safeguards for transparency could be offered if these licenses would be provided by an independent platform, informing users of the 'costs' in terms of privacy and data protection of providing very broad licenses.

So what kind of standardized licenses would be useful? When we have a look at the CC approach to licensing, we see that the solutions to IP problems are reduced to a few basic conditions, which we can call the "license atoms", and the combinations of these. Currently there are six CC license types, which build on three basic distinctions. The first two distinctions are commercial versus non-commercial and derivative versus non-derivative (i.e. no modification of the original work) use. These two distinctions have even been integrated in Google image search where you can filter your results according to the following five 'usage

[75] https://creativecommons.org/licenses/?lang=eng

rights' types: (a) not filtered by license, (b) labeled for reuse, (c) labeled for commercial reuse, (d) labeled for reuse with modification, (e) labeled for commercial reuse with modification. The third distinction underlying CC licenses is whether or not derivative works, based on works licensed under a CC license, have to be licensed under the same licensing conditions as the original work. This is called "share alike". Furthermore, each of the six CC licenses has the basic condition of attribution ("CC BY"), which means that you mention the author of the original creation. This results in the following six licenses[76], ordered from the most accommodating license (the "*CC BY*" license: "lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you [the author] for the original creation") to the most restrictive one (the "*CC BY-NC-ND*": "only allowing others to download your works and share them with others as long as they credit you [the author], but they can't change them in any way or use them commercially"):

Attribution

CC BY

Attribution-ShareAlike

CC BY-SA

Attribution-NoDerivs

CC BY-ND

Attribution-NonCommercial

CC BY-NC

Attribution-NonCommercial-ShareAlike

CC BY-NC-SA

Attribution-NonCommercial-NoDerivs

CC BY-NC-ND

[76] License images reproduced from: https://creativecommons.org/licenses/?lang=eng

What kind of default licenses (i.e, pre-set licensing settings) would make sense for mutual agreements with regard to re-use or sharing of personal data, based on the initially specified purpose or one that is compatible? Before proposing a set of gPDLstandardized licenses it is good to note that the CC licenses have been created over time, in a bottom-up way. Our proposal should therefore also be seen as a first step in a long debate. A second caveat is that the licensing would have to stay within the limits of the purpose set out by the data controller and of EU data protection law. For example, it would not be possible for a data controller and data subject to come to the agreement that the data subject will give an open license allowing data use for an unspecified set of purposes, because this would be at odds with the purpose limitation principle (Art. 6 DPD 95/46). We think that the gPDL license types could be developed according to lines described in section 5.2: that is, that the data subject should have the possibility to exclude any processing which is not strictly necessary for the requested service. We distinguished: sharing data with third parties, targeting by data controller, targeting by third parties. In the final version of this deliverable (D3.10) we also explore the possibility for further fine-grained divisions along the following lines:

- *Thematic*: This license type would give users the possibility to license the sharing of their personal data or becoming the object of targeting according to themes: e.g. for medical research, commercial personalization, scientific research, consumer products, etc. This license type could have the format of a list of subjects with tick boxes: e.g. one could tick ‘medical science’ and ‘market research in consumer products’ while leaving ‘market research in consumer services such as insurances’ unchecked. The drawback of thematic licensing might be that it could become quite complex, offering too many fine-grained options, and that it could cause many qualification problems (e.g., is market research into fitness habits scientific, medical research?).
- *Institutional*: This license type would give the possibility to users to license sharing of their personal data or becoming the object of targeting according to the type of institution: e.g. license the use of data processing for NGO, non-profit, commercial, governmental, etc. An obvious distinction could be the one that is made by CC licenses: commercial and non-commercial use. It should be noted that with regard to CC licenses what should be qualified as “non-profit” and what as “commercial use” has caused quite some controversies. For example, when an NGO with a charitable purpose uses copyright protected content in a campaign – is this “commercial” use? Most likely this would qualify as commercial use but one could imagine many boundary cases.
- *According to the nature of the data type*: This license type would give the possibility to users to license the processing of their personal data according to the nature of the data e.g. no sharing or argeting involving sensitive data in the sense of Art. 8 DPD 95/46 or no processing of certain data types based on personal preferences in terms of privacy, e.g. data related to income or social status. See annex 7.4 for an alternative classification of data types.
- *Action based licensing*: This license type would give the possibility to users to license the sharing of their personal data or becoming the object of targeting according to the kind of processing actions that can be performed with the data, e.g. no personal inquiries, but only categorical applications such as in targeted ads; or: no profiling, only raw data use. The latter could correspond to what in CC licenses is called

'derivative' and 'non-derivative', though this would have a different meaning (namely whether the license allows for inferences based on the raw data)

- *Quantitative limitations on re-use by the data controller or other actors:* Any sharing of personal data or becoming the object of targeting would have to be in accordance with the purpose limitation principle: i.e. the purpose of the re-use would have to be compatible with the original purpose for which the data was collected, would have to be reasonably foreseeable, assessment of consequences of re-use, etc. (Article 29 Data Protection Working Party 29, 2013). However, within the boundaries of the purpose limitation principle, further 'quantitative' limits could be set, e.g. that the re-use is only allowed for a set period of time or that re-use by other actors is only allowed within a limited 'degree of separation' (e.g. maximum two transfers of the personal data) allowed.
- *Integrity based licensing: re-use by actors with a certain certificate of trustworthiness, 'green label',* etc.: This license type would give the possibility to users to license the the sharing of their personal data or becoming the object of targeting according to certain certificates of trustworthiness, 'green labels', etc. This could also relate to how the actor responds to requests from police, secret services, etc. (does the company simply comply or have a lawyer look at the request?), working conditions in a company (e.g. only companies with fair working conditions), etc.

3.

A third aspect we propose in our gPDL system is a layered format for the licences: a layer specifying the legal intricacies, a layer which can be easily grasped by a layperson and a machine readable layer. This is what would make the license *granular* in a second sense[77], that is, operating on three levels of details and with three different 'audiences' in mind. When we enumerated the six CC licenses (see above) we only reproduced the human readable layer: an iconic depiction and a set of abbreviations.

The gPDL system would look very much like the CC licensing system. Like in the CC licensing system, each license would consist of a layer of "legal code", of human readable icons and text and machine-readable tags. The layer of « legal code » is pretty straightforward – it is the traditional legal formulation of a license. The human and machine readable layers require more creative thinking. The human readable layer is the "common sense" layer, consisting of a user friendly interface of human understandable *text* and *icons* (which could actually be considered as two sublayers). Here legal options are translated and communicated to the user in an understandable way. This also correlates with a series of transparency obligations introduced in the new General Data Protection Regulation, on the data controller for providing transparent information and communication and for effective procedures and mechanisms. First, the controller needs to offer "transparent and easily accessible policies" for the exercise of the data subject's rights, and for the processing of personal data in general. Furthermore any information and communication with regard to the processing of personal data has to be provided "in an intelligible form, using clear and plain language, adapted to the data subject, in particular for any information addressed specifically

[77] The first granularity is the break-down in different options for data handling by the data controller, instead of 'all-or-nothing' consent.

to a child' (article 11.1-2 GDPR).[78] Here social sciences like media studies can be very relevant giving shape to this "accessibility" and "intelligibility"of these communications.[79]

Then there is the machine readable.layer[80] : this is the technological layer. Wouldn't it be great if sets of personal data or individual pieces of personal data would be tagged in such a way that one could easily filter personal data according to the kind of permitted uses ? In the case of content licensed with a CC license there is a kind of digital marking of copyrightable works with the content usage policy, travelling with it. This could also be done with regard to personal information, though the tagging would probably be easier at the level of a set of data than tagging each individual data piece. The OSN would have to implement this, since they are often the one's creating the data in the first place, or facilitating their creation. Within the USEMP project we are discussing concrete possibilities with the technical partners.

Because certificates of trustworthiness or eco-friendliness are not very uniform, and thematic licenses might be too fine-grained, we propose a set of basic combinations of the institutional, 'nature of the data'-based, action-based and quantitative licenses. The first licensing atom (institutional) would be the distinction between a commercial and a non-profit license. The second licensing atom (nature of the data) would be whether the license excluding sensitive (Art. 8) personal data from the license. The third licensing atom (action based) would either (a) merely license the processing of raw data without any further inferences, or (b) processing for general inferences (statistics offering global insights), or (c) personalization-targeting. The fourth licensing atom would limit compatible re-use further by a set period of time or amount of actors.

It goes beyond the scope of the DataBait tool to give the end-user the possibility to license their personal data through a granular licensing system (USEMP is not a databroker, nor do we have any power over how an OSN like Facebook uses personal data, so all of this would have to be a simulation, which might confuse the DataBait user – and this would be misplaced in a profile *transparency* tool!). However, we are currently discussing with our USEMP partners if DataBait could contain a tool which would offer a "menu" to the DataBait user, from which she could choose what her "ideal" granular license would look like. After the user has provided her desired settings, the DataBait tool could provide feedback about how these desired licensing settings correspond with the data policy of the OSNs and browsers she is using.

.

[78] See the discussion in Heyman, R., van Dijk, N., Who can Afford Users as Targets? Interfaces, Transparency and the Commodification of Relations in Online Social Networks, Report on differences between user and legal perspectives on privacy and profiling (D3.3.1). EMSOC Project. 2013, at: http://emsoc.be/wp-content/uploads/2013/11/D322-SMIT-and-LSTS.pdf

[79] One possibility is to experiment with the idea of *mental models* of privacy and data protection in this case. See Camp, L. J., (2009) Mental Models of Privacy and Security, IEEE Technology and Society Magazine.

[80] See also the W3C work done in the past as part of P3P initiative http://www.w3.org/P3P/

5.4. gPDL as contractual clauses

A big difference between gPDL and CC licenses is that the latter are not directed to one particular user of the license but to *anyone* who wants to use the copyrighted content. Granular personal data licenses could be part of the service contracts between data controllers (aka service providers) and data subjects (aka end users of a service). This means that a data subject who chooses very restrictive gPDL settings might simply be denied to use a service. The purpose and ways of the processing personal data are determined by the data controller; a gPDL system could make the controller negotiate with the user, but the licensing of personal data cannot be a one-sided affair of the subject of those data (the data subject).

In a gPDL system it takes two to tango. A user will not always be able to oversee which settings are optimal when signing up for a service. An optimal scenario would be that a service contract would allow a user to adjust the granular personal data licensing settings – within certain limits- during the use of the service. Software could be used to aid the user: automated suggestions for licensing settings could be made (e.g. based on settings of friends in a OSN) and a clear graphic interface could help the user with picking the desired settings. Within the USEMP project we explore software possibilities in collaboration with the technical partners. Clearly, this adjustable software solution poses problems in terms of legal certainty and foreseeability: how can a company process data if a user continuously changes the license settings. This is an issue deserving further investigation. One possible problem following from the fact that the licensing of personal data would be part of a contract between a commercial internet business and an end-user, is that 'cunning' lawyers could undo the transparency of the licensing system by surrounding it with complex clauses which weaken the position of the end-user.

However, it should be noted that the possibility of contractual 'trickery' to disadvantage the end-user is legally limited by considerations of fairness with regard to the content of contractual clauses. While contractual freedom is pivotal for the flourishing of any economic market, this contractual freedom is not unlimited: unfair contract terms can be detrimental for a healthy functioning of the market and are thus curtailed by national and EU legislation. Directive 93/13 EC on unfair terms in consumer contracts[81] posits in Art. 3(1) that a clause which causes a 'significant imbalance' between a seller or supplier and a consumer should be considered as unfair:

> "A contractual term which has not been individually negotiated shall be regarded as unfair if, contrary to the requirement of good faith, it causes a significant imbalance in the parties' rights and obligations arising under the contract, to the detriment of the consumer."

Further limitations to contractual clauses can be found in consumer rights law.[82] The fact that a service, such as the use of a browser or an OSN, is rendered for 'free', that is

[81] See in particular: Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts, OJ L 95, 21.4.1993, p. 29–34 Online available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31993L0013

[82] See in particular: European Parliament and Council Directive (EU) 2011/83 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council [2011] OJ L304/64 (Consumer Rights Directive).

without direct financial remuneration from the user, does not change the fact that the business model of most major OSNs and browsers is commercial and money is being made. As argued by Wauters e.a. (2014), this commerciality brings along that the terms and conditions under which this service is rendered should be as compliant with consumer regulation as any other commercial service. In D3.8, in relation to the IP license which end-users have to provide to an OSN in relations to the copyright protected content they produce, we elaborate further on which contractual clauses should be considered unfair. In this respect it is also interesting to note that in the Parliament version of the proposed GDPR[83] the 'significant imbalance' criterion from EU consumer law is incorporated in Art.7(4) with regard to consent:

> "Consent shall not provide a legal basis for the processing, where there is a significant imbalance between the position of the data subject and the controller be purpose-limited and shall lose its validity when the purpose ceases to exist or as soon as the processing of personal data is no longer necessary for carrying out the purpose for which they were originally collected."
>
> Moreover the second sentence of Art. 7(4) specifies :
>
> "The execution of a contract or the provision of a service shall not be made conditional on the consent to the processing of data that is not necessary for the execution of the contract or the provision of the service […]"

Obviously, industry has not been very pleased with this clause, as it would make it impossible to use the provision of a service as a leverage to make users consent with data processing which is unrelated to the provided service. It is impossible to predict the fate of Art.7(4) in the final version of the proposed GDPR, but if this clause was to survive in the negations between Council and Parliament, the possibilities of contractual 'trickery' into data processing for marketing and other commercial services "not necessary for the execution of the contract or the provision of the service" would be significantly limited.

Next to the protection offered against unfair contractual clauses by consumer law and, possibly, data protection law, Art. 8 ECHR (right to respect for one's private life) can also function as a buffer of fairness. In D3.8 we explore how personality rights, particularly the protection of one's portrait ("portrait right"), deduced from Art. 8 ECHR could supplement the DLA licensing structure. Art. 8 ECHR could function as a life-vest in situations where 'cunning', 'wicket' lawyers might make a contract where users "sign away" their rights through licenses and only later realize the consequences. The portrait aka personality right would be something that makes sure there is a core of the intimate sphere (portrait right protection) which cannot be contracted away.

Moreover, gPDL would not necessarily have to be something which just involves a service provider and an end-user : sometimes it takes *three* to tango. The granular personal data licences could be mediated through an independent body (see research strand three on the modular DLA : independent bodies could, for example, be a third-party commercial provider, an NGO, a private nonprofit organization with a public goal, or a scientific

83 https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-07-27_GDPR_Recommendations_Annex_EN.pdf

consortium). An independent third party, acting like an intermediary or very special type of ‘data broker’ (whose purpose is to support EU end-users in the exercise of the data protection rights), could leverage the field between service providers and end-users[84].

[84] In previous work from W3C (P3P) the user agent (browser) would enforce a policy dictated by the the user by filtering out data but this was meant for web browsing data NOT data explicily shared via OSNs. See: P3P initiative http://wwww3.org/P3P/

6. Conclusion – legal requirements based on this deliverable

In this deliverable we followed four research strands.

The first research strands looked at how *technological and organizational transparency* can be provided and supported for data processing in general, and for profiling in particular. The legal analysis resulted in the requirement that the DataBait tool should support the user in exercising her informational rights by providing her insight in her own raw data (what information do I share? who is tracking me?) and by showing what can be derived from these data. In terms of compliance of the DataBait tool itself with data protection law, the analysis resulted in a list of information provided in the "DataBait: what, how, why?" section of the tool.

The second strand gave *a legal clarification* of which data should be considered *'sensitive'* in the sense of Art. 8 DPD 95/46, and which data can be considered *anonymous* (i.e., not personal data and therefore outside the scope of DPD 95/46). We concluded that in assessing whether data are truly anonymous their potential for de-anonymization should be taken into account. Similarly, in assessing whether data are to be considered sensitive, the potential to extract sensitive data should be taken into account. We propose that in non-obvious cases the crucial notion in assessing whether this potentiality is relevant for the legal qualification 'sensitive' is the *intended use* (realistic possibility and significant chance that sensitive information will be extracted and used, given the concrete circumstances). The DataBait user should be given realistic examples of the intended uses for which the informaton could be used. Also the protected grounds from EU anti-discrimination law should be used in designing the informational functionalities of the DataBait tool.

The third strand presents the Data Licensing Agreement (DLA), which offers an alternative to the 'take-it-or-leave-it-approach' of consent as a legal ground for processing personal data by engaging the data subject in the process of profiling. Before using DataBait each user should sign the DLA. The DLA creates more of a level playing field between data controller (service provider) and data subject (end user) by means of an obligatory agreement that entails clear and mutual quid quo pro, while providing transparency about all the relevant legal issues when using *DataBait*.

The fourth strand looked at granular licensing of personal data. We explored how the distinction between functional and non-functional cookies and the format and functioning of Creative Commons licenses could inform a contractual *granular licensing system*. This could be a form of DPbD based on the requirements of purpose specification and transparency with regard to the processing. In a later version of DataBait the user could be offered the possibility to pick preferred granular licensing settings and compare them with the data policy of the OSNs and/or browsers she is using. In the final version of this deliverable granular licensing will be explored in more detail.

7.Annexes

7.1. Profiling in two versions of the proposed GDPR

This table compares the original text of the GDPR (proposed by the EU Commission) and the amended GDPR (by the EU Parliament) with regard to profiling. Everything that is **bold** indicates differences between the two versions. The words 'profile' and 'profiling' have been underlined to make it easier to see where they are discussed.[85]

	Text proposed by the Commission, submitted to the European Parliament on 25 January 2012[86]	***Amendments adopted on 12 March 2014 by the European Parliament***[87]
+ *Monitoring* data subjects: tracking application of a profile	(Recital 21) In order to determine whether a processing activity can be considered to 'monitor ***the behaviour' of*** data subjects, it should be ascertained whether individuals are tracked ***on the internet with*** data processing techniques which consist of applying a 'profile' ***to an individual*** , particularly in order to take decisions concerning her or him or for analysing or predicting her or his personal preferences, behaviours and attitudes.	(Recital 21) In order to determine whether a processing activity can be considered to 'monitor' data subjects, it should be ascertained whether individuals are tracked***, regardless of the origins of the data, or if other data about them is collected, including from public registers and announcements in the Union that are accessible from outside of the Union, including with the intention to use, or potential of subsequent use of*** data processing techniques which consist of applying a 'profile', particularly in order to take decisions concerning her or him or for analysing or predicting her or his personal preferences, behaviours and attitudes.

[85] In the final version of this deliverable (D3.10) we will look at the Council version.
[86] *Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, Brussels, 25 January 2012, COM(2012) 11 final.
[87] *European Parliament legislative resolution on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, Strasbourg, 12 March 2014, COM(2012)0011 – C7-0025/2012 – 2012/0011(COD). Ordinary legislative procedure: first reading. Online available at : <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2014-0212+0+DOC+XML+V0//EN>

***Right of access to the logic of the data* in relation to the rights and freedoms of others, such as IP rights.**	(Recital 51) Any person should have the right of access to data which has been collected concerning them, and to exercise this right easily, in order to be aware and verify the lawfulness of the processing. Every data subject should therefore have the right to know and obtain communication in particular for what purposes the data are processed, for what period, which recipients receive the data, what is the logic of the data that are undergoing the processing and what might be***, at least when based on profiling,*** the consequences of such processing. This right should not adversely affect the rights and freedoms of others, including trade secrets or intellectual property ***and in particular*** the copyright protecting the software. However, the result of these considerations should not be that all information is refused to the data subject.	(Recital 51) Any person should have the right of access to data which has been collected concerning them, and to exercise this right easily, in order to be aware and verify the lawfulness of the processing. Every data subject should therefore have the right to know and obtain communication in particular for what purposes the 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 and what might be the consequences of such processing. This right should not adversely affect the rights and freedoms of others, including trade secrets or intellectual property, ***such as in relation to*** the copyright protecting the software. However, the result of these considerations should not be that all information is refused to the data subject.
Right to object to profiling ; Prohibition of –especially discriminatory - profiling with legal or similar significant effects.	(Recital 58) Every natural person should have the right ***not to be subject*** to ***a measure which is based on*** profiling ***by means of automated processing***. ***However, such measure*** should be allowed when expressly authorised by law, carried out in the course of entering or performance of a contract, or when the data subject has given his consent. In any case, such processing should be subject to suitable safeguards, including specific information of the data subject and the right to obtain human ***intervention*** and that such measure should not concern a child.	(Recital 58) ***Without prejudice to the lawfulness of the data processing*** , every natural person should have the right ***to object*** to profiling. ***Profiling which leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject*** should ***only*** be allowed when expressly authorised by law, carried out in the course of entering or performance of a contract, or when the data subject has given his consent. The In any case, such processing should be subject to suitable safeguards, including specific information of the data subject and the right to obtain human ***assessment*** and that such measure should not concern a child. ***Such measures should not lead to discrimination against individuals on the basis of race or ethnic origin, political opinions, religion or beliefs, trade union membership, sexual orientation or gender identity.***

Presumption that profiling based *pseudonomous*		***(Recital 58a)*** ***Profiling based solely on the processing of pseudonymous data should be presumed not to significantly affect the interests, rights or freedoms of the data subject. Where profiling, whether based on a single source of pseudonymous data or on the aggregation of pseudonymous data from different sources, permits the controller to attribute pseudonymous data to a specific data subject, the processed data should no longer be considered to be pseudonymous.***
Restrictions on data protection rights may be imposed by Union or Member State law, as far as necessary and proportionate in a democratic society to safeguard public	(Recital 59) Restrictions on specific principles and on the rights of information, ***access,*** rectification and erasure or on the right to data ***portability*** , the right to object, ***measures based on*** profiling, as well as on the communication of a personal data breach to a data subject and on certain related obligations of the controllers may be imposed by Union or Member State law, as far as necessary and proportionate in a democratic society to safeguard public security, including the protection of human life especially in response to natural or man made disasters, the prevention, investigation and prosecution of criminal offences or of breaches of ethics for regulated professions, other public interests of the Union or of a Member State, in particular an important economic or financial interest of the Union or of a Member State, or the protection of the data subject or the rights and freedoms of others. Those restrictions should be in compliance with requirements set out by the Charter of Fundamental Rights of the European Union and by the European Convention for the Protection of Human Rights and Fundamental Freedoms.	(Recital 59) Restrictions on specific principles and on the rights of information, rectification and erasure or on the right of ***access and to obtain*** data, the right to object, profiling, as well as on the communication of a personal data breach to a data subject and on certain related obligations of the controllers may be imposed by Union or Member State law, as far as necessary and proportionate in a democratic society to safeguard public security, including the protection of human life especially in response to natural or man made disasters, the prevention, investigation and prosecution of criminal offences or of breaches of ethics for regulated professions, other ***specific and well-defined*** public interests of the Union or of a Member State, in particular an important economic or financial interest of the Union or of a Member State, or the protection of the data subject or the rights and freedoms of others. Those restrictions should be in compliance with requirements set out by the Charter of Fundamental Rights of the European Union and by the European Convention for the Protection of Human Rights and Fundamental Freedoms.

The power to adopt more specific acts to fulfill the objectives of the GDPR is delegated to the Commission.	(Recital 129) In order to fulfil the objectives of this Regulation, namely to protect the fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data and to ensure the free movement of personal data within the Union, the power to adopt acts in accordance with Article 290 of the Treaty on the Functioning of the European Union should be delegated to the Commission. In particular, delegated acts should be adopted in respect of ***lawfulness of processing*** ; specifying ***the criteria and conditions in relation to the consent of a child; processing of special categories of data; specifying the criteria and conditions for manifestly excessive requests and fees for exercising the rights of the data subject; criteria and requirements for the information to the data subject and in relation to the right of access;*** the right to ***be forgotten and to*** erasure; ***measures based on profiling*** ; ***criteria and requirements in relation to the responsibility of the controller and to data protection by design and by default; a processor; criteria and requirements for the documentation and the security of processing; criteria and requirements for establishing a personal data breach and for its notification to the supervisory authority*** , ***and on the circumstances where a personal data breach is likely to adversely affect the data subject; the criteria and conditions for processing operations requiring a data protection impact assessment; the criteria and requirements for determining a high degree of specific risks which require prior consultation; designation and tasks of the data protection officer;*** codes of conduct; criteria and requirements for certification mechanisms; criteria and requirements for transfers by way of binding corporate rules; ***transfer derogations;*** administrative sanctions; processing for health purposes; processing in the employment context	(Recital 129) In order to fulfil the objectives of this Regulation, namely to protect the fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data and to ensure the free movement of personal data within the Union, the power to adopt acts in accordance with Article 290 of the Treaty on the Functioning of the European Union should be delegated to the Commission. In particular, delegated acts should be adopted in respect of specifying ***conditions of icon-based mode for provision of information;*** the right to erasure; ***declaring that*** codes of conduct ***are in line with the Regulation*** ; criteria and requirements for certification mechanisms; ***the adequate level of protection afforded by a third country or an international organisation;*** criteria and requirements for transfers by way of binding corporate rules; administrative sanctions; processing for health purposes ***and*** processing in the employment context. It is of particular importance that the Commission carry out appropriate consultations during its preparatory work, including at expert level ***in particular with the European Data Protection Board*** . The Commission, when preparing and drawing-up delegated acts, should ensure a simultaneous, timely and appropriate transmission of relevant documents to the European Parliament and ***to the*** Council.

	and processing for historical, statistical and scientific research purposes . It is of particular importance that the Commission carry out appropriate consultations during its preparatory work, including at expert level. The Commission, when preparing and drawing-up delegated acts, should ensure a simultaneous, timely and appropriate transmission of relevant documents to the European Parliament and Council.	
Definition of profiling		***(Article 4-3a)*** ***'profiling' means any form of automated processing of personal data intended to evaluate certain personal aspects relating to a natural person or to analyse or predict in particular that natural person's performance at work, economic situation, location, health, personal preferences, reliability or behaviour;***
General principles for data subject rights : includes the right to object to profiling		***(Article 10a)*** ***General principles for data subject rights*** ***1. The basis of data protection is clear and unambiguous rights for the data subject which shall be respected by the data controller. The provisions of this Regulation aim to strengthen, clarify, guarantee and where appropriate, codify these rights.*** ***2. Such rights include, inter alia, the provision of clear and easily understandable information regarding the processing of his or her personal data, the right of access, rectification and erasure of their data, the right to obtain data, the right to object to profiling, the right to lodge a complaint with the competent data protection authority and to bring legal proceedings as well as the right to compensation and damages resulting from an unlawful processing operation. Such rights shall in general be exercised free of charge. The data controller shall respond to requests from the data subject within a reasonable period of time.***

Duty to inform about existence of profiling.		**(Article 14-ga)** Information to the data subject. 1. Where personal data relating to a data subject are collected, the controller shall provide the data subject with at least the following information, ***[…]***: […] ***ga) where applicable, information about the existence of profiling, of measures based on profiling, and the envisaged effects of profiling on the data subject;***

Profiling which leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject shall not be based solely or predominantly on automated processing and shall include human assessment, including an explanation of the	(Article 20) ***Measures based* on** profiling 1. Every natural person shall have the right ***not to be subject*** to ***a measure which produces legal effects concerning this natural person or significantly affects this natural person, and which is based solely on automated processing intended to evaluate certain personal aspects relating to this natural person or to analyse or predict in particular the natural person's performance at work, economic situation, location, health, personal preferences, reliability or behaviour.*** 2. Subject to the other provisions of this Regulation, a person may be subjected to ***a measure of the kind referred to in paragraph 1*** only if the processing: (a) is ***carried out in the course of*** the entering into, or performance of, a contract, where the request for the entering into or the performance of the contract, lodged by the data subject, has been satisfied ***or where*** suitable measures to safeguard the data subject's legitimate interests have been adduced***, such as the right to obtain human intervention*** ; or (b) is expressly authorized by a Union or Member State law which also lays down suitable measures to safeguard the data subject's legitimate interests; or (c) is based on the data subject's consent, subject to the conditions laid down in Article 7 and to suitable safeguards. 3. ***Automated processing of personal data intended to evaluate certain personal aspects relating to a natural person*** shall not be based solely on the special categories of personal data referred to in Article 9. ***4. In the cases referred to in paragraph 2, the information to be provided by the controller under Article 14 shall include information as to the existence of processing for a measure of the kind referred to in paragraph 1 and the envisaged effects of such processing on the data subject.*** 5. ***The Commission shall be***	(Article 20) Profiling ***1. Without prejudice to the provisions in Article 6*** every natural person shall have the right ***to object*** to ***profiling in accordance with Article 19. The data subject shall be informed about the right to object to profiling in a highly visible manner.*** 2. Subject to the other provisions of this Regulation, a person may be subjected to ***profiling which leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject*** only if the processing: *(a)* is ***necessary for*** the entering into, or performance of, a contract, where the request for the entering into or the performance of the contract, lodged by the data subject, has been satisfied***, provided that*** suitable measures to safeguard the data subject's legitimate interests have been adduced*;* or (b) is expressly authorized by a Union or Member State law which also lays down suitable measures to safeguard the data subject's legitimate interests; (c) is based on the data subject's consent, subject to the conditions laid down in Article 7 and to suitable safeguards. 3. ***Profiling that has the effect of discriminating against individuals on the basis of race or ethnic origin, political opinions, religion or beliefs, trade union membership, sexual orientation or gender identity, or that results in measures which have such effect, shall be prohibited. The controller shall implement effective protection against possible discrimination resulting from profiling. Profiling*** shall not be based solely on the special categories of personal data referred to in Article 9. 5. ***Profiling which leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject shall not be based solely or predominantly on automated***

	empowered to adopt delegated acts in accordance with Article 86 for the purpose of further specifying the criteria and conditions for suitable measures to safeguard the data subject's legitimate interests referred to in paragraph 2.	***processing and shall include human assessment, including an explanation of the decision reached after such an assessment. The*** suitable measures to safeguard the data subject's legitimate interests referred to in paragraph 2 ***shall include the right to obtain human assessment and an explanation of the decision reached after such assessment***. ***5a. The European Data Protection Board shall be entrusted with the task of issuing guidelines, recommendations and best practices in accordance with point (b) of Article 66 (1) for further specifying the criteria and conditions for profiling pursuant to paragraph 2.***
Risk analysis: **Profiling resulting in measures with legal or similar significant effects is**		***(Article 32a) Respect to Risk*** ***1. The controller, or where applicable the processor, shall carry out a risk analysis of the potential impact of the intended data processing on the rights and freedoms of the data subjects, assessing whether its processing operations are likely to present specific risks.*** ***2. The following processing operations are likely to present specific risks:*** ***[…]*** ***(c) profiling on which measures are based that produce legal effects concerning the individual or similarly significantly affect the individual;*** ***[…]***

***Binding corporate rules* shall include the right not to be subject to a measure based on profiling.**	(Article 43) Transfers by way of binding corporate rules 1. A supervisory authority shall in accordance with the consistency mechanism set out in Article 58 approve binding corporate rules, […]: […] […] 2. The binding corporate rules shall at least specify: […] (e) the rights of data subjects and the means to exercise these rights, including the right not to be subject to a measure based on profiling in accordance with Article 20, the right to lodge a complaint before the competent supervisory authority and before the competent courts of the Member States in accordance with Article 75, and to obtain redress and, where appropriate, compensation for a breach of the binding corporate rules;	(Article 43) Transfers by way of binding corporate rules ***1. The*** supervisory authority shall in accordance with the consistency mechanism set out in Article 58 approve binding corporate rules, […]: […] 2. The binding corporate rules shall at least specify: […] (e) the rights of data subjects and the means to exercise these rights, including the right not to be subject to a measure based on profiling in accordance with Article 20, the right to lodge a complaint before the competent supervisory authority and before the competent courts of the Member States in accordance with Article 75, and to obtain redress and, where appropriate, compensation for a breach of the binding corporate rules;
No profiling in the employment context	(Article 82) Processing in the employment context […]	(Article 82) ***Minimum standards for*** processing ***data*** in the employment context […] ***1a. The purpose of processing such data must be linked to the reason it was collected for and stay within the context of employment. Profiling or use for secondary purposes shall not be allowed.***

7.2. The annotated PDPA and DLA

7.2.1. Annotated 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 data of DataBait Users, who are defined as the USEMP end-users who have signed the Data Licensing Agreement with the USEMP Consortium Partners.	This links the PDPA to the DLA.

(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 stored at the premises of CEA, that will be authorized to run its algorithms on the data stored at HWC.	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 compliant with EU data protection law.

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 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	Because all the USEMP partners

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.	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.

7.2.2. Annotated 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],[88] provider of the USEMP platform and services, joint data controllers, from hereon called 'USEMP consortium partners'.[89] 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	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

[88] 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/.

[89] 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."

(storing what You did on the Internet and on FaceBook).	
(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	This establishes the legal ground (art. 7 DPD) for the collection and processing of the survey data. This is necessary in order to be

and targeted profiles and (2) the effectiveness, usability and utility of the USEMP tools.	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: ☐	

7.3. Creative Commons Licenses

Attribution
CC BY

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

Attribution-ShareAlike
CC BY-SA

This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to “copyleft” free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

Attribution-NoDerivs
CC BY-ND

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

Attribution-NonCommercial
CC BY-NC

This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

Source: https://creativecommons.org/licenses/

7.4. Typology of data based on the mode of data capture

The social semantic web is only one of the ways that the commodification of the social becomes enabled in online networked technologies. A more encompassing view can be obtained when turning to the affordances offered by the interfaces Facebook offers to advertisers. Through these interfaces users become socially reassembled according to series of fundamental relational categories, which form a broader conception of the social ontology operative in these network technologies.[90] These categories are themselves made possible by the different channels of information flow within this technological infrastructure. We can distinguish the following **five modes of data capture**[91] and their correlated objectification into some of the ***basic concepts of the commodity ontology*** of online social networks:

- **Registration data & page content** are basically data obtained by the ways of user engages in self-categorization either through the processes of joining the OSN, or the data disclosed on the pages of the user or others.[92] In the advertisement interface these encompass most of the data categories for targeting users. Many of them include classical ***demographics*** like age, gender, education, languages, workplaces, relationship status, but they also include the ***specific interests*** indicated by the user.
- **Incidental data** capture information about a user through the behavior of other users. This relates primarily to the direct actions that the Facebook platform performs like tagging, posting, etc. In advertisement however incidental data plays a role on a different level through the "***connections***" category, especially by enabling the targeting of "friends of connections". This is a kind of social network analysis by which data about someone can be derived through their degrees of connectedness to others.[93]
- **Traffic data** are basically meta-data not about the content of a online behavior but which are often necessary for the carrying out of these behaviors.[94] On Facebook these traffic data include both the **technical/logging data** about the type of computer, type of operating software, type of browser of the user and the more **browser behavioral data** made possible by all kinds of online identifiers like

[90] Broader in the sense that it goes beyond the vocabularies of the social semantic web in order

[91] Some of these data categories overlap with the 6 types of data used in online social networks as distinguished by Schneier 2010: service data, disclosed data, entrusted data, incidental data, behavioural data and derived data. Whereas the user actively discloses the first three types of data, this is not the case with the latter three types of data. Especially derived data are relevant in the context of this article, since these refer to the data have been inferred about a user by the use of group profiling through data mining techniques like association and clustering algorithms. See: https://www.schneier.com/blog/archives/2010/08/a_taxonomy_of_s_1.html

[92] This correlate with what Schneier calls service data and disclosed data.

[93] For an early account of social network techniques for online consumer profiling, see Olcay Cirit, F., Nikravesh, M., Alptekin, S., Consumer Profiling Using Fuzzy Query and Social Network Techniques, in M. Nikravesh, L. Zadeh, J. Kacprzyk, Soft Computing for Information Processing and Analysis, Springer, 2006

[94] The e-privacy Directive 2002/58/EC provides the following examplary list: "data referring to the routing, duration, time or volume of a communication, to the protocol used, to the location of the terminal equipment of the sender or recipient, to the network on which the communication originates or terminates, to the beginning, end or duration of a connection. They may also consist of the format in which the communication is conveyed by the network."(recital 15)

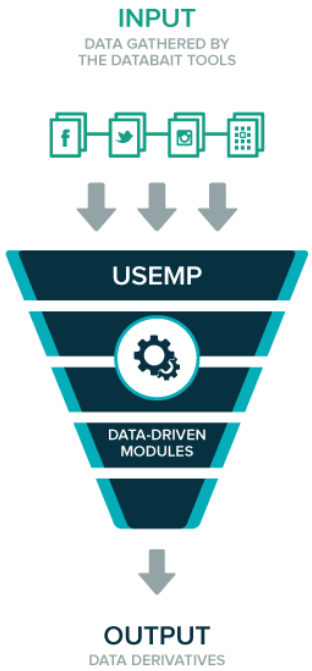
cookies, session trackers, IP addresses, etc. This is a standard basis for web advertisement and plays a crucial role in Google's Analytics program. For advertisements on Facebook these data are relevant in the category ***location*** which can be determined on the basis of IP address[95], but also as one of the "broad categories" see image hereunder] pertaining to what we can call the ***traffic medium*** used, which enables Facebook to extend to the mobile market and finetune the "placement" of its ads.

- **Interaction data** play a very important role in OSNs. They include most of the social actions a user can perform on a networking platform.[96] For advertisement purposes they play a crucial role in the category of "interest targeting". These interests are taken from several indicators. The most significant action is the crucial function of ***liking*** that has become afforded through the design of the like button for direct preference indication and its plug-ins on other sites. Also very important is the ***subscription*** to applications that plug into Facebook and, as we have seen above, can render stories about the user through their underlying web semantics. Furthermore ***membership*** of groups or events is also interpreted as an indicator for interest.[97]
- **Inferred data** are data about a user derived from all these previous data types of the users and of other users obtained through data mining techniques in order to learn new information about people. We have become acquainted with these techniques in the discussion about profiling in this article. These data play several roles in the advertising interface. Firstly, Facebook offers a few pre-mined profiles included in the "***broad interests categories***" which especially relate to one's "family status" [see image hereunder] Secondly, when advertisers have selected certain likes and interests as targets Facebook automatically offers "suggested likes and interests". These are "the terms that are most common among the people your targeting criteria already includes."[98] These ***conjectured interests*** are thus likely derived through clustering methods or association rules, in order to aggregate group profiles with shared features. Lastly, we could probably also include Facebook "topic targeting" under inferred data.[99] Certain interest keywords include overlapping precise interests. These terms can be called ***topical interests***.

[95] It could also be argued that the category of location is actually inferred data, since these data have to be derived from IP addresses which themselves do not yet directly indicate location. Furthermore, user location can also be obtained on the basis of self-categorization by the user.
[96] This correlate with what Schneier calls behavioural data.
[97] "Interest targeting helps advertisers target people based on information they've added to their timeline. This considers information such as the Pages they like, apps they use" and groups to which they belong, or "may be drawn from their listed interests, activities, education and job titles". This function thus also makes use of registration and profile data. https://www.facebook.com/help/www/453530464730606/
[98] https://www.facebook.com/help/www/453530464730606/
[99] https://developers.facebook.com/docs/reference/ads-api/topic-targeting/

7.5. Text and flow-charts of ‘DataBait at a Glance’

DataBait shows you who tracks you on the internet and which guesses and predictions about you and your personality can be derived by analyzing your digital trail with smart software. Moreover, it gives an indication of the economic value of your profile and your Facebook friends. Databait is a tool created in a European research project, USEMP.

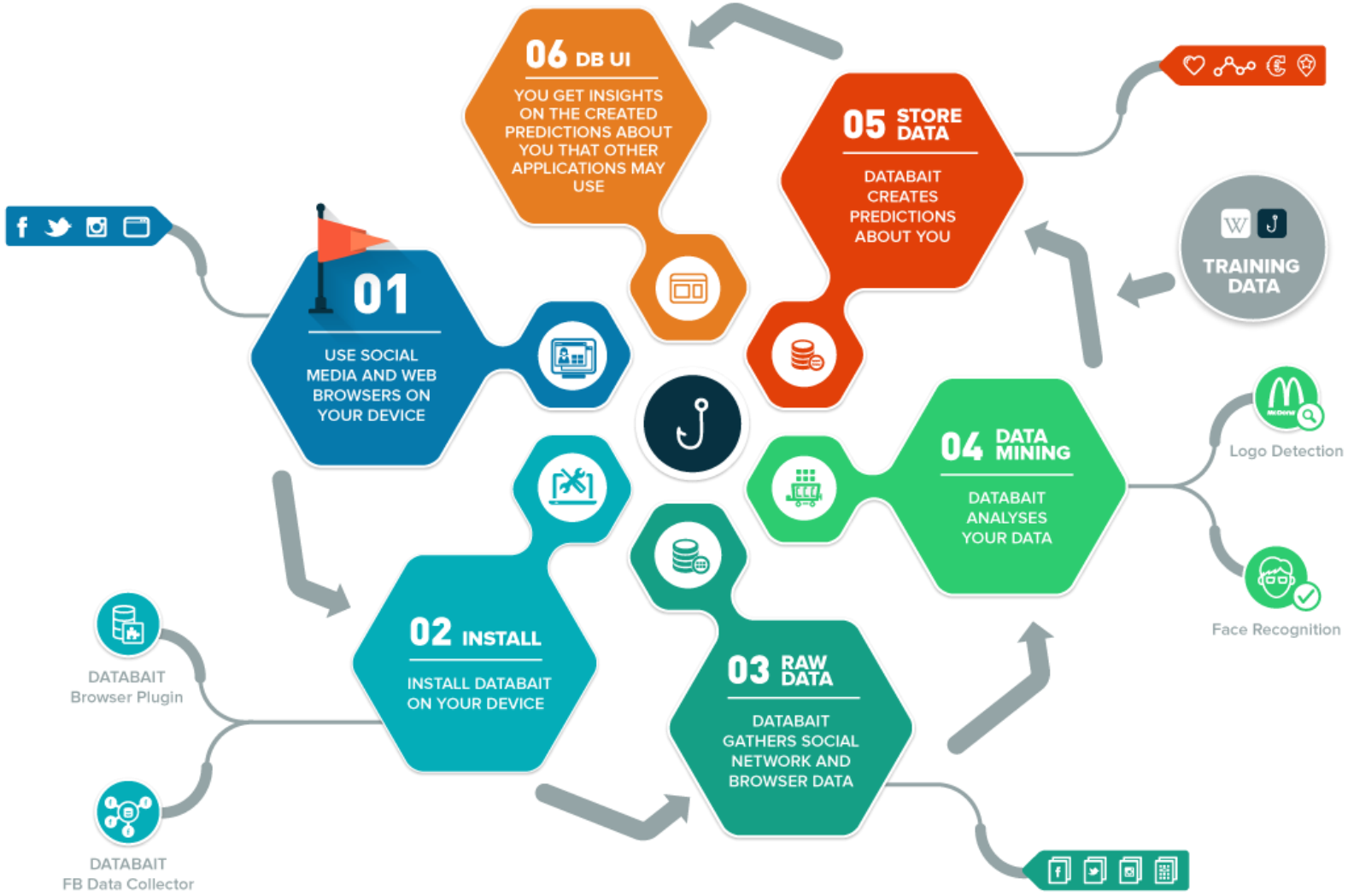


In order to give you these insights, we collect, with your permission, data from your Facebook profile and (if you choose to install the browser plugin) data from your online surfing behaviour. All (raw) data are safely stored in a database on a server in the UK at the premises of one of the USEMP partners – HWC. You can find more about HWC here. On this server, your data are analyzed by Databait analytic software in order to investigate if it's possible for companies and institutions to derive insights about you and your personality from the raw data you (consciously and unconsciously) generate through your activities on the Internet. The DataBait analytic software can do a lot of things, like guessing what's depicted on a picture and extracting a mood from a bit of text (happy? sad?). You can find more information here.

Wondering how we taught our DataBait software to make smart guesses and predictions about you? For example, how does our software know that it's you standing in front of the Eiffel tower on your holiday picture? How does the software conclude that the building depicted on the picture is the Eiffel tower? How can it know that this means that you have been in Paris? How can the USEMP softwarederive from your posts that you are unhappy? Or what your sexual orientation is? One way we made our software intelligent is by letting it learn from lots of examples (e.g. pictures from different angles of the Eifel tower). Here you can find a list of databases we used for training purposes: Databases.

We only used public available databases, such as Wikipedia and databases with Flickr pictures. During the pilot phase of DataBait we also ask you to fill in a survey where we ask you to tell us something about yourself: basic information like age and gender, but also more personal stuff like religious views and personality type. By comparing the self-reported data from the survey with the predictions made by the DataBait software we can see how well our software works – and adjust it to make it work (even) better. Here you find a list of all the data we try to derive.

The derived data are used to present to you how third parties (such as trackers) might create a 'hidden profile' of you, which consists of features you might not even realize you disclosed on the Internet. Moreover, we want to show what economic value your profile and your Facebook friends can present for others (like advertising agencies). All these derived data are stored in a second database at the premises of HWC. Here you can find more about the security of your data and how we anonymize your data.



Below you find a list of all the project partners involved in the USEMP project and how they will reason on your data:

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 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.

iMinds will also participate in the preparation of the survey with LTU and CERTH gathering the required data from registered users of the USEMP platform and the DataBait tools to enable testing of how the inferences drawn from DataBait Users' postings, social graphs and behavioural data match their real preferences and background. iMinds can only access the result of the survey (stored at HWC database) based on secured authorization, where 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.

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

8.Bibliography

Article 29 Data Protection Working Party 29. (2013). Opinion 03/2013 on purpose limitation. Adopted on 2 April 2013.

Article 29 Data Protection Working Party 29. (2015). Press release on Chapter II of the draft regulation for the March JHA Council [Press release]. Retrieved from http://ec.europa.eu/justice/data-protection/article-29/press-material/press-release/art29_press_material/20150317_wp29_press_release_on_on_chapter_ii_of_the_draft_regulation_for_the_march_jha_council.pdf

Article 29 Data Protection Working Party. (2012). Opinion 04/2012 on Cookie Consent Exemption.

Article 29 Data Protection Working Party. (2014). Opinion 05/2014 on Anonymisation Techniques (Adopted on 10 April 2014).

Article 29 Data Protection Working Party. (2015a). Annex to the letter (Brussels, 05 February 2015) to the European Commission on health data in apps and devices.

Article 29 Data Protection Working Party. (2015b). Letter (Brussels, 05 February 2015) to the European Commission on health data in apps and devices.

Asgharpour, F., Liu, D., & Camp, L. J. (2007). Mental models of security risks *Financial Cryptography and Data Security* (pp. 367-377): Springer.

Camp, L. J. (2006). Mental models of privacy and security. *Available at SSRN 922735*.

European Commission. (2012). FREQUENTLY ASKED QUESTIONS RELATING TO TRANSFERS OF PERSONAL DATA FROM THE EU/EEA TO THIRD COUNTRIES.

European Data Protection Supervisor. (2015). Opinion 3/2015. Europe's big opportunity. EDPS recommendations on the EU's options for data protection reform.

European Union Agency for Fundamental Rights, C. o. E. (2014). *Handbook on European data protection law*. Luxembourg: Publications Office of the European Union.

Hildebrandt, M. (2014). Location Data, Purpose Binding and Contextual Integrity: What's the Message? *Protection of Information and the Right to Privacy-A New Equilibrium?* (pp. 31-62): Springer.

Koning, M. (2014). *Purpose Limitation and Fair Re-use*. Paper presented at the Computers Privacy and Data Protection 2014, Brussels.

Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large datasets (how to break anonymity of the Netflix prize dataset). 2008. *University of Texas at Austin*.

Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review, 57*, 1701.

Wauters, E., Lievens, E., & Valcke, P. (2014). Towards a better protection of social media users: a legal perspective on the terms of use of social networking sites. *International Journal of Law and Information Technology, 22*(3), 254-294. doi: 10.1093/ijlit/eau002