D3.7

Profile transparency, Trade Secrets and Intellectual Property Rights in OSNs – v2

v 1.3 / 2015-03-31

Katja de Vries and Sari Depreeuw (ICIS-RU)

This document analyses whether the end users' right to profile transparency and the way in which DataBait supports this (see D3.6) could be obstructed by the protection of trade secrets or the Intellectual Property Rights [IPRs] of OSNs or other actors. The IPRs which are discussed are patents, database rights, copyrights and trademarks. This report makes an inventory of IP rights which protect content collected and analysed through DataBait, and studies the likelihood that this would infringe on exclusive rights on the content, the OSN databases to which the content belongs, and/or the OSN graphic user interfaces in which the content is represented. Due to the particular architecture of DataBait, the Data Licensing Agreement (DLA) signed by DataBait users, and the existence of exceptions for scientific research, the likelihood of infringement is not very large. However, there are several issues that deserve careful attention and continuous monitoring during the remainder of the USEMP project. In terms of IPRs this report also analyses the relationship between DataBait software and software protected by patents or copyrights of OSNs or other rights holders. In this regard we conclude that the risk of infringement is very small due to the fact that DataBait has created its own independent software and only simulates the overall profiling process without mimicking or reproducing any specific methods employed by others (such as the studied OSNs). This report includes some design implications for the DataBait tools, and concludes with a research agenda for the next version of this deliverable (D3.11) in month 36 of the project.

Project acronym	USEMP
Full title	User Empowerment for Enhanced Online Presence Management
Grant agreement number	611596
Funding scheme	Specific Targeted Research Project (STREP)
Work program topic	Objective ICT-2013.1.7 Future Internet Research Experimentation
Project start date	2013-10-01
Project Duration	36 months

Workpackage	WP3
Deliverable lead org.	USEMP
Deliverable type	Report
Authors	Katja de Vries, Sari Depreeuw, Mireille Hildebrandt (iCIS)
Reviewers	Symeon Papadopoulos (CERTH) Ali Padyab (LTU)
Version	1.5
Status	Final
Dissemination level	**PU: Public**
Due date	2015-12-31
Delivery date	2016-02-01

Version	Changes
1.0	Initial Release (De Vries, ICIS)
1.1	Adjusted version (Depreeuw, ICIS)
1.2	Review (Papadopolous, CERTH)
1.3	Review (Padyab, LTU)
1.4	Adjusted version (Depreeuw, ICIS)
1.5	Minor adjustments (De Vries, ICIS) after consulting Popescu (CEA)

Table of Contents

Summary

Profile transparency is a legal right under current and upcoming data protection law. It is, however, subject to limitations (see recital 42 of the DPD 95/46) due to trade secrets and Intellectual Property Rights (IPRs) (notably copyright in a database or in a computer program and the so-called database right *sui generis*) of those who engage in profiling. Though the latter cannot entirely erode the substance of the right to profile transparency, it is conceivable that OSN providers could claim either trade secret or IPRs against end users that claim their right to profile transparency. Similarly, an OSN could invoke the same rights to the provider of a profile transparency tool (like DataBait) that has distinct interests from its users (data subjects). The technical partners in the USEMP project provided extensive input concerning the algorithms, databases and data exploited as well as the software used in creating DataBait. All of this contributes to ensuring that DataBait does not infringe on any IPRs of OSNs or other actors (e.g., the creators of the databases used to train and test the DataBait algorithms). The analysis presented in this deliverable also aims to explore how the rights of commercial profilers can (partly) oppose claims to profile transparency, and how to inform USEMP end-users through the DataBait graphic user interface about the possible tensions between IPRs of profilers and their right to profile transparency.

Within this deliverable we also elaborate on how DataBait shows end users what *could* be extracted from their data (which makes DataBait both speculative about how a user might be currently profiled as well as forward looking with regard to profiling to which she might be subjected in the near future: it thus mimics the profiling 'reality' in general without copying any particular profiling algorithm): this fundamentally differs from reproducing existing code or 'reverse engineering' it. USEMP does not reproduce the actual code or other protected elements of computer programs, owned by OSN providers; instead it creates own software to present end users with *potential* inferences by those with access to similar data.

Another question which is addressed in this deliverable is what kind of copyright protection should be granted to the so-called "banal" creations that are not exploited as original creations but only as "content" driving "data traffic" (e.g. a picture of breakfast cereal is not appreciated for its "originality" and not exploited individually but it attracts data traffic that is used for targeting and profiling practices). We raise the question whether original elements of a copyright protected work are reproduced in case of data analysis, where the data are copied for functional reasons and how extraction and reutilization for profiling purposes in database protection should be qualified. We will explore which 'tolerances' within IPRs protection (such as those existing for scientific referencing) could be applicable to the profiling process and which not.

1. DataBait: a profile transparency tool which does not infringe on OSN's trade secrets or IPRs

1.1. Role of this deliverable within WP3

The overall goal of the legal input in Work Package 3 (*Legal Requirements and the Value of Personal Data*) is to elicit/engineer legal requirements that should inform the development of the various USEMP tools. Thus, the legal deliverables in WP3 are not just theoretical legal treatises on data protection, anti-discrimination, and IPRs in relation to the profiles built in and through OSNs, but they aim to provide concrete, hands-on legal requirements which are translated into technical specifications for the architecture of DataBait (i.e. the profile transparency tool created by the USEMP consortium).

All of the legal input in WP3 is quite hands-on. Yet, with respect to the deliverables produced within WP3 we distinguish between 'research' deliverables (see Table 1), presenting the research in which the legal requirements for DataBait are based, and the 'coordination and integration' deliverables (D3.4, D3.9 and D3.13) that report on how the legal requirements are interfaced with the tasks at hand in the other WPs and transposed into the DataBait architecture.

	Version 1	Version 2	Version 3
Fundamental Rights Protection by Design for OSNs	D3.1 (delivery date: M12)	D3.6 (delivery date: M21)	D3.10 (delivery date: M36)
Profile transparency, trade secrets and Intellectual Property rights in OSNs	D3.2 (delivery date: M12)	D3.7 (delivery date: M24)	D3.11 (delivery date: M36)
Copyrights and portrait rights in content posted on OSNs	D3.3 (delivery date: M12)	D3.8 (delivery date: M24)	D3.12 (delivery date: M36)

Table 1: Overview of the deliverables in WP3 containing original legal research

As shown in Figure 1, the legal research (D3.1-3.3, D.3.6-3.8 and D3.10-12) and the integration of the legal requirements into the design of the USEMP tools (D3.4, D.3.9 and D3.13) are intertwined with each other. D3.1-3.3, D.3.6-3.8 and D3.10-12 reflect the work done in T3.1-3.5 [M1-M36]. D3.4, D.3.9 and D3.13 reflect the work done in T3.6 [M1-M36], which implements legal coordination.

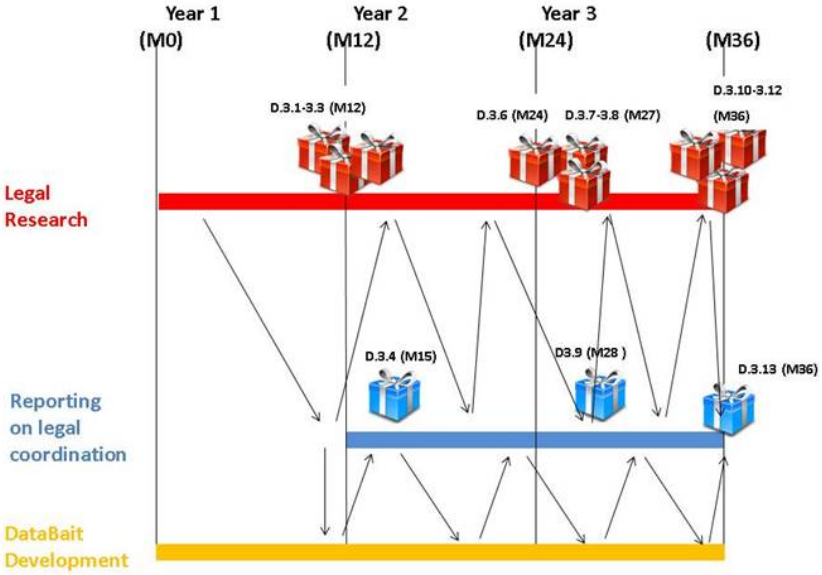


Figure 1. Timeline for the legal deliverables in WP3

This deliverable, which builds on the research performed in task T3.3. (*Relevant EU Intellectual property rights regarding databases and software employed by the OSN*) looks at how the IPRs and trade secrets held by Online Social Networks (OSNs) could impact on the DataBait tool. A main concern is to ensure that DataBait does not infringe on any IPRs or trade secret held by an OSN. In terms of the organisation of WP3, this means that the research of T3.3 (resulting in this deliverable) looks at how IPRs and trade secrets could interfere with the possibilities of empowerment of OSN users studied in T3.1 and T3.2. The research from T3.1 ("Fundamental Rights Protection by Design for OSNs") and T3.2 ('Relevant EU legal framework for non-discrimination') generated both requirements of empowerment and compliance. The *empowerment requirements (*derived from data protection and antidiscrimination law) showed how DataBait can best support informational rights of OSN end-users by providing them transparency about the additional knowledge and value which could be derived from the digital trail they leave behind when using a particular OSN. The *compliance requirements* showed how the data processing by DataBait should take place order to be in accordance with data protection law.

In this deliverable we show how the DataBait architecture is not merely compliant with data protection law but also refrains from infringing on trade secrets and IPRs of OSNs. We make an inventory of the relevant requirements (derived from IP and trade secret law) which ensure that DataBait does neither expose any OSN trade secrets nor commit any prohibited acts with regard to IP protected matter belonging to an OSN (such as the algorithms used by the OSN to derive additional information from the data generated by OSN users or the way these data are structured by the OSN). Three crucial elements for this analysis are, firstly, the role of DataBait (and the USEMP consortium) in the relation between the 'OSN-as-data-controller' and the 'OSN-user-as-data-subject'; secondly, a correct legal qualification of the data deriving activities ('profiling') of the OSN and those of DataBait; and finally, an assessment of how these two data deriving activities relate to each other from a technical and legal perspective. In the following section (1.2) we begin by looking at how the role of

DataBait should be defined – namely as a facilitator, that is: an independent and supportive actor, for the OSN end-user in her relation towards the OSN. Section 1.3 and 1.4 describe how the profiling activities of DataBait should be qualified, and section 1.5 gives a first indication of how these activities relate to those performed by the OSN.

1.2. DataBait as a facilitator for the OSN end-user

EU data protection law recognizes that informational rights of the data subject could clash with the protection of trade secrets or IPRs (copy- and database rights) of the data controller (who controls the system or practice which tracks and profiles its users). *Data Protection Directive 95/46* states in Recital 41 that:

> Whereas any person must be able to exercise the right of access to data relating to him which are being processed, in order to verify in particular the accuracy of the data and the lawfulness of the processing; whereas, for the same reasons, every data subject must also have the right to know the logic involved in the automatic processing of data concerning him, at least in the case of the automated decisions referred to in Article 15 (1); whereas this right must not adversely affect trade secrets or intellectual property and in particular the copyright protecting the software; whereas these considerations must not, however, result in the data subject being refused all information;

The protection of trade secrets and IPRs held by the data controller might thus necessitate that the right of access and the right to be informed about the logic involved in a profiling practice are limited. However, such considerations can never fully eradicate these informational data protection rights of the data subject. The heavy weight that has to be attributed to the right to respect for private life and data protection when balancing it with regard to the commercial interests of a data controller (Art. 16 of the EU Charter : the right to conduct a business) under the DPD was stressed in *Google Spain v AEPD and Mario Costeja Gonzalez*[1] (sections 56-58) :

> « *…the activities relating to the advertising space constitute the means of rendering the search engine at issue economically profitable and that engine is, at the same time, the means enabling those activities to be performed. […] Since […] [the] display of results is accompanied, on the same page, by the display of advertising linked to the search terms, it is clear that the processing of personal data in question is carried out in the context of the commercial and advertising activity […]. That being so, it cannot be accepted that the processing of personal data carried out for the purposes of the operation of the search engine should escape the obligations and guarantees laid down by Directive 95/46, which would compromise the directive's effectiveness and the effective and complete protection of the fundamental rights and freedoms of natural persons which the directive seeks to ensure […], in particular their right to*

[1] Decision of the CJEU (Grand Chamber) of 13 May 2014, C-131/12, ECLI:EU:C:2014:317.

> *privacy, with respect to the processing of personal data, a right to which the directive accords special importance […].* »

While the outcome of a balancing act always depends on the particulars of a case, it would be likely that, if a court had to strike a balance between fundamental informational rights of a data subject and the protection of IP rights and trade secrets of a data controller, the protection of the former would not lightly be put aside.

In Recital 51 of the proposed *General Data Protection Regulation* one can find a similar approach: while the necessity to strike a balance between informational rights of the data subject and the protection of trade secrets and IPRs of the data controller is recognized, the result of this balancing act can never result in a complete obliteration of the former in favour of the latter:

> 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.

How does the information provided by a profile transparency tool, like DataBait, fit in this balancing between profile transparency and the protection of trade secrets and IPRs of OSNs? After all, DataBait is merely *a third actor* with regard to the relationship between the 'OSN-user-as-data-subject' and the 'OSN-as-data-controller'. DataBait only *supports* the data subject in her relation to the OSN (the data controller) and does not act as a stand-in for either the 'OSN-user-as-data-subject' or the 'OSN-as-data-controller'. DataBait may facilitate the exercise of the data-subject's informational rights, but it does not exercise these rights on the data-subject's behalf. DataBait does not have the rights the 'OSN-user-as-data-subject' has towards the 'OSN-as-data-controller'.

Also, DataBait cannot fulfil the informational duties of the 'OSN-as-data-controller' towards the 'OSN-user-as-data-subject'. The 'profile transparency' DataBait provides is something additional: it is independent, and fundamentally different, of the profile transparency the 'OSN-as-data-controller' is obliged to provide to the 'OSN-user-as-data-subject'. It does not exhaust or replace the duties of the OSN towards its end-users (as their data subject): an OSN can never fulfil its informational duties by simply referring to the information generated by DataBait.

With regard to 'factual' empowerment, the role of DataBait is to provide information which helps the OSN end-user to make better informed decisions (for example: deciding whether to remove a picture from the OSN after DataBait has informed the user that health information could be derived from it). With regard to 'legal' empowerment, the role of DataBait is to enable the OSN end-user to pose relevant questions to the OSN and make use of

informational rights such as the right of access or the right to be informed about the logic of the profiling to which she is subjected. A user who is unaware of the data-driven business ecology of OSNs or the technical possibilities to derive additional data from her digital trail, is unlikely to make any attempt to effectuate her informational rights towards a data controller. By giving a user oversight over the data she has posted on an OSN, the trackers which follow her, the data which can be derived from her digital trail, their possible value and the influence she has on her OSN connections ('friends', 'followers', etc.), she can become more empowered (both in the meaning of 'factual' control and the effectuation of her legal rights).

1.3. How 'real' is the DataBait disclosure score? A disclaimer.

It is important to underline that the information provided to the user by DataBait concerning what can be inferred from her digital trail, particularly the so-called 'disclosure score' (later in this section we discuss this notion in more detail), is in some sense 'speculative'. DataBait 'simulates' a potential scenario of user profiling by a third party (such as an OSN or another commercial profiler): it does not simulate their methods, nor their outputs, but the overall process.

DataBait shows what additional information can be derived from one's digital trail based on the state of the art of data analytics: it cannot tell whether this information actually is derived or not. DataBait shows what is possible. This 'speculative' aspect of the data derivation in DataBait has important implications both in terms of expectation management towards the DataBait user, as well as towards the OSN.

In order not to give the DataBait user (or the OSN) the false impression that DataBait can tell exactly what an OSN (such as Facebook, Twitter or Instagram) 'knows' about her, a 'disclaimer' will be added to DataBait in the "DataBait: how, what, why"-section. This disclaimer will also have to clarify that the 'speculative' nature of the derived knowledge is a strength of DataBait, not a weakness. It means that DataBait is forward looking, not tied to one single OSN, and that it educates the user about the inherently constructive nature of profiling practices (i.e. enhancing media literacy with regard to profiling). The constructive nature of profiling entails that not only will different machine learning algorithms analysing the same data result in quite different outputs, but even the same machine learning algorithm will generate different outputs depending on the training data used (see section 4 of D6.1, where this conclusion was reached with regard to the myPersonality dataset[2]).

The USEMP consortium has created its own original algorithms to derive additional data. The most 'high level' additional derived data are the DataBait disclosure scores: these build on 'lower level' additional derived data, that is, values (66 years old, heterosexual, Christian Democrat) on a set of personal attributes (e.g. age, sexual orientation, political beliefs, etc.). For example, if DataBait can derive the political beliefs of a DataBait user from a textual analysis her OSN post with some particular level of confidence, this impacts her disclosure score. This 'data model' (which includes the definition of a particular 'hypothesis space': see below, section 1.4, for a more detailed discussion of these terms) used to calculate the

[2] Kosinski e.a (2015). See online: mypersonality.org

disclosure scores of DataBait users is the *disclosure dimensions framework*: a hierarchical structure (consisting of a set of dimensions, attributes, values and links to data) describing which additional derived data are calculated, how they relate to each other and what they contribute to the overall DataBait disclosure score (see figure 2).

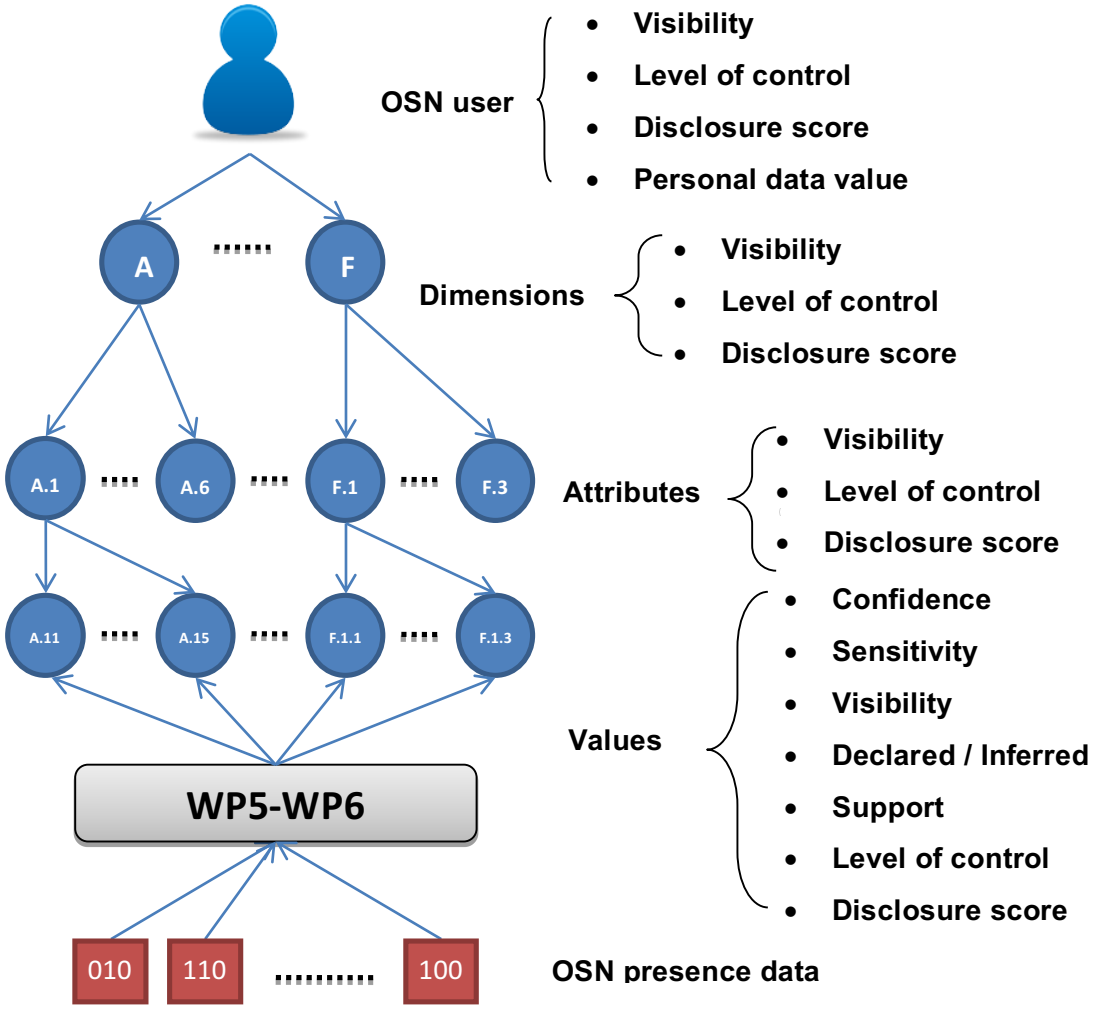


Figure 2. Overview of disclosure scoring framework (reproduced from D6.4)

This data model is original work developed within the USEMP project (see D6.4, which contains an update of the work of D6.1 and D6.2). The consortium does not have access to the internal data model used by OSNs, and fully relies on its own data model. On top of that, the consortium only has access to very limited training data in comparison to large OSNs.

Both elements (own data model/hypothesis space/algorithms and different training data) mean that the outputs ('derived data') produced by USEMP and the ones produced by OSNs (only used internally) are not comparable: DataBait will never be able to produce the same results as Facebook or another large online service operator. This is not a problem – the USEMP consortium does not strive to replicate the exact outputs of the data analysis of OSNs or to reverse engineer their exact methods: this would only result in legal problems in terms of IPRs. Nevertheless, the fact that DataBait operates based on algorithms which differ from the ones used by OSNs and which are trained on data which also differ, some overlap

in the output is to be expected: in the same way as asking fifty different people what is depicted on a particular picture (the 'input') will result in both differences and overlaps in their descriptions (the 'output'), the DataBait 'outputs' will most likely also have a partial overlap (potentially high in a number of cases) with what an OSN derives from the original user data ('inputs').

The exact text of the disclaimer in DataBait is still to be decided upon during the last year of the USEMP project, but it could be something like this:

> **DataBait shows you what *can* be extracted from your data.** The DataBait disclosure score aims to give you insight in what you disclose and gives you the possibility to have more control and make better informed decisions about what you share on a social network (OSN) like Facebook or Twitter. However, based on the information that you explicitly share or others share in relation to you (posts, pictures and likes), other information may be derived or guessed – what we call '*derived* data'. We need to make a big disclaimer with regard to these 'derived data'. DataBait only gives you insight about which information about you *might* be derived by OSNs and about which information *might* be transmitted to its commercial partners. We show you what, given the state of the art, in data technologies (called machine learning or data mining), *can* be extracted from your data. We do not claim that the information which our DataBait algorithms extract from your data is exactly or nearly the same as what is extracted by OSNs such as Facebook, Twitter, Instagram, etc. Nor do we claim that this information is actually being used. For instance, if DataBait tells you that our DataBait algorithms conclude, based on an automated analysis of the content of some of your posts, that you are homosexual, this does not mean that Facebook or Twitter have drawn the same conclusion, nor that this information is used. We don't have spies inside the walls of Facebook or Twitter, nor do we try to reverse engineer the technologies which such businesses use to analyze your data.
>
> **Does this mean that the information which our DataBait algorithms extract about you are useless or bogus?** No, definitely not!
>
> **Does this mean that the DataBait 'data derivatives are 'speculative?** Yes *and* no.
>
> **Yes:** We do not show exactly what happens but what *can* be extracted from user data, what the value *could* be of that data, what the value of your 'audience' (Facebook friends and Twitter followers with whom you interact) could be.
>
> **No:** DataBait has developed its own algorithms to derive additional information from your data. The methods we use to derive the 'data derivatives' are cutting edge methods similar to the ones which are actually used by commercial actors.
>
> The **benefit** of our approach is that DataBait is forward looking and not lagging behind technology (it would be impractical if DataBait became outdated just because the commercial actors changed their data analytics or business model a bit; and the fact that a certain bit of information is not actually extracted or used today does not mean it will not be tomorrow).
>
> **Still confused as to which part your DataBait disclosure is 'speculative' and why we call it 'speculative'?**
>
> The DataBait disclosure score aims to give you insight in what you disclose and gives you the possibility to make a better informed decision about what you disclose. **Part**

of the disclosure score is based in 'hard facts' about what you disclose. For example, if you state on your Facebook profile that your religion is Catholic and you have shared this information publicly, DataBait will alert you about the fact that you reveal sensitive information, that it is highly visible and that your level of control is high (you posted it yourself, so you can also remove it). When you write on your profile that your religion is Catholic, this is **declared information** – no further interpretation is required.

However, part of the disclosure score is based on what we call "derived data" – information that is derived (one could also say: 'extracted ', or 'inferred') from textual posts, likes or pictures. Derived data are speculative in two ways:

(1) When you derive information from some data this **always requires some level of interpretation.** Imagine, for instance, that a group of people is presented with a set of pictures and textual posts from a person they don't know and are asked: "What kind of person is this?" Not everybody will give the same interpretation. Some might say: "A happy person - because there are so many pictures of this person partying". Others might say: "An alcoholic – look at the amount of pictures where she is drinking some alcoholic beverage". Some might focus on something very different: "This is a rich person – look at the posh cars in the background of these pictures", while others might dispute that this is the right interpretation: "Posh cars in the background do not mean that they belong to this person – she might be very poor as well". In DataBait the derivation of information is not done by humans but in an automated way, by a computer (by using a number of algorithms). Like with human interpretation, **different algorithms will derive different information from the same data and even the same algorithm will result in different outputs when trained on different data.** Thus we cannot guarantee that what we derive from your data is the same as what OSNs like Facebook or Twitter derive from your data. However, like with human interpretation there is likely to be some consensus in the more obvious cases: a person who endlessly posts about the progress of her various illnesses in a depressed way is unlikely to be categorized as a "healthy, happy" person by any algorithm.

(2) An interpretation of data is never 100% certain – there is always a level of uncertainty. If you write on your profile: "Sexual preference - interested in members of the same sex" this is declared data and there is no uncertainty of interpretation. However, if an algorithm has to "guess" your sexual preference based on your likes, textual posts and or your pictures, this guess will be made with a varying level of confidence (or to put it the other way around: a varying level of uncertainty). A textual post like "So happy to be gay", might give away more than "I really enjoyed the gay pride today", which in turn is a more solid base for a guess than a picture where you embrace another person of the same sex. **This is why DataBait scores the level of confidence (a number between 0 and 1) for the information we derive from your data.** A confidence level of 1 means there is no uncertainty (which is only the case with declared data, and never with derived data), whereas a confidence level of 0 means no confidence whatsoever (a completely random guess). **In this way you get a sense of how confident we are about our guess and thus how "speculative" our guess is.**

In the disclaimer with regard to the 'speculative' nature of the data derived in DataBait further reference could be made to the DataBait disclosure scoring framework and its constituents (see deliverable 6.4).

1.4. Which parts of the 'profiling' process are covered by IPRs?

OSNs analyse data of OSN end-users in an automated way to derive additional data from them. This process of inference of additional information from original raw data is called 'profiling'. DataBait also analyses the data of OSN end-users. A first step in assessing whether DataBait infringes on any IPRs of the OSN is to describe the similarities and differences in these two profiling processes. It seems to us that there are at least five 'objects' in a profiling process which can be relevant from the perspective of IPRs: the set of training and testing data, the algorithm which is 'trained', the hypothesis space, the resulting 'trained algorithm' (or: 'predictive data model' or 'classifier'[3]), and the data analysed by the trained algorithm. Each of these 'objects' in the training process involve some element of 'labour' (sometimes 'creative' labour bearing a mark of authorship, sometimes 'just' the investment of time, money and mental energy) which the maker might wish to protect.

Let us clarify this with an example. Imagine an OSN would like to know which of its users is a smoker. To begin with, the OSN will need to define its question more precisely: does it simply want to distinguish between 'smokers' and 'non-smoker', or also between 'heavy smokers', 'occasional party smokers' and 'non-smokers'? This is the definition of the *hypothesis space*: it defines which hypotheses need to be considered. Now, let's assume that the OSN keeps its hypothesis space simple: just "smokers" and "no-smokers". This is information which is not included in the basic profile information volunteered by OSN end-users, so the OSN will have to derive this information in an indirect way, for example by analysing pictures and textual posts of the user. This means that the OSN will need some kind of 'predictive data model' to distinguish between smokers and non-smokers. Such a model would incorporate some mathematical rule that says: 'if a picture contains element x, y or z, than the person depicted in that picture is likely to be a smoker', or 'if a textual post contains elements a, b or c, then the author is likely to be a smoker'. When a human observer looks at pictures or textual posts, she might be able to make some intelligent guesses about whether somebody is a smoker: a picture where somebody is seen with a cigarette is a good indicator that the depicted person is a smoker. Similarly, a post saying "nothing beats a first smoke in the morning" is a good indicator that the author of the post is a smoker. For a human observer these inferences are not very complicated to make. However, to explain to a computer how to make such an inference is way more complex. How to explain to a computer what a cigarette looks like in a picture? And which words indicate that

[3] We will use the terms 'predictive data model', 'classifier' and 'trained algorithm' as synonyms in this deliverable. However, because in computer science the term 'data model' can also refer to the notion 'relational database', which is a particular way to organize a database, and 'classifier' is a more narrow term than 'trained algorithm', we will predominantly use the latter term.

the author has a positive attitude about smoking? Let's say that a picture can be described by two hypotheses: the first hypothesis is that the depicted person is a smoker, the second hypothesis is that the depicted person is a non-smoker. How can we teach the computer to pick the best fitting hypothesis? This is where machine learning algorithms and training and testing data come in. An untrained machine learning algorithm contains a definition of the hypothesis space (possible outcomes: smoker or non-smoker) and a mathematical 'recipe' which a computer can use to construct a predictive data model (i.e, a 'trained algorithm' or 'classifier') based on labelled examples (pictures and posts labelled by a human as representing a "smoker" or a "non-smoker"). This is called supervised learning (in contrast to 'unsupervised learning', where the algorithm is not presented with any labelled examples, but 'simply' searches for interesting patterns). An algorithm which has learned a predictive model to classify new data is a 'classifier' or '*trained algorithm*'. Such a trained algorithm can 'sieve' through *other, new data* in an automated way and categorize them (i.e. transform raw input data into derived output data). The trained algorithm is thus created by training and testing an *untrained algorithm* (this algorithm is, one could say, the 'recipe' for creating a 'data sieve') on a *data set of labelled examples*. When applied to new data the trained algorithm can predict which hypothesis is more likely ('smoker' or 'non-smoker') to be applicable.

What is the 'labour' that goes into each of the five named 'objects'? Making a hypothesis space requires some intellectual labour: which distinctions are useful? Creating a dataset which can be used for training and testing an algorithm requires the labour of labelling (e.g., 'this is a picture of a smoker') and organizing the database. Producing an algorithm which can be 'trained', that is, use training and testing data to create a 'predictive data model', requires intellectual labour, machine learning knowledge and programming skills. There are some well-known basic algorithms[4] but a particular problem (such as: distinguishing between smokers and non-smokers based on OSN pictures) will often require that such algorithms are tailored and/or combined with each other. And then there is the final result, the trained algorithm, which is constructed through the labour of fine-tuning the first three elements towards each other, until the best possible output (correct 'smoker' and 'non-smoker' classifications) are generated. Finally, somebody has to make an effort to generate new data (e.g. an OSN user posting on her wall) and organize them in such a way and format that they can be analysed by the trained algorithm (e.g. the OSN provides a structured platform which stores the OSN data in an orderly and accessible manner). Whether each of these five elements[5] could be protected by IPRs will be discussed in chapter 2 of this deliverable (where we discuss the various IPRs which could be applicable).

---

[4] Examples of such algorithms are *lineair regression*, that is, a 'recipe' to make a formula/function/line which allows you to divide a space of data points, or a *support vector machine (SVM)* which is a 'recipe' to divide a space of data points with a with a very particular type of function (namely a 'hyperplane'), or *C4.5*, that is a 'recipe' to create a particular type of decision tree to classify data, or a *neural network*, that is a 'recipe' to calibrate the weight which should be attributed to certain input in a structure of connected, layered processing units which are connected by either positive and/or negative feedback, in order to get the best possible output.

[5] It should be noted that algorithm and trained algorithm cannot always be distinguished. For example, in the kNN method (which looks at an k amount of 'nearest neigbours' to determine how to classify data) there is no 'seperate' predictive model next to the kNN-algorithm. Moreover, the hypothesis space can often be considered as an element of the algorithm. Thus, while this distinction into four elements might be a bit of a simplification, it is a useful instrument of analysis this distinction

As explained above (and further clarified below, in section 1.5), only the *overall process* of the profiling performed by DataBait is similar to the profiling process performed by large OSNs like Facebook. This *overall process* is that a certain type of machine learning algorithms is used to create predictive data models (that is, 'trained algorithms') which can categorize new data (supervised learning) or discover interesting patterns in data (unsupervised learning)[6]. That means that neither hypothesis space, training and testing data, nor the algorithms, nor the trained algorithms are the same. They only bear an *overall* likeness: both profiling processes consist out of the four aforementioned characteristic elements and build on similar types of algorithms. Important differences can be found in the purpose of the processing (in contrast to the OSN, DataBait has no commercial purpose – its purpose is scientific and aims at informing and empowering the user by showing what could be extracted from her digital trail), the hypothesis space (though some of the DataBait categories are likely to overlap with the ones analysed by the OSN), the used training and testing data, the exact algorithms, the trained algorithms ('predictive data models') and the outputs of the profiling.

Nevertheless, DataBait's profiling process may raise some issues in terms of the protection of trade secrets and intellectual property of the OSN and other third parties. In order to assess whether DataBait infringes such rights, the relevant elements of this profiling process should be examined. Firstly, DataBait copies part of the user data from an OSN, and uses this as 'input' for its own profiling process. It should thus be assessed whether these input data are protected by any IPRs. Secondly, it should be examined whether the algorithms (or more precisely: the hypothesis space, the algorithms and the trained algorithms) used by DataBait to analyse these input data are protected by any IPRs. A third point that should be studied is whether the training and testing data, which are partly derived from external databases, partly from the OSN, are protected by any IPRs. In chapter 2 of this deliverable we explore to which extent these elements of the DataBait profiling process are compatible with the protection of trade secrets, patents, copyrights and sui generis rights on databases of OSNs and other possible rights holders. Based on the analysis in chapter 2 (see Annex B for a summary of the conclusion) we conclude that overall it is unlikely that DataBait infringes the aforementioned rights, especially due to the fact that DataBait is in no way mimicking or reverse engineering software code. In addition, the DataBait user licenses the USEMP consortium to use all her data (including copyright protected materials) in the Data Licensing Agreement. Finally, should any protected use be found, it may be exempted on the basis of several research exceptions.

1.5. 'Soft' indicators that there are similarities between DataBait's 'profiling' processes and those of big OSNs

[6] For now, the majority of DataBait algorithms are supervised. However, there are some bits of unsupervised learning in WP6.

How does the USEMP consortium know that the overall DataBait profiling process is similar to the one performed by OSNs like Facebook? The answer to this question is necessarily a bit vague, because the exact profiling process of big OSNs is unknown to the consortium. In fact, if the consortium would try to get to know this, this might result in infringements on IPRs or trademark protection. Thus the only indicators for such similarities are 'soft' ones. The USEMP consortium bases itself on the state of the art in machine learning, the scientific publications of an OSN like Facebook, and the fact that the machine learning expertise of the researchers employed by such OSNs is akin to the one possessed by the USEMP consortium members. For instance, in terms of the like-based user profiling, the DataBait approach could be considered as being similar to the one used in the widely popular[7] study by Kosinski, Stillwell and Graepel (2013), though in the DataBait version some variations and additions have been tried over it (e.g. feature selection, topic-based modelling). For image-based profiling, CEA used image features extracted from extensions of the Convolutional Neural Networks (CNNs).

It is beyond doubt that Facebook has huge expertise in the area of deep learning, as attested by their very relevant publications in this field[8] and by the fact that they employ some of the most well-known researchers in the area. What the USEMP consortium does not know is whether OSNs like Facebook, LinkedIn, Twitter, Instagram, etc. actually already use these kinds of algorithms in their operational settings. However, if they do not do this currently, it is likely that such techniques will be used in the future – and that DataBait is forward looking.

[7] See http://fivethirtyeight.com/datalab/this-algorithm-knows-you-better-than-your-facebook-friends-do/ for a popularized rendition of the study.
[8] https://research.facebook.com/publications/ai/

2. Tensions between profile transparency and the rights of the profilers

2.1. The profile as subject matter protected under IPRs

Both the OSN and the USEMP consortium (through DataBait) are profilers: they profile the OSN user based on the data she provides to the OSN (and as a corollary to DataBait). Each profiler may also have one or more autonomous legal relations towards a "profile": for example, a copyright towards the profile, or a trade secret or database right towards the way in which a profile is organised. Legally speaking several legal claims could thus exist simultaneously on a profile in multiple ways: it can be a profile as defined by data protection law, a ground for prohibited discrimination, it can result from a trade secret, constitute a database, constitute or contain copyright protected works, be the object of a contract, etc. (see Figure 3).

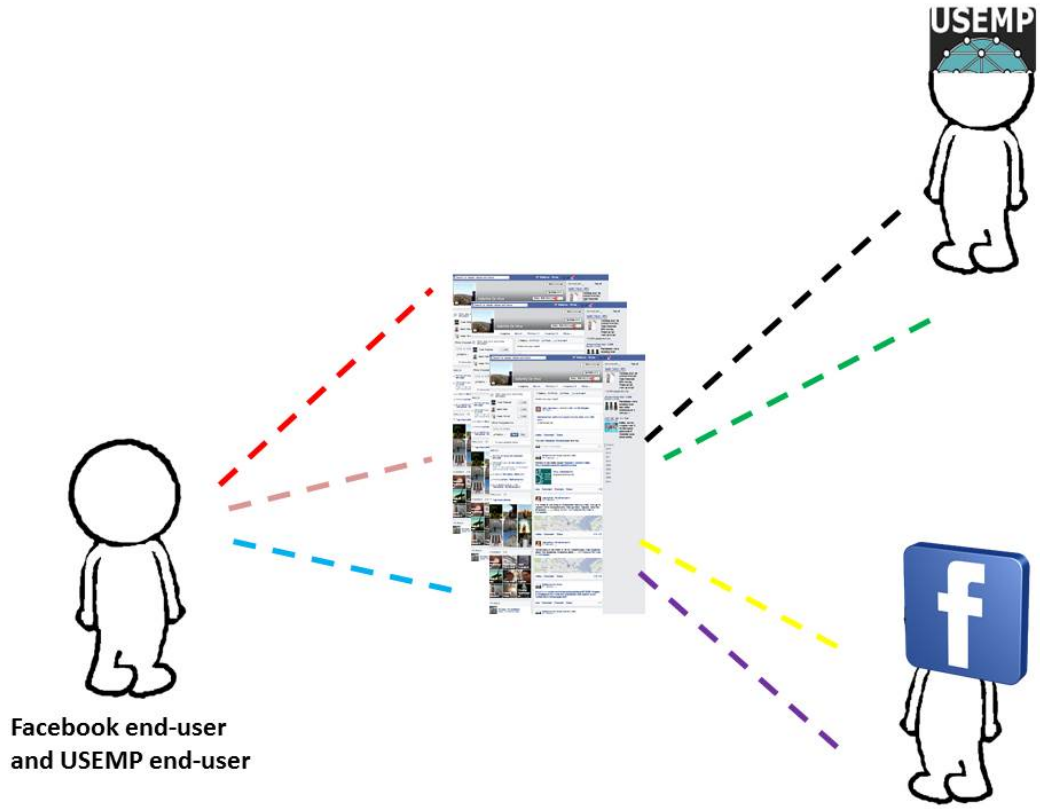


Figure3: The "same" profile can be the object of various legal relations with multiple actors.

Often several legal relations can co-exist but sometimes they will clash: for example, the IPRs of a commercial profiler who wants to protect the software used to offer targeted ads might seem incompatible with the rights of the data subject to have access to the logic of the

profiling. Anticipating these kinds of conflicts, Recital 41 of the DPD 95/46 states that although the right of access "must not adversely affect trade secrets or intellectual property rights in particular the copyright protecting the software [..] these considerations must not, however, result in the data subject being refused all information."

Even though there is quite an abundance of case law in which a balance had to be struck between an IP right and another fundamental right (for example cases involving parodies of copyrighted works, where a balance had to be struck between copyright protection and freedom of expression[9]), up until now there is no case law where IP rights in profiling and data protection law are confronted with each other[10]. This is not surprising, given the highly unclear IP status of profiles: whether a "profile" can be legally qualified as a copyrighted work, as a database protected by either copyright or the *sui generis* database right, or as the object of trade secrets is far from undisputed (Custers, 2009, section 5.3; Van Dijk, 2009 2010a, 2010b).

A first problem to be solved when asking if "a profile" can be qualified as the object of a trade secret or the aforementioned IPR, is that the noun "profile" is even more equivocal than the verb "to profile". "Profiling", as explained in D3.1, is defined in the proposed GDPR as:

> … 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*[11] [pGDPR], the successor to DPD 95/46)

---

[9] *Ashby Donald and others v. France*, Appl. nr. 36769/08, ECtHR (5th section), Strasbourg 10 January 2013 ; *Deckmyn v. Vandersteen*, C-201/13, EU:C:2014:2132.

[10] Van Dijk names three cases of which the subject matter might be extended in an analogical manner to a potential clash between IP-rights on a profile and profile transparency rights : ECHR, *Gaskin v. UK*, Application no. 10454/83, 7 July 1989 [scope of the right of access to care records kept by the public authorities with regard to the time Gaskin spent in public care during his childhood]; *Dexia*, The High Court of the Netherlands (Hoge Raad) [scope of the right of access to one's financial file at *Dexia* bank], 29 June 2007, LJN: AZ4664, R06/046HR; and *Opinion of the Dutch Data protection Authority (CBP) regarding the right of access to the raw data of a psychological test and the IP rights protecting such a test*, 15 July 2008, online available at http://www.cbpweb.nl/downloads_overig/NIP.pdf.

[11] The proposed *General Data Protection Regulation* (pGDPR) is currently being created in the so-called *ordinary legislative procedure* (formally known as the *co-decision 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>]. Currently, the amended text is examined by the Council of the European Union. If Parliament and Council disagree on a proposed legislative text, it can go back and forth between Parliament and Council up to three times. A clear infographic clarifying the ordinary

Contrary to the verb "profiling" (which is already hard to define`, see e.g.: Hildebrandt, 2008; Ferraris, 2013), there is no legal definition of what "a profile" is. However, there are two meanings which stand out: in the first place it can refer to **an individual set of characteristics** (e.g., a Facebook profile consisting of volunteered data on the frontend, but including observed data at the backend), and secondly it can refer to what could be termed an **algorithm** (see above, section 1.4, for a discussion of two constitutive elements of a trained algorithm: an untrained algorithm, which includes the definition of a particular hypothesis space, and the data used to train the algorithm) which classifies individuals according to certain traits or preferences, e.g., an algorithm which predicts a user's political preferences based on Facebook posts). The profile of an individual on an OSN can be protected under IPRs, such as copyright or database rights[12], which implies that the holder of the IPR can exercise exclusive rights on certain uses of the profile. The act of gathering data from individual profiles may result in databases. These databases themselves (as a structured unit of data) may also be subject to IPRs. Moreover, the content of these databases can contribute to the training of machine learning algorithms. The initial algorithms, the trained algorithms (which one could also call general inferred 'profiles'), the computer programs in which these trained algorithms are embedded, and the 'output' (classification of input data) of a trained algorithm (which one could call individual inferred 'profiles'), could also be subject to IPR protection.

Thus, exploring whether *profiling* amounts to an infringement on trade secrets or certain IP rights in fact entails three questions:

(1) does the **profiling process** involve infringements on intangibles that are traditionally qualified as trade secrets or the objects of certain IP rights (e.g. a set of pictures from a Facebook profile which is copied in order to make profiling possible)?,
(2) can an **individual profile** (e.g. the complete Facebook profile of a user, potentially including both user generated content and behavioural data) be qualified as a trade secret or the object of certain IP rights?,[13]
(3) can **trained algorithms** (e.g. image classifiers) and some of their constitutive elements (the untrained algorithm, training/testing data, the definition of the hypothesis space) be qualified as trade secrets or the objects of certain IP rights?

In order to answer these questions we will have to take a closer look at the notions "trade secret", "patent", "copyright", and "*sui generis* database right". The answers to these questions may have large implications for the USEMP project, because they might either support or interfere with the goal of the DataBait tools to provide profile transparency and

---

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 procedure which are still ahead of it, the GPDR will most likely enter into force by 2016.

[12] It is not very likely that a user profile be protected as a trade secret, since all information from the user is actually visible to others and thus not very 'secretive' (see below section 2.2). If, however, an OSN develops user profiles that contain behavioural data to which users have no access, such profiles will probably be kept a secret.

[13] Note that on the foreground a user profile consists of volunteered data (user generated content), whereas the profile at the backend of the system will probably also consist of observed data (machine readable behavioural data).

give insight into possible discriminatory differentiations and illegitimate negative stereotyping. For USEMP it is pivotal to know if the user rights it aims to support are “trumped” by IPR rights.

The analysis of IPRs commonly takes the following issues into consideration (i) the protected subject matter, (ii) ownership issues (first owner, transfer of rights), (iii) scope of protection: protected acts and exceptions, term of protection; (iv) enforcement. In D3.7 the main issues are whether the operation of DataBait entails *protected acts* relating to the *protected subject matter* (e.g. are there protected databases or other works? Are protected quantities of data extracted and reutilised? Is any protected element of the OSNs computer code reproduced?). The issues in D3.8 are more related to the IRPs the OSN user might be able to invoke to reinforce her position towards the OSN, including questions regarding the ownership and the valid transfer/licensing of rights by OSN users/authors to the OSN. When signing up for Facebook you sign the terms and conditions in which you agree to a number of IP issues (including the non-exclusive license to Facebook for all your IP-matter). Thus, when studying the issues related to copyrights/database rights of the users it is important to look at the terms and conditions of the agreement between the OSN and the users (Wauters e.a., 2014).

Thus, in this deliverable we explore the different types of rights that OSNs, browsers and third-party profilers might have in profiles. We discuss five possible legal qualifications with which these actors might protect the economic, intellectual and creative efforts which they have invested in ‘profiles’ of OSN and browser users: trade secrets, patentable inventions, copyrighted ‘expressions’, the IP protection of databases (through copyrights or sui geris rights) and trademarks. It should be born in mind that this analysis does not only examine how these legal means allow OSN providers and other profilers to act towards the users of their tools and services, but also towards makers of empowering transparency tools such as the Databait tools.

2.2. Profiles as trade secrets?

Let us begin by explaining what is (seemingly) the most straightforward term: a trade secret. A trade secret is in the first place the result of a factual action: it is a secret which is kept by a company in order to keep an economic advantage over competitors.

The protection of trade secrets is provided at the national level but the European Commission has proposed a draft directive to harmonise the national protection rules. This first draft defines “trade secrets” as:

> Information which meets all of the following requirements:
>
> (a) is **secret** in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;
>
> (b) has **commercial value** because it is secret;

> (c) has been subject to **reasonable steps** under the circumstances, by the person lawfully in control of the information, to **keep it secret**. (Art. 39(2) TRIPS[14]; Art. 2(1) of the proposed *Trade Secret Directive*[15] - our emphasis)

Meanwhile an amended definition has been proposed in the Report of the European Parliament (Committee on legal affairs)[16]:

> *trade secret' means **know-how and business information** which meets all of the following requirements:*
>
> *(a) is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;*
>
> *(b) has commercial value because it is secret;*
>
> *(c) has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.*
>
> ***Experience and skills honestly acquired by employees in the normal course of their employment shall not be considered a trade secret.***

The exact recipe of *Coca Cola* is an example of a trade secret. Keeping a trade secret is a practical solution which avoids the legal complexities and the high costs and publicity of a patent.

However, if a trade secret is stolen or used without the trade secret holder's consent, the law might get involved after all. It is in this stage that a judge might be called upon to decide whether something was a true trade secret or not. The TRIPS Agreement obliges Member States to provide a minimum protection for undisclosed information, including trade secrets, but there is currently no unified EU legislation with regard to trade secrets and national laws differ very much in their definitions, in the type of legislation that affords protection and the

---

[14] World Trade Organisation's 1994 Marrakesh Declaration, Annex 1C *Agreement on Trade-Related Aspects of Intellectual Property Rights* (TRIPS). The TRIPS Agreement is a multilateral agreement on intellectual property which was drafted by the World Trade Organisation and came into effect on 1 January 1995. It defines a set of minimum standards for many forms of intellectual property rights (e.g. copyrights, trademarks, and trade secrets) which binds all 158 WTO members. As such it is a very important and comprehensive instrument with regard to all kinds of IPRs. When comparing the TRIPS agreement with other important international IPR agreements, such as the *Berne Convention for the Protection of Literary and Artistic Works* ("the Berne Convention") from 1886, it is not only its extremely broad geographical reach but especially the fact that (a) it covers almost all forms of IPRs (for example, the aforementioned Berne Convention only covers copyright), and (b) that it incorporates most substantial provisions from several other important IPR agreements (such as the aforementioned Berne Convention), which makes it stand out. As such the TRIPS agreement is an extremely *comprehensive* legal IPR instrument.

[15] Proposal for a Directive of the European parliament and of the Council on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure ("Proposed Trade Secret Directive"). COM/2013/0813 final - 2013/0402 (COD). Brussels, 28 November 2013.

[16] Report of the European Parliament Committee on Legal Affairs on the proposal for a directive of the European Parliament and of the Council on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (COM(2013)0813 – C7-0431/2013 – 2013/0402(COD)) , 19 June 2015, A8-9999/2015, available on http://ec.europa.eu/DocsRoom/documents/14622. The text in bold result was amended by the EP.

scope of protection granted[17]. Member States provide protection under specific laws on trade secrets, unfair competition, intellectual property, civil law, tort law, labour law, contract law, criminal law or common law provisions.

The proposed EU Directive on Trade Secrets tries to bring more unity. The legal definition of a trade secret in the (amended) proposed EU Directive is very broad: a trade secret can be basically know-how and business information which has commercial value and provided that it can be shown that the trade secret holder (and persons lawfully in control of the information) has made appropriate efforts to keep it a secret. One cannot claim protection for something that one has not tried to keep secret by taking "reasonable steps" (technical measures, e.g. passwords, contractual and organisational measures). Futile steps or mere *pro forma* measures are not sufficient.

The broadness of the definition of a trade secret means that, for example, a trained profiling algorithm (or 'predictive data model', which can refer to one type of "profile"), but also the "training set" as structured in a relational database ("the ingredients" in their respective "containers") on which an algorithm is trained (Ateniese, 2013), the hypothesis space (definition of the possible outputs, which is an essential pasrt of the untrained algorithm) and the untrained machine learning algorithm (the "recipe" which is used to construct the trained algorithm), could very well be trade secrets. Thus, an untrained or trained algorithm, a hypothesis space and a training set all bear a likeness to a recipe such as the one for Coca Cola: while everybody knows what the approximate ingredients are, the competitive advantage is exactly in the details ("the secret ingredients", their measurement and how they interact). While the main ingredients of the Facebook news feed algorithm are well known, the specifications can be trade secrets (provided that they remain secret and reasonable measures are taken to maintain the secret character).

Under the national laws, the **scope of trade secrets protection** and the available remedies are quite divergent. Generally, the owner of the trade secret must establish that the trade secret has been infringed and that the information was used or misappropriated in an unlawful way. The specific conditions depend however on the legal instrument that the trade secret owner relies on, e.g. labour law or tort law against a (former) employee or unfair competition law against a competitor.

The proposed Trade Secrets Directive intends to harmonise the protection against the "unlawful acquisition, use or disclosure of a trade secret" (art. 3 proposed Directive)[18]. The **acquisition** of trade secrets is considered unlawful if it is carried out, without the consent of the trade secret holder, intentionally or with gross negligence by (a) unauthorised access to files under control of the trade secret holder that contain the trade secret, (b) theft, (c) bribery, (d) deception, (e) breach or inducement to breach a confidentiality agreement or any other duty to maintain secrecy, or (f) any other conduct which, under the circumstances, is considered contrary to honest commercial practices.

[17] See Baker & McKenzie 2013.

[18] The Directive has not been adopted yet – let alone transposed in the internal legal order of the Member States. Considering the divergence among the national regimes on this point, we will restrict the analysis for now to the provisions of the proposed Directive. Should more specific questions arise, we can analyse these according to the applicable law.

Furthermore, the **use** or **disclosure** of such acquired information is unlawful if it is carried out, without the consent of the trade secret holder by a person who (a) has acquired the trade secret unlawfully; (b) is in breach of a legally valid confidentiality agreement or any other duty to maintain secrecy of the trade secret; or (c) is in breach of a legally contractual or any other duty to limit the use of the trade secret (art. 3(3) proposed Directive as revised). More generally, the acquisition, use or disclosure of a trade secret is considered unlawful "in the second degree", when the user of the secret information, at the time of acquisition, use or disclosure, knew or should, under the circumstances, have known that the trade secret was obtained directly or indirectly from another person who was using or disclosing the trade secret unlawfully (art. 3(4) proposed Directive as revised). Finally, it is also considered an unlawful use of a trade secret to engage in the production, offering or placing on the market of infringing goods, or in the import, export or storage of infringing goods for those purposes, "in cases where the person engaging in such activity was, or depending on the circumstances, should have been, aware of the fact that unlawful use had been made of the trade secret" (art. 3(5) proposed Directive as revised). This last provision can be seen as an example of an unlawful use of a trade secret (under art. 3(3) of the proposed directive): not only is it unlawful to publish a trade secret belonging to another party without the latter's consent, it is also not permitted to apply trade secrets of a third party in new products. For example, if the algorithms developed by an IT solutions provider are protected as trade secrets under the contracts with its consultants, these consultants are not entitled to reveal those algorithms when they work for a competitor later on but they are also not allowed to apply these algorithms in new products of their own and commercialise these products.

In contrast, under the proposed Directive (as amended), the holder of the trade secret has no legal basis if the information is acquired in a lawful way, i.e. by independent discovery or creation, by observation, study, disassembly or test of a product or object that has been made available to the public or that it is lawfully in the possession of the acquirer of the information, by exercise of the right of workers representatives to information and consultation in accordance with European Union and national law and/or practices or by any other practice which, under the circumstances, is in conformity with honest commercial practices (art. 4 proposed Directive[19]).

Finally, the proposed Directive contains provisions that limit the rights of trade secret holders, in favour of *inter alia* the legitimate exercise of the right of freedom of expression and information or in order to address the misconduct, wrongdoing or illegal activity of the trade secret holder (art. 5 proposed Directive – art. 4 of the amended proposal[20]).

---

[19] In the draft EP legislative resolution, these considerations are moved to the definition of the scope of the Directive (art. 1(3) amended proposal); "For the purposes of this Directive, the acquisition of a trade secret shall be considered lawful when obtained by any of the following means: (…)".

[20] The text is amended slightly :

« Member States shall ensure that there shall be no entitlement to the application for the measures, procedures and remedies provided for in this Directive when the alleged acquisition, use or disclosure of the trade secret was carried out in any of the following cases:

*(a) for making legitimate use **in accordance with the Charter of Fundamental Rights of the European Union** of the right to freedom of expression and information, **including media freedom**;*

*(b) for ▌revealing a ▌misconduct, wrongdoing, **fraud** or illegal activity, provided ▌that the respondent acted in the public interest; (…) »*

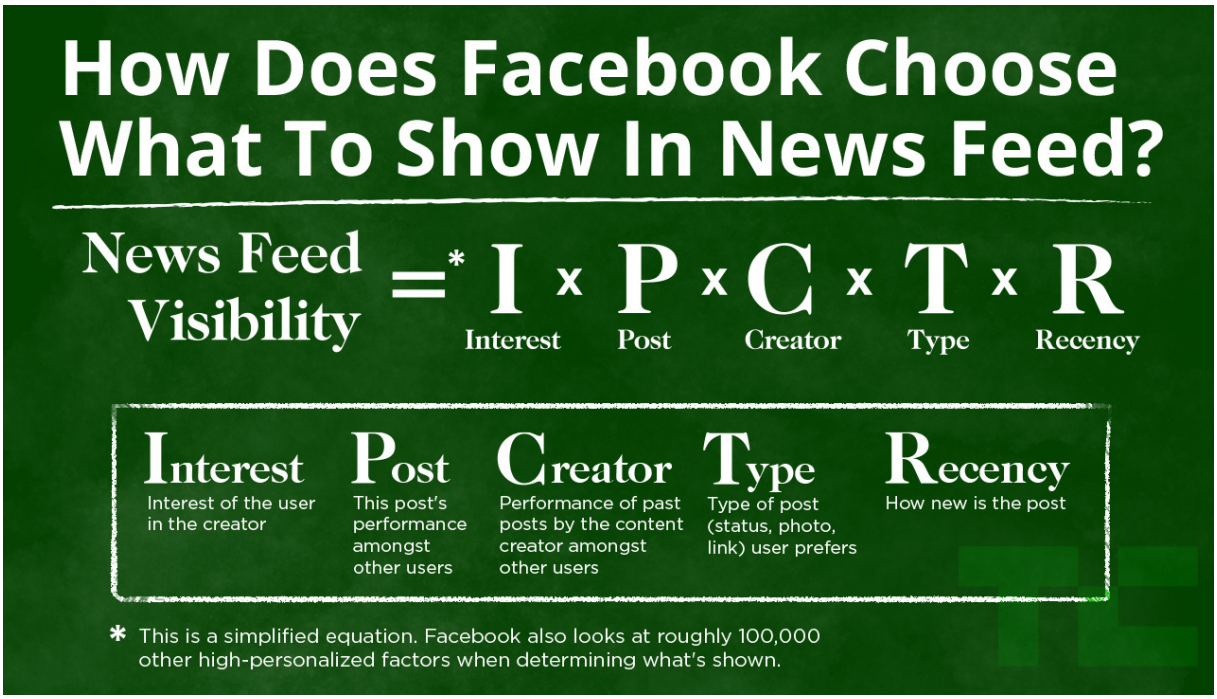


Figure 4: A simplified depiction of the algorithm which is used by Facebook to decide what is shown in the News Feed of a user. Image source: <http://techcrunch.com/2014/04/03/the-filtered-feed-problem/>

If we apply these rules to (i) the profiles and (ii) the act of profiling, we come to the following provisional conclusions.

Regarding the data derived from an individual OSN profile: as long as this profile is just a small amount of (relatively) publicly accessible "raw data" (volunteered data) it is not very likely that it would qualify as a trade secret. Firstly, this is the OSN profile that is published by the OSN user, hence it is not secret. Secondly, taken on their own, these profiles are not likely to represent a commercial advantage because of their secret nature. The OSN user cannot claim protection for her profile as a trade secret.

To the extent that the OSN processes individual user profiles (e.g. by adding machine-readable behavioural data) the OSN might claim indeed a trade secret. The combination of volunteered, behavioural and inferred data contained in the individual profile that is only accessible to the OSN thus results in valuable know-how that could qualify for protection as a trade secret, notably when an individual Facebook profile contains historical data which neither the Facebook account holder nor others can see. Such information could have commercial and technical value and, if Facebook takes reasonable steps to keep these data secret, then Facebook could indeed claim protection of this information, as a trade secret. Since Facebook offers access to such data via a documented API and regulates access to this API, one might be inclined to say that Facebookcannot be said that it takes steps to keep these data secret[21]. However, things might be a bit more complicated because one also has

[21] Though a bit of nuance could be added here: there could still be protected trade secrets if Facebook would add a contractual condition to keep the secret (nondisclosure). If an API is freely available there are probably no "reasonable" steps to keep it secret. If an API only made available after having accepted a secrecy obligation, this could be seen as 'reasonable' steps to keep the data secret. In the

to consider contractual clauses regulating the use of an API. If a browser operator or OSN operator makes APIs freely available without specifying any non-disclosure clauses, the condition of "reasonable steps to keep a secret" are not met. Does Facebook specify any non-disclosure clauses that could qualify as "reasonable steps" to preserve secrecy? This is not completely clear. Facebook's policy for app developers contains the following clauses that might be interpreted as some form of non-disclosure clause:

> 6. Keep private your secret key and access tokens. You can share them with an agent acting to operate your app if they sign a confidentiality agreement.
>
> 7.If you use any partner services, make them sign a contract to protect any information you obtained from us, limit their use of that information, and keep it confidential.
>
> 8.Keep Facebook user IDs within your control. Contract with any providers who help you build or run your app to ensure that they keep the user IDs secure and confidential and comply with our policies. If you need an anonymous unique identifier to share with third parties, use our mechanism.
>
> 9.Don't sell, license, or purchase any data obtained from us or our services.
>
> 10.Don't transfer any data that you receive from us (including anonymous, aggregate, or derived data) to any ad network, data broker or other advertising or monetization-related service.
>
> 11.Don't put Facebook data in a search engine or directory, or include web search functionality on Facebook.
>
> 12.If you are acquired by or merge with a third party, you can continue to use our data only within your app.
>
> 13.If you stop using Platform, promptly delete all user data you have received from us (absent explicit consent from people). You can keep Account Information if you have presented your privacy policy within your app.
>
> (https://developers.facebook.com/policy/#data; clauses 3.6-13, last accessed 1 Dec 2015)

The question is whether any or several of the above clauses should be considered as nondisclosure clauses which are "reasonable steps" to preserve secrecy that indicate that the data should be treated as trade secrets. Thus it depends on the interpretation of these clauses whether the data should be considered a trade secret.

Obviously, next to the data which are provided through the Facebook API, an OSN like Facebook might also possess data that are kept fully secret (and these are clearly objects of trade secrets)but the USEMP consortium does not refer to those. Moreover, the algorithms applied to the data to derive more data constitute important know-how, which the OSN will want to keep secret. The USEMP consortium does not have direct access to such algorithms. Even if one would reconstruct such algorithms on the basis of observations (which USEMP does not do) this should not be considered unlawful.

---

case of Facebook, use of the API has to be approved by Facebook. We didn't see any contractual nondisclosure clause in Facebook's conditions but we will look closer into this in the final year of the USEMP project.

So what about extracting large amounts of data from a browser or a social network site? If such data can be used to train a competitive algorithm, they are likely to have commercial value, along with the precise training method and the analysis of the results. However, the crucial question will be to which extent the commercial value results from the fact that the data are secret. Moreover, an OSN relying on the protection of its trade secrets should also demonstrate that (i) the information is not generally known (which is probably not the case) and the data are not "readily accessible to persons within the circles that normally deal with the kind of information" (cf. definition in the Proposed Directive) and (ii) the information has been subject to reasonable steps to keep it secret. Thus, here it is important whether the extraction is authorised/enabled by the browser/OSN. A large data set can be the object of a trade secret if it is kept secret, but in order to infringe on it this would require that the extraction is in some way illegal (e.g. hacking into a database) or uses the data in ways not permitted by the OSN policy.

When creating profile transparency tools like those developed in USEMP it is important not to infringe on trade secrets.

Because the inferred knowledge presented in the USEMP tools is based on untrained and trained machine learning algorithms developed within the USEMP consortium, it seems highly unlikely that the USEMP algorithms could be qualified as an infringement on a trade secret. A more obvious risk of infringing trade secrets (depending on the applicable law) is present, rather when one would try to hack into protected information, to obtain secret information by illegally accessing secured systems or by manipulating employees or service providers to gain access to such information. It is unlikely that the ongoing research of the USEMP consortium amounts to an infringement of Facebook's trade secrets. USEMPwill not approach Facebook employees to share secret information in breach of their confidentiality agreements. Instead, the inferences made in USEMP are hypothetical ("this is the kind of information which *could* be extracted from your data trail and this is what it *could* be used for"). After all, these inferences are based on independent discovery, observation and study (see art. 4(1) proposed Directive, art. 1(3) amended proposal).

Thus, while the USEMP *algorithms* are not likely to infringe on trade secrets, there is a possibility that the use of data used for constructing these algorithms do infringe on a trade secret. This would be the case if the OSN policy contains non-disclosure clauses. Even if a policy does not say explicitly: "you're not allowed to use our data to build an competitive algorithm which will give you a commercial advantage", one could imagine that an OSN would claim before a court that policy clauses requiring non-disclosure indicate that the data are to be considered a trade secret and that using them to create a commercial advantage is thus not permitted. If an OSN was to invoke trade secret infringement, the legitimate exercise of the right to freedom of information and expression could be invoked in defence (art. 4(2) proposed Directive, art. 4(a) amended proposal). The contradictory interests would have to be balanced against each other by a court.

2.3. Profiles as patentable inventions?

As an alternative to keeping a data model or profiling algorithm as a trade secret, one could also try to patent it. Can data models and profiling algorithms, such as the ones used within

DataBait or large OSNs such as Facebook or Twitter, be patented? And if this first question were answered affirmatively: is it possible that DataBait infringes on an existing patented invention? Finally, the question should be raised if DataBait could and/or should be patented.

To answer these three pertinent questions we first have to look at the protected subject matter of a patent. Article 52(1) of the *European Patent Convention* (EPC) reads:

> "European patents shall be granted for any inventions, in all fields of technology, providing that they are new, involve an inventive step, and are susceptible of industrial application."

Thus, according to this article, there are four conditions for a successful patent application: the object of the patent has to be an invention, which is a novel and inventive (i.e., non-obvious to a person skilled in the art) product, apparatus or a process that has industrial applicability (i.e. if it can be made or used in any kind of industry, including agriculture).

Whether algorithms (both 'untrained' algorithms, including the part of the algorithm defining a particular hypothesis space, and 'trained' algorithms) that are written down in computer code (software) can be patented is a highly contentious legal topic, which is further complicated by differences between patent law in, for example, the US and the EU. Profiling algorithms or profiling data models are specific types of software. Patents on software are in general more broadly accepted in the US than in European jurisdictions. For example, the famous Google *PageRank* algorithm is a patented invention within the US[22], while it is not patented in Europe. Looking at the text of the European Patent Convention, which names computer programs and mathematical methods explicitly as being *outside* the scope of the protected subject matter (Art. 52(2a) and (2c) EPC), this is not surprising.

Art. 52. Patentable inventions.

(1) European patents shall be granted for any inventions which are susceptible of industrial application, which are new and which involve an inventive step.

(2) The following in particular shall not be regarded as inventions within the meaning of paragraph 1:

a. discoveries, scientific theories and mathematical methods;
b. aesthetic creations;
c. schemes, rules and methods for performing mental acts, playing games or doing business, and programs for computers;
d. presentations of information.

[22] *U.S. Patent No. 6,285,999 on a method for node ranking in a linked database*, invented by Lawrence Page, assigned to Stanford University, filed for on January 9, 1998. See : http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=6,285,999.PN.&OS=PN/6,285,999&RS=PN/6,285,999.

(3) The provisions of paragraph 2 shall exclude patentability of the subject-matter or activities referred to in that provision only to the extent to which a European patent application or European patent relates to such subject-matter or activities as such.

However, this does not mean that untrained algorithms (which can be described as a software expressing a mathematical 'model' for creating a trained algorithm) and trained algorithms (which can be described as software expressing a 'predictive data model') can never be patented within the EU – it only means that software (computer programs) or mathematical models 'as such' cannot be patented. Consequently, software or a mathematical model which is *not* 'as such', but functional to a technical solution *can* be the object of a patent:

> "…computer languages or codes are considered computer programs as such and receive copyright protection. The technical solution to a technical problem that a computer program may provide is not considered to be the computer program as such, but refers to its function. If it has a technical function or "character" it is patentable as an invention." (Custers, 2009, p. 48)

In practice, the European Patent Offices grants patents to "computer implemented inventions" (in contrast to "computer programs as such"), a criterion that is not easily applied. Thus, while algorithms and data models might under certain circumstances be patented within in Europe, their patentability depends on whether they are merely computer programs or mathematical models 'as such' or whether they are 'functional' to a technical solution to a technical problem.

Large OSNs, like Facebook, have several patents and patent applications in Europe (as well as in the US) on various aspects of the complex functioning of the OSN. Why would an OSN want to patent such inventions? The answer to this question can be found by looking at the *protected acts* with regard to a patented invention: the common denominator in most European jurisdictions is that the exclusive right of a patent holder entails that she can exclude others commercially making, using, selling, importing, or distributing a patented invention without permission – and that this permission can be given in exchange for a financial remuneration. The rationale of patenting law is technological innovation benefits from the openness of the patent, while at the same time ensuring that an inventor can economically benefit from her invention. OSN providers, such as Facebook, often use a combination of copyright, patent and trade secret protection to maximise protection of its services.

So, if an inventor of an 'algorithm-related invention' would like to patent it at the European Patent Office, how would she proceed? She would probably patent it as a "computer implemented invention" – which is distinguished from a "computer program as such". A "computer implemented invention" involves "the use of a computer, computer network or other programmable apparatus, where one or more features are realised wholly or partly by means of a computer program."[23] A computer implemented invention can be a hybrid between software and hardware, i.e., "system and methods"[24], or merely consist out of

[23] https://www.epo.org/news-issues/issues/software.html
[24] Four examples of algorithm-related inventions patented by Facebook in this way :

software. The implementation in hardware ("system") is not decisive: in the *Vicom* case[25] (European Patent Office [EPO], Decision T208/84; OJ EPO 1/1987, 14) the Technical Board of Appeal held that:

> "... a claim directed to a technical process which process is carried out under control of a program (be this implemented in hardware or in software), cannot be regarded as relating to a computer program as such ... it is the application of the program for determining the sequence of steps in the process for which in effect protection is sought".

In recent case law[26] the European Patent Office (EPO) has confirmed the restrictive interpretation of 'as such'[27], which means that many software programs are in fact patentable (if they have a technical character and can thus be qualified as 'computer implemented inventions'). Moreover, software programs and mathematical methods which are part of a hardware invention – and are thus not 'as such' - might also fall within the scope of patent law (provided they fulfil the other requirements of functionality within a technical solution, novelty, inventiveness and industrial applicability).

According to the EPO's case law, a computer program is not excluded from patentability if « the computer program resulting from implementation of the corresponding method is capable of bringing about, when running on a computer or loaded into a computer, a "further technical effect" going beyond the "normal" physical interactions between the computer program and the computer hardware on which it is run »[28]. Importantly, "schemes, rules and methods for (...) doing business" are not patentable; « but a new method which solves a technical, rather than a purely administrative, problem may indeed be patentable »[29].

This is particularly relevant for, e.g., artificial neural networks, which are often a hybrid of hardware and software, and may thus indeed be patentable elements under the EPC, since they may provide technical solutions and have a technical character. It is also interesting to

---

(a) "Systems and methods for identification based on clustering" (http://ep.espacenet.com/publicationDetails/biblio?DB=ep.espacenet.com&II=54&ND=3&adjacent=true&locale=en_EP&FT=D&date=20150408&CC=EP&NR=2858013A1&KC=A1) or,
(b)"Systems and methods for providing privacy settings for applications associated with a user profile" (http://ep.espacenet.com/publicationDetails/biblio?DB=ep.espacenet.com&II=83&ND=3&adjacent=true&locale=en_EP&FT=D&date=20100210&CC=EP&NR=2150885A1&KC=A1), or
(c) "Performing actions based on metadata associated with objects in a set of objects associated with a social networking system user" (http://ep.espacenet.com/publicationDetails/biblio?DB=ep.espacenet.com&II=101&ND=3&adjacent=true&locale=en_EP&FT=D&date=20140820&CC=EP&NR=2767946A1&KC=A1), or
(d) 'Targeting social advertising to friends of users who have interacted with an object associated with the advertising' (http://ep.espacenet.com/publicationDetails/biblio?DB=ep.espacenet.com&II=117&ND=3&adjacent=true&locale=en_EP&FT=D&date=20131023&CC=EP&NR=2652690A1&KC=A1)

[25] https://www.epo.org/law-practice/case-law-appeals/recent/t840208ep1.html (accessed 1 Nov 2015).
[26] See for an overview, e.g.: http://en.swpat.org/wiki/Software_patents_exist_in_Europe,_mostly
[27] Art. 52(2a), (2c) and (3) EPC posit that mathematical methods and programs for computers are not patentable *as such*.
[28] http://www.epo.org/news-issues/issues/software.html.
[29] http://www.epo.org/news-issues/issues/software.html.

note that the distinction between computer science and electrical engineering that seems to underlie the restrictions of the EPC, is crumbling, as wearables, sensor-technologies, and the Internet of Things integrate with back-end systems that include neural nets, thus further hybridizing software and hardware.

As stated above, it is not sufficient to demonstrate that a computer program or an algorithm has a technical nature and is not excluded as an invention.The 'computer implemented invention' should also meet the conditions of novelty, providing an inventive step and industrial applicability. It seems that the required "inventive step" raises the most important hurdle, software implemented invention being assessed following the "problem-solution approach"[30].

In the context of USEMP this is relevant, because this means that patents may exist in the research field where DataBait is being developed. Where other actors have patented (parts of) their solutions, it is theoretically not excluded that DataBait is affected. Making a thorough check if this is indeed the case goes beyond the resources the USEMP consortium has. A fully-fledged patent check would require a dedicated team of lawyers and engineers who check all patents (not just the ones patented by large OSNs such as Facebook or Twitter, but by *any* inventor) and published patent applications that could possibly overlap with DataBait. Moreover, the vast majority of the technical work done by the USEMP consortium is based on methods that are described in scientific publications, which by definition are not patented.

Yet, the possibility cannot be excluded that an OSN could try to use its patent rights to oppose the development, offer and use of transparency tools (such as DataBait). However, the chances that an OSN could effectively oppose DataBait based on patent protection are limited. In most jurisdictions the exclusive rights of a patent holder covers only *commercial* use of the patented invention. The practical implications of this limitation for a profile transparency tool which uses patented technology and is controlled by a non-commercial entity, depends on how 'commercial use' is defined in a particular jurisdiction. Moreover, many European national patent legislations contain a research exception.[31] Such a research exception to patent protection entails that the patent holder cannot prevent the use of the invention when the use is for scientific purposes. The USEMP consortium seems to meet this condition.

---

[30] Three questions are asked : (i) determining the "closest prior art", (ii) establishing the "objective technical problem" to be solved, and (iii) considering whether or not the claimed invention, starting from the closest prior art and the objective technical problem, would have been obvious to the skilled person. Only the technical elements contributing to the inventive step should be considered ; when the computer implemented invention is essentially a non-technical creation, no inventive step will be established. See Janssens 2011.

[31] See e.g. Belgium, Art. 28 (b) of the Belgian Patent Law ('Wet op de uitvindingsoctrooien'): 'The rights following from a patent do not cover actions on and/or with the object of the patented invention, which have a scientific purpose' ('De uit een octrooi voortvloeiende rechten strekken zich niet uit tot : handelingen die op en/of met het voorwerp van de geoctrooieerde uitvinding worden verricht, voor wetenschappelijke doeleinden'). Online available at : http://www.ejustice.just.fgov.be/cgi_loi/loi_a1.pl?language=nl&caller=list&cn=1984032835&la=n&fromtab=wet&sql=dt=%27wet%27&tri=dd+as+rank&rech=1&numero=1#Art.27quinquies> 28 <http://www.ejustice.just.fgov.be/cgi_loi/loi_a1.pl?language=nl&caller=list&cn=1984032835&la=n&fromtab=wet&sql=dt=%27wet%27&tri=dd+as+rank&rech=1&numero=1#Art.29

Finally we raise the question if DataBait could be patented.

Patenting brings along two characteristics which may render it an unattractive option for the "inventor" of a data model or profiling algorithm. Firstly, patenting an invention requires a considerable investment of time, money and work to file a patent (contrary to copyright, which comes into existence automatically whenever a work is created). Moreover, as we have discussed, the outcome is uncertain. Secondly a successful patent application requires that the invention is disclosed to the public in a patent document which contains a description of the invention explaining how the invention is made, how it functions, and the 'claims' which define what the inventor seeks to protect with her patent. Incidentally, such publicity reduces the possibilities for the inventor to rely on trade secrets (although a clever patent agent knows how to limit the publication to what is necessary for the patent application to succeed, while keeping valuable know-how secret).

Incidentally, it can be observed that the publication of the invention suggests that the rights on a patented data model or profiling algorithm are unlikely to interfere with the right to profile transparency. It could be argued that a data subject who is subjected to such patented profiling could simply access the patent and read about the logic underlying the profiling. In reality, the technical language used in a patent description will not be easy to grasp for every user. The claims and the description in the patent documents could be useful for technically skilled providers of transparency tools as a basis for a more comprehensible and easy description of the logic underlying the profiling. However, some patents are formulated in quite general and cryptic terms, so even a technically skilled person might find it difficult to understand the exact functioning of such 'algorithmic invention'. Clearly, one could question if a general (and, sometimes, cryptic) description suffices to provide the necessary profile transparency. In fact it might be more interesting to know which specific data of the data subject have been used: the usefulness of a general description of the logic underlying the profiling will largely depend on whether the data subject has access to such additional information. On the other hand, any provider of transparency tools should be careful not to develop a solution within the scope of the patented invention.

As far as DataBait is concerned, the answer to the patenting question is: probably not. One of the requirements for a patent is the novelty of the invention. Given the fact that DataBait is already used online, it is unlikely that the USEMP consortium (or somebody else) could patent DataBait. However, as we will show in the following section, USEMP partners can claim *copyright* on the code they write as part of DataBait.

2.4. Profiling and copyright?

2.4.1. How does copyright function in the context of profiling?

In this section we examine, firstly, how copyrights in content created by an OSN user can be opposed to anyone who wants to mine this content without consent and what kind of license an OSN provider, like Facebook, or a profile transparency tool provider, like the USEMP consortium, would need to be able to mine this user generated content. We study this question from the perspective of the profiler. In deliverable D3.8, which is complementary to

this one (D3.7), we also look at the question of IP licensing of user generated content but from the opposing perspective – we focus on the copyrights and personality rights on the side of the user of a social networking service and explore whether she can invoke these rights to strengthen her legal position.

A second question explored in this section is whether an OSN operator, which holds copyright on elements constituting the OSN (e.g., the graphic user interfaces, computer programs, databases and user generated content which has been licensed to the OSN), could rely on these exclusive rights to prohibit transparency efforts.

In order to address these two issues we first have to define what copyright is. Copyright is, like patents, *sui generis* data base rights or trademarks, an intellectual property right. Contrary to "ordinary" property rights, IPRs are not based on something "material" but on an "intangible" product of the mind like a particular expression (copyright) or invention (patent). Being the *owner* of a book only means that one owns the book as a "material object" and does not imply that one also has the IPRs on the novel contained by the book, or that one is entitled to copying the book, to sharing it with one's friends or adapting it into a play or a film (though exceptions are often made for sharing within a small set of people).

The **subject matter** protected under copyright is not explicitly defined but indications can be found in various legal instruments, such as the Berne Convention, the 1996 WIPO Copyright Treaty and, at EU level, the Directives in the field of copyright. These international and European instruments harmonise the national copyright legislations of the Member States of the EU. Copyright can offer protection for diverse types of creations in the literary, scientific and artistic domain, including books, theatre plays, operas, music and lyrics, dance choreographies, press articles or scientific publications (art. 2 BC). Moreover, computer programs are considered literary works and therefore protected under copyright (art. 4 WCT; art. 1 CPD[32]) and certain aspects of a database may also be protected under copyright[33].

Copyright cannot protect a mere idea (e.g., a guy and a girl fall in love with each other but their respective families have a feud), but only on a particular expression of an idea (Shakespeare's *Romeo and Julia* is a very unique *expression* of the aforementioned idea, as are the subsequent (and more recent) adaptations for theatre and cinema).

> Copyright protection extends to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such.[34]

Some differences may subsist among Member States in the definition of the "work", i.e. the protected subject matter of copyright, but following the decision of the CJEU in *Infopaq I* one can say that in order to be a protected under copyright, the subject matter should be "**original**" in the sense that it is its author's own "intellectual creation"[35] and reflects the

---

[32] **Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (Codified version),** *OJ* L 111, 5.5.2009, p. 16–22 (hereafter CPD).

[33] **Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases,** *O.J. L 077 , 27/03/1996 P. 0020 – 0028 (hereafter DBD);*

[34] Art. 2, *WIPO Copyright Treaty*, adopted 20 December 1996, Geneva.

[35] Judgment in *Infopaq International A/S v Danske Dagblades Forening*, C-5/08, ECLI:EU:C:2009:465, para. 37.

author's personality[36]. More specifically, this is the case if the author was able to express his or her creative abilities in the production of the work by making free and creative choices[37].

In the context of profiling, two types of copyrights are relevant: copyrights held by those subjected to profiling and copyrights held by profilers. In OSNs, users are the subjects of profiling. These users often hold copyright in, at least part, of the content they post. Users of OSNs may post elements that are reflect their personalities (criterion of "originality") and are consequently protected by copyright. Such "user generated content" can consist out of status updates (in as far as it is not just a factual statement like "It is warm in London", but contains some basic touch of 'authorship'), pictures – even "selfies" – videos or music. As a matter of principle, if this content is reproduced or communicated to the public (e.g. by providers of transparency tools such as DataBait), the user's prior consent is required. In its general terms and conditions Facebook requires the user to grant a broad licence, which could mean that Facebook is entitled to exercise copyright rights on content submitted by its users (this aspect is discussed in D3.8).

The second category consists of the copyright held by profilers (which includes OSNs and other commercial profilers but also scientific or non-profit profilers like the USEMP consortium). Profilers may claim copyright protection for, for example, the graphic user interfaces of the system, the computer programs used to profile, operate the databases and perform the data analyses, the structure of databases and user generated content which has been licensed to them by users (although this may be disputable).

2.4.2. What kind of IP license does a profiler need to 'mine' copyright protected content?

Copyright was initially meant to protect authors from certain forms of exploitation of their work without their consent, commonly expressed as the acts of "reproduction" or "communication to the public" (art. 2 and 3 of Directive 2001/29). An author of a novel who holds the copyright over it has the right to prohibit its reproduction (i.e., copies without the author's consent are 'pirated' copies – unless some other exception or limitation applies). Similarly, immaterial forms of exploitation are protected, e.g. live performances in presence of an audience, broadcasting, the "publication" on a website or the massive transmission over peer-to-peer networks. This way copyright allows authors, or anyone who is licensed by the author, to exploit the fruits of copyright protected content. Other uses – typically private uses – are not restricted under copyright (e.g., copyright does not prevent anyone from reading a copyright protected work).

In the last decades an avalanche of new technologies and corresponding new business models has stretched the scope of copyright protection to all kind of new fields of application: e.g., the copies ("reproductions") made by "search engines, either for indexing, for the

[36] 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 ;*

[37] Judgment in *Eva-Maria Painer v Standard VerlagsGmbH and Others*, C-145/10, ECLI:EU:C:2011:798, para. 89.

display of thumbnails in search results or for the dissemination of news articles"; "the use of works in 'user created content'"; copies made in "cloud computing"; or the copies made in "data mining" (Van Der Noll e.a. 2012). From a technical perspective the "copies" made in these new fields of application are still copies and consequently "reproductions" (cf. CJEU decision in Infopaq I), even if they are technically different from the "pirated copies" of a novel[38] and despite the fact that the function of these copies and the modes of exploitation differ fundamentally. Moreover, uses that are not restricted in an analogue world (reading, retrieving ideas rather than copying their concrete expression) risk being protected in the digital world, because of digital technologies are indeed based on "copies" (in the technical sense).

Copyright in the EU is currently based on wide notions of reproduction and communication to the public and an exhaustive list of exceptions (art. 5 Directive 2001/29), cover inter alia certain uses in private circles, for scientific purposes or new expressions such as parodies. New forms of use are not always easily squeezed in the existing list of exceptions.

This large and technical understanding of "reproductions" raises the question if certain of these practices need to be excluded from copyright protection or at least treated in a different way. At the level of the EU[39] this had also led to the intention "to adapt copyright rules to new technological realities so that the rules continue to meet their objectives". Some scholars have pleaded for creating a more flexible copyright protection and expanding on the list of existing exceptions on copyright, or make an open-ended list of exceptions:

> "[A] decisive argument against an exhaustive list of limitations, is that a fixed list of limitations lacks sufficient flexibility to take account of future socio-economic and technological developments. A dynamically developing market, such as the market for online content, requires a flexible legal framework that allows new and socially valuable uses that do not affect the normal exploitation of copyright works to develop without the copyright owners permission, and without having to resort to a constant updating of the Directive, which might take years to complete". (Van Der Noll e.a. 2012, p. 7)

The issue with copyright and analyzing data in an automated way ("profiling" or "data mining") fits into the pattern of copyright problems with other digital technologies: mining data is based on massive copying of data, including possibly copyright protected works. These copies resulting in protected reproductions, the data miner or profiler should then demonstrate that her practice falls within one of the exceptions provided in Directive 2001/29 – as implemented in the applicable national law. This will be the case for certain profiling practices, in particular when the (national) exceptions for temporary acts of reproduction and for scientific purposes are found to apply (Triaille, 2014). If it cannot be established that the

[38] For example, the fact that a search engine makes the copies of images which are typically of much lower quality (thumbnails) or copies of text which are merely summarized verisons of the original text does not exclude these copies from copyright law.

[39] 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Towards a modern, more European copyright framework' Brussels, 9.12.2015, COM(2015) 626 final, p. 3.

mining falls within one of the exceptions, the data miner or profiler must obtain the right holders' prior consent – absent which there will be a copyright infringement.

During the last years especially scientists have opposed publishers' analysis that a licence is required for the text and data mining (TDM) of copyright protected content or databases. This has given rise to the argument voiced in the Hague declaration[40] that "the right to read is the right to mine":

> "One of the key principles recognized in the [Hague] declaration is that intellectual property law does not regulate the flow of facts, data, and ideas–and that licenses and contract terms should not regulate or restrict how an individual may analyze or use data. It supports the notion that "the right to read is the right to mine", and that facts, data, and ideas should never be considered to be under the protection of copyright. To realize the massive, positive potential for data and content analysis to help solve major scientific, medical, and environmental challenges, it's important that intellectual property laws and private contracts–do not restrict practices such as text and data mining"[41].

In its recent Communication the Commission also acknowledges the problem with too many copyright restrictions with regard to text and data mining for scientific purposes:

> "The need to better reflect technological advances and avoid uneven situations in the single market is also clear with text-and-data mining (TDM), through which vast amounts of digital content are read and analysed by machines in the context of science and research. The lack of a clear EU provision on TDM for scientific research purposes creates uncertainties in the research community. This harms the EU's competitiveness and scientific leadership […]." (Commission 2015, p. 7)

The problem is exacerbated by the fact that other jurisdictions outside the EU are much more permissive in this respect:

> "Copyright comes into play because text and data mining begins with the unavoidable organisation of data so that it can be analysed. It is the subject of fierce debate whether, for researchers, this act of 'organisation' amounts to copying within the meaning of copyright law. In Europe, some Member States have already adopted an exception or limitation to copyright rules applying generally to academic research, but this exception is both uneven in its application and less permissive than the legal regime in the United States, where the 'fair use' defence appears to offer significantly greater comfort to researchers about what they can and cannot do without fear of provoking successful legal action from rights holders. With its reference point of the

[40] http://thehaguedeclaration.com/
[41] http://creativecommons.org/tag/tdm

> First Amendment to the US Constitution, forbidding any abridgement of the right to free expression, and its explicit reference to scholarly research in its 'fair use' doctrine, American jurisprudence in copyright continues to evolve in a more permissive direction, from the point of view of researchers." (p.12)

Making TDM easier for researchers within the EU can be achieved in three ways (Hargreaves e.a., 2014). The first way can be achieved by the proprietors of copyrights and sui generis database rights: the owners of these exclusive rights could grant liberal licenses (allowing TDM for research purposes). The second way would be through a change in copyright legislation, namely by creating an exception for TDM for research purposes. If the legislator was to create a copyright exception for TDM for scientific purposes, this would be important for research projects like USEMP. The third way would be up to the courts and/or the legislator: by making the notion of "reproduction" more restrictive (so that the copies made for TDM fall outside the scope of copyright and database law).

> "The reproduction right in copyright law, as the right of extraction under the database regime, has traditionally received a broad interpretation encompassing any direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part of his/her work. After years of expansive interpretation, it seems timely to ask whether this broad interpretation of the reproduction/extraction right should be reconsidered. Instead of a functional approach to the reproduction/extraction right where all acts of reproduction or extraction that are technically possible fall within the scope of the owner's exclusive right, the legislator could take a normative approach and only recognise protection for acts of reproduction or extraction that actually entail an act of 'expressive' exploitation. Is TDM a form of copyright or database exploitation that should be under the control of the rights owner? Is TDM (in all its forms) an act of reproduction (and eventually of communication to the public) that affects the interests of the rights owner?" (Hargreaves, 2014, p. 53)[42]

Would such a normative interpretation of the word "reproduction" mean that the commercial exploitation of data would also fall outside the scope of copyright protection? And that OSNs like Facebook, Google plus and LinkedIn could mine their users data without users giving them a broad IP license when signing up for the service (as is currently the case)? This depends on the normative interpretation, but most likely not:

> "[Th]e normative approach to the definition of the right of reproduction/extraction [could be]: if an act of reproduction of a work gives rise to no exploitation of that work, then this act of reproduction should not fall under the control of the rights owner." (Hargreaves, 2014, p. 53)

[42] Such exception has been adopted in the UK Copyright, Designs and Patents Act 1988 (section 29A).

The legislator could take the normative stance that the fact that a profiler *exploits* – in whatever way – copyright protected works (even as part of aggregated data) would entail that the copies made for the mining/profiling are still considered as "reproduction" in the sense of copyright law. The legislator could also differentiate between different types of exploitation. While the normative stance is not very common, differentiating between types of exploitation, as we will argue below, might also be useful in a completely different context, namely when drafting licensing conditions.

Let's first take a closer look at the exploitation model of commercial profilers (which includes large OSNs like Facebook, LinkedIn, Google Plus, etc.). Such commercial profilers exploit "data" and "content" at an aggregated level (and not at the level of the individual content or based in the 'originality' of such work). Arguably, such exploiters use digital content, including copyright protected content, but their model is evidently different from the exploitation model of, let's say, a film distributor or music label. Thus, the exploitation model of commercial profilers covers copyright protected content but in a sense it is not founded on the original character of the creations. They are interested in the copyright works as "data" or "driving "data traffic". For example, a picture made by an OSN user of her breakfast cereal is not appreciated by the OSN for its "originality", nor exploited individually by the OSN. However, such picture has value for the OSN on an aggregated level and as an attractor of data traffic for targeting and profiling practices (e.g. the likes and the comments of the friends of the OSN user). If a legislator considers taking a normative stance towards the notions of "reproduction" (copyright protection), it will also have to consider whether protected elements are reproduced when a work it is not exploited for its originality, when it is not used as a work "as such" and has little value as a copyright work.

This question has not received a lot of attention because large OSNs, in contrast to researchers who want to mine IP protected content or databases, have had no legal difficulties getting extremely broad IP licenses from their users, allowing them to use the user generated content in whatever way they please – including reproduction, extraction and reutilization for profiling purposes[43]. If we assume for now that such general licence is valid, then the OSN may argue that it has acquired the OSN users' copyrights and can prevent developers of transparency tools (who are also miners and profilers) from doing the same. If the OSN requires a licence for its mining activities, alternative miners can be expected to do

[43] This could change if the far-reaching non-exclusive licenses granted on the basis of a non-specific clause in the general terms and conditions of an OSN (without defined object, scope of rights, duration) were challenged from the perspective of consumer and contract law (Wauters e.a, 2014). This would require of profilers (not only commercial ones but also scientific or non-profit ones like the USEMP consortium) to reconsider the exact form of an IP license needed to enable reproduction, extraction and re-utilization for profiling purposes. In the case of a commercial profiler the type of exploitation might become relevant. Under many European copyright laws, copyright licenses with the author have to meet certain requirements (as a matter of substance or for evidence purposes) such as specificity with regard to the intended uses. The ratio for such specific copyright contract rules is generally to offer more protection to the author, who is considered the weak party in a negotiation with a professional party that will commercially exploit the work. The fact that OSN providers exploit IP-protected content in a way which differs from "traditional" exploiters (they exploit the content on an aggregated level) and offer their OSN to the user/author without requesting a fee, might shift the equation in favour of the OSN provider, resulting in a lesser protection for the user/author. We will examine this argument in further detail in D3.8.

the same. It is however unlikely that an OSN can exercise such prohibitive rights on the basis of a use licence (as far as copyright on the users' content is concerned). By contrast, the OSN may exercise its database rights in this sense (*infra*).

In the deliverables relating to empowering IP rights on the side of OSN user (D3.8 and D3.12) we explore this issue in more detail.

2.4.3. An IP license granted to USEMP by DataBait users

In the hypothesis that the development and operation of DataBait entails restricted acts (reproduction) relating to protected content and no exception applies to such activities, then USEMP should seek the consent of the rights holder of the used content. If the users have validly transferred their copyrights in the posted content to Facebook, then Facebook could grant or refuse such licence. If the copyrights have not been validly transferred by the Facebook general terms and conditions, then the consent of the individual rights holders (not necessarily the user) should be acquired. This would entail many practical complications.

Thus it is crucial to establish whether a valid transfer of the copyright by the OSN users to the OSN has taken place. In Article 2 of the *Facebook Statement of Rights and Responsibilities*[44] (version of November 15, 2013) every Facebook user gives a non-exclusive, transferable, sub-licensable license to Facebook. This means that a Facebook user continues to be the copyright holder over her own IP content[45] and that she can license others next to Facebook (the license is non-exclusive)[46]. Facebook acquires a licence to use the user's content but does not become the sole owner of the rights to the posted content. The question in this section is then whether this general IP clause entails that Facebook's prior consent is required for the acts of reproduction (to a lesser extent communication to the public) performed for the construction of a transparency tool (absent an exception for data mining[47]).

Considering that the general IP clause in Facebook's general terms and conditions provides a non-exclusive licence, it could be argued that USEMP does not need Facebook's consent to process protected content from users (or third parties). A licence from the Facebook user to USEMP is then sufficient, Facebook cannot prohibit the creation of the DataBait tools on the ground that its prior consent is required (in addition to its users').

In the USEMP Data License Agreement (see D3.6) signed by every USEMP user and the USEMP Consortium Partners, a license is given to the USEMP consortium to use all data gathered through the DataBait Facebook app and the browser plug-in for the specific purpose of USEMP research. Moreover, in the last year of the USEMP project we will include a specific clause in the DLA in which the DataBait user licenses the USEMP consortium for

---

[44] Online available at: <https://www.facebook.com/legal/terms>.
[45] The matter is more complicated where users share works to which they do not hold the copyright, such as pictures, news articles or videos.
[46] Such licences raise many legal questions on the relation between the user and the OSN. These will be addressed in D3.3.
[47] Trialle et al (2014),122
http://ec.europa.eu/internal_market/copyright/docs/studies/1403_study2_en.pdf.

the use of her copyright works. The exact formulation of this clause is based on the outcomes of the analysis (see D.3.8 and above, previous section) with regard to the best way of formulating the licensing conditions for profiling.

Considering that the IP licence is transferable and sub-licensable, could Facebook's consent be sufficient? This depends on whether the licence from the user to Facebook is valid in the first place. Should it be considered that the Facebook user gives a valid licence to Facebook on the basis of the general terms and conditions, then Facebook is entitled to use the copyright works for its own mining and profiling activities. Moreover, it may then also be entitled to sublicense this right (as provided in the general terms and conditions) to a subcontractor. In that case, Facebook can (implicitly or explicitly) authorise third parties (such as app developers to use the users' works, as a form of sublicensing. Whether this is allowed on the basis of the general terms and conditions is not answered by the Art. 9 (*Special Provisions Applicable to Developers/Operators of Applications and Websites*) or Art.10 (*About Advertisements and Other Commercial Content Served or Enhanced by Facebook*) of the *Facebook Statement of Rights and Responsibilities*[48] (see the Annex for the full text of these two articles).

2.4.4. Can an OSN oppose profiling transparency based on copyright?

Any OSN will present several elements that qualify for copyright protection. Firstly, the presentation of the OSN may be protected under copyright, in particular the graphic user interfaces. In addition, its computer programs (i.e. source code, object code, and interfaces) are likely to be original and therefore protected under copyright. Its databases may enjoy some degree of protection under copyright as well. In short, any OSN deals with copyrights on various types of creations and from various sources (its own creations, user contributions, third party content shared by users). Furthermore, protected subject matter from other sources than the OSN may be used during the development of the DataBait tools, e.g. images that are part of "training sets". It has been verified in D3.4 and in D3.9 whether such third sources are used and on which legal basis (exception, consent).

The **scope of protection** of copyright is determined by the exclusive rights granted and the exceptions. Holding the copyright over a work means, according to the *EU Copyright* or *"Infosoc" Directive*[49], to hold the right over its reproduction, communication to the public and distribution. The Computer Programs Directive[50] provides rights of reproduction, adaptation and distribution (art. 4 CPD). The Database Directive provides rights of reproduction, adaptation, distribution and communication to the public (as far as copyright protection is concerned).

---

[48] Online available at : <https://www.facebook.com/legal/terms>.

[49] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, O.J. L 167, 22/06/2001 P. 0010 – 0019.

[50] **Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (Codified version) (Text with EEA relevance)** OJ L 111, 5.5.2009, p. 16–22.

1. OSN's creations

It is unlikely that the USEMP consortium infringes any rights to the computer programs developed by the OSNs. The USEMP consortium has developed its own computer programs in an independent way. It has not had access to the OSN computer programs and has not attempted to reverse engineer their computer programs, hence no infringements of copyright on OSN software are to be expected.

As far as the computer programs of the OSN are concerned, we verify in D3.4, D3.9 and D3.13, based on the technical description of the development and use of the DataBait tools, whether any protected part of the computer programs running the OSN will be used and, if so, an exception can be relied on.

Based on our consultation with the technical partners in the USEMP project, it is unlikely that any parts of an OSN's graphic user interfaces will be reproduced. However, given that the fact that the final DataBait visualizations are still under development, we will continue to closely monitor that no elements of the graphic user interface of OSNs are reproduced. Considering that GUIs are not protected under the Computer Programs Directive but as other copyright works (cf. CJEU's decision in *BSA*), it should be verified (at a later stage) whether any exception provided in the InfoSoc Directive can apply. At this stage of the USEMP project, it is likely that the exception for scientific purpose can apply.

2. OSN user's creations

The DataBait tools copy user generated content (like posts and pictures) in order to analyse it. This probably constitutes an act of reproduction in the sense of copyright law. The circumstance that the copy is not a lasting one (it is discarded once the processing is done) does not have an impact on this qualification; it may however affect the application of the exceptions.

The reproduced and annotated content (e.g., "this information can be derived from your post") is then communicated to the individual DataBait user. This individual communication may be qualified as "communication to the public", which includes the making available of works for transmission to individual members of the public (such as DataBait users.. We will explore this in more detail in the next version of this deliverable (D3.11).

The exceptions to the exclusive rights are listed in art. 5 InfoSoc Directive. During the phase of **development** of the transparency tool, several exceptions may exempt the USEMP partners from acquiring the right holders' prior consent. The exceptions for temporary acts of reproduction (art. 5(1) InfoSoc Directive) and for scientific research may be interesting for the USEMP consortium (art. 5(3)(a) InfoSoc Directive[51,52]. It should however be verified in the applicable national law which exceptions have been transposed and under which conditions.

[51] Article 5 Exceptions and Limitations (…)
3. Member States may provide for exceptions or limitations to the rights provided for in Articles 2 and 3 in the following cases:
(a) use for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author's name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved;

[52] The InfoSoc Directive does not provide an equivalent exception to the exception for decompilation (in view of developing an interoperable software) in the Computer Programs Directive (art. 6 CPD).

Considering the condition that the exception only covers the use of works "to the extent justified by the non-commercial purpose to be achieved", it is unlikely that the exception provides a legal basis for commercial usage of the DataBait tools.

However, because each DataBait user will provide USEMP with an IP license in the DataBait agreement, infringement of the users' copyright is avoided. The licence from the rights holder (OSN users) should cover all the USEMP activities.

In summary, an OSN may hold copyright on various aspects of the OSN. Firstly, there are the copyrights on its own creations (it can be assumed that these have been acquired from the creators such as employees or consultants), such as computer programs, interfaces and perhaps also databases (see next section). Secondly, the OSN users post protected works (their own works or third party works, with or without consent). Facebook, for example, has provided a general IP clause in the form of a non-exclusive, transferable and sub-licensable licence. In D3.4 and its successors, D3.9 and D3.13, we verify on the basis of the technical description of the DataBait tools (development and operation of the tools) which elements are processed and reproduced. For each element it should be verified (i) whether the consent of the right holder is required or if an exception applies (so no consent is required), (ii) who holds the exclusive rights and is authorised to consent (cf. D3.8) and (iii) whether such consent can be acquired (USEMP's data licence, Facebook's implicit licence, or an alternative solution).

DataBait is extremely unlikely to infringe on the copyright on user generated content, considering firstly the applicable exceptions for temporary acts of reproduction and use for scientific research and secondly the licence granted by DataBait users in the DLA.

2.5. Profiling and the IP protection of databases

Up until now we have focused on copyrighted content that is part of one's profile (a status update, a video, a picture, etc.). In addition, the profile as a whole could be the subject matter of another layer of intellectual property protection. The Member States of the European Union indeed provide protection for databases, following the adoption of the Database Directive (DBD)[53].

A database is defined as "a collection of independent works, data or other materials arranged in a systematic way or methodical way and individually accessible by electronic or other means" (Art. 1(2) DBD).

The DBD provides a two-tier protection for databases: the database may be protected under copyright (structure) or the "sui generis" protection on the content of the database.

Firstly, there may be **copyright** protection for databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation (Art. 3 DBD). It is important to underline that in such a case the copyright is not on the content of the database (one particular status update or one individual profile) but on its particular structure

[53] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases ("Database Directive"), *Official Journal* L 077, 27/03/1996 P. 0020 – 0028.

("selection or arrangement"). The structure of the database can be protected under copyright provided that it meets the originality requirement, i.e. it is the author's own intellectual creation[54]. It can be reminded here that protection under the Database Directive does not extend to the algorithms or computer programs used to make or operate the database (art. 1(3) DBD).

Holding a copyright over the structure ("expression") of a database gives the author of the database the right to permit or prohibit reproduction, publication and distribution (Art. 5 of the Database Directive).

Article 5. Restricted acts

In respect of the expression of the database which is protectable by copyright, the author of a database shall have the exclusive right to carry out or to authorize:

(a) temporary or permanent reproduction by any means and in any form, in whole or in part;

(b) translation, adaptation, arrangement and any other alteration;

(c) any form of distribution to the public of the database or of copies thereof. The first sale in the Community of a copy of the database by the right holder or with his consent shall exhaust the right to control resale of that copy within the Community;

(d) any communication, display or performance to the public;

(e) any reproduction, distribution, communication, display or performance to the public of the results of the acts referred to in (b).

Secondly, next to the classical copyright protection of databases, there is also a ***sui generis* database right** in favour of the maker of the database (art. 7 DBD). Such protection is available for databases provided that there has been qualitatively and/or quantitatively a substantial investment, either in the obtaining, or in the verification or the presentation of the contents. The investment in the creation of the content is not taken into account[55].

A substantial investment…

> "… may consist in the deployment of financial resources and/or the expending of time, effort and energy." (Recital 40 of the DBD)

Where a substantial investment in the obtaining, verification or presentation of the contents of the database can be demonstrated, the maker of the database has an exclusive right covering the extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database (art. 7 DBD). In the next version of the report, depending on the further development of USEMP tools, we may elaborate an analysis of these exclusive rights, based on cases decided by the CJEU on re-utilisation/extraction (e.g. the Sportradar-case).

---

[54] CJEU, Football Dataco Ltd and Others v Yahoo! UK Ltd and Others, C-604/10, ECLI:EU:C:2012:115.

[55] See *inter alia* Fixtures Marketing Ltd v Oy Veikkaus Ab, C-46/02, ECLI:EU:C:2004:694 ; The British Horseracing Board Ltd and Others v William Hill Organization Ltd, C-203/02, ECLI:EU:C:2004:695.

A database can simultaneously be protected by copyright (protecting the author from unauthorized reproduction, adaptation, communication and distribution of the database structure) and by the sui generis right (protecting the maker of the database from to unauthorized extraction and/or re-utilization of the whole or of a substantial part of the database).

The copyright and sui generis right on databases is of particular interest to the USEMP project – do profile transparency tools like the ones created by USEMP reproduce (parts) of the overall structured way in which data are organized by, for example, Facebook? After all, we cannot be sure that Facebook will not invoke exclusive database rights. Although Facebook does not invest in the creation or verification of the content of the database per se (this is added by the users), it arguably makes substantial efforts for the presentation of the content. It could also be argued that the structure of the database shows a certain degree of originality (cf. the subsequent changes to the presentation of the user's profiles, e.g. "walls", "timelines", "newsfeeds"). In this case, it is not the Facebook user who decides what her profile looks like; she uses the mould defined by Facebook.

DataBait collects user data in two ways: through a browser plug-in and a Facebook application ('app')[56].

The data collected through the browser plug-in are not structured in any way, so copying of these data coming from the user on the DataBait server (at the HWC premises in Lancaster) will not infringe any database rights.

Next to the collection of data through the browser plug-in, DataBait also collects data through a Facebook app. A Facebook app is a computer program created by a third-party programmer (in this case: the USEMP consortium) which runs on the Facebook platform and which Facebook users can choose to add (or remove) to their account. What is characteristic of a Facebook app is that it is not hosted by Facebook and that it is an optional extension of the features of Facebook: it enables the user to perform and/or allow certain actions which do not belong to the 'basic' package offered by Facebook itself. For example, there are Facebook apps which can enable a user to play a game (with Facebook friends) and post the results on the user's wall, apps that allow the user to post music on her wall and let others purchase it, apps that enable the user to share movie ratings on her wall and apps which review the content of the user's wall for malicious content. Everyone who develops a Facebook app, has to submit the app for review[57] to Facebook before it goes 'live'.

> "In order to use Facebook Login in your app and access additional elements of a person's Facebook profile, you will need to submit your app for review. If your app is

[56] The USEMP consortium is also exploring the possibility to deliver profile transparency about other OSNs than Facebook – this could, of course involve that data is collected through other means and that the Data Licensing Agreement has to be adjusted accordingly.
[57] https://developers.facebook.com/docs/facebook-login/review

> not approved or you don't submit for review, people will not be able to use Facebook Login in your app."[58]

The DataBait app will be submitted for review in the course of 2016. When submitting an app for review to Facebook, a developer has to specify which Facebook data would be needed to make the app to function. One of the functions of the review process is to ensure that apps do not ask for more data than they actually need. As Facebook explains on its developer site:

> "People tell us that some apps ask for too many permissions. To address this, we're extending our existing App Center and Open Graph review process to login. During login review, we'll look at and approve any permissions that an app requests beyond public profile, email and friend list."[59]

An important review criterion of Facebook is 'utility'[60] of the requested data and writing permissions: app developers only are allowed to access data ('read permission') and post things ('write permission') on users' walls if this is of direct use for the app.

Next to the individualized review, Facebook also has certain general rules (which they regularly change) about the permissions they grant. For example, in 2014 Facebook removed the possibility to request access to the full list of friends of a user. An app developer can now only ask to see friends of a user who are already using this same app. Thus, while it is not expected that the DataBait app is rejected during the Facebook review process, it is possible that Facebook allows less permissions than the ones requested by the USEMP consortium. In the worst case scenario the permissions are so limited that it would interfere with a proper functioning of the DataBait app (see below for the way in which USEMP would deal with this scenario).

The data which the consortium gets through the Facebook app, in contrast to those collected through the browser plug-in, are in some way structured by Facebook (the OSN) and could thus be protected by both the copyright in the database structure or sui generis right of the OSN. However, in as far as the data one gets through the API Facebook are based on the explicit permissions to access certain data (and the structure in which they are offered), the USEMP consortium cannot be said to infringe on either the copyright in the database structure or sui generis right of the OSN.

If Facebook would not grant DataBait the requested permissions (or if the permission would be too limited, that is, only for a very small subset of the data that we require to make DataBait work), a work-around would be needed. A possibility in this scenario would be to collect Facebook data through the browser plug-in. In this way, the withheld permissions of Facebook would be circumvented. Because the DataBait user licenses the USEMP consortium to use her data in the Data Licensing Agreement, this would be legitimate. The browser plugin method has some drawbacks, however: collecting data through the Facebook

---

[58] https://developers.facebook.com/docs/facebook-login/review/what-is-login-review
[59] http://www.adweek.com/socialtimes/login-review-update/439294
[60] https://developers.facebook.com/docs/facebook-login/review/what-is-login-review

application programming interface (API) is technologically a much easier and smoother way, and the data collected through the browser plug-in do not cover any historical data from the OSN profile (the browser plug-in only captures data which are posted after the user has signed up for DataBait). The latter issue could be circumvented by asking the user to download her historical data and send the PDF to DataBait in order to be analysed – but this is, obviously, a rather clumsy method, requiring lots of additional effort from the user (downloading data from the OSN, uploading data to DataBait) and from the consortium (transforming the data in the right format). Moreover, in as far as the historical data are organised in a particular way, the copyright or sui generis right on databases might apply, prohibiting any extraction or re-utilisation / reproduction, distribution or communication to the public of elements.

Here it is important to note that the European Court of Justice (CJEU) in recent case law[61] uses a rather broad interpretation of what constitutes a database under Database Directive 96/9, thereby guaranteeing that investments in the information society (by producing a database) are well protected. Consequently it is important to take the possibility that a tool like DataBait could infringe on a database right (the copyright or sui generis) when extracting data structured by an OSN seriously.

However, as far as the sui generis database right is concerned, the reutilisation and extraction are only protected if "the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database" is used in this way. This is not necessarily the case. The USEMP consortium does not extract or reutilise the whole content of Facebook's databases and, considering the fairly small number of DataBait users, it is unlikely that even a substantial part of their contents are used.

Moreover, in the case of the USEMP project, both the regimes of copyright and sui generis right provide exceptions with regard to scientific research: reproduction (copyright) and extraction or re-utilization of substantial parts of a database (sui generis right) for the sole purpose of scientific research[62] to fall under the exceptions in Art 6(2b) and Art. 9(b) of the Database Directive.

Article 6 of the Database Directive

Exceptions to acts restricted by the copyright on a database

1. The performance by the lawful user of a database or of a copy thereof of any of the acts listed in Article 5 which is necessary for the purposes of access to the contents of the databases and normal use of the contents by the lawful user shall not require the authorization of the author of the database. Where the lawful user is authorized to use only part of the database, this provision shall apply only to that part.

2. Member States shall have the option of providing for limitations on the rights set out in Article 5 in the following cases:

(a) in the case of reproduction for private purposes of a non-electronic database;

[61] CJEU, *Verlag Esterbauer*, C-490/14, 29 October 2015 ; CJEU, *Ryanair*, C-30/14, 15 January 2015.
[62] See for a more nuanced and detailed discussion: Traille et al., 2014.

(b) where there is use for the sole purpose of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved;

(c) where there is use for the purposes of public security of for the purposes of an administrative or judicial procedure;

(d) where other exceptions to copyright which are traditionally authorized under national law are involved, without prejudice to points (a), (b) and (c).

Article 9 of the Database Directive

Exceptions to the sui generis right

Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorization of its maker, extract or re-utilize a substantial part of its contents:

(a) in the case of extraction for private purposes of the contents of a non-electronic database;

(b) in the case of extraction for the purposes of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved;

(c) in the case of extraction and/or re-utilization for the purposes of public security or an administrative or judicial procedure.

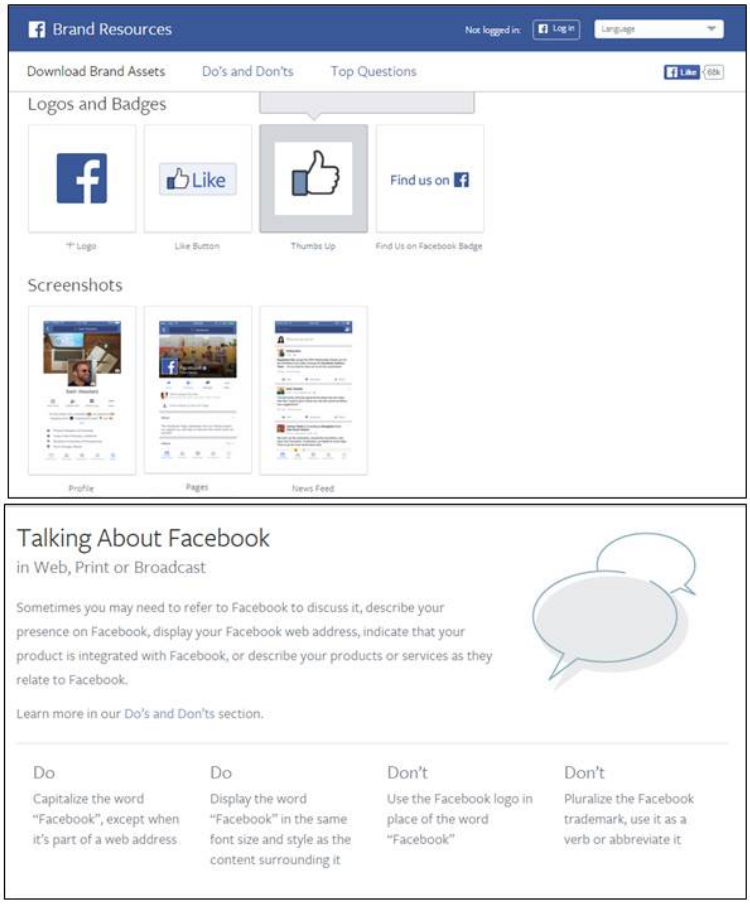
There are two caveats about the protection offered by the three aforementioned exceptions for scientific research.

Firstly, the exceptions are optional – not every Member State has opted to implement them in their national legislation[63]. Secondly, tools similar to the ones developed by USEMP which are used outside a scientific context are more likely to infringe database rights. The merit of any hypothetical infringement claims will be examined in more detail, should the OSNs not consent to USEMP's use of the data (cf. API discussions).

[63] Triaille e.a. (2014) studied the implementation of the scientific exceptions in the following member states : Netherlands, Germany, Poland, Luxembourg, Denmark, Hungary, Belgium, Spain, the UK and Italy. "The exception to copyright for scientific research in relation to databases contained in Article 6(2)(b) of the Database Directive has been implemented in four Member States among those considered in this Study: Belgium, Spain, the UK and Italy. […] Other Member States – the Netherlands, Germany, Poland, Luxembourg, Denmark and Hungary – have not implemented the exception for scientific research to the copyright protection of databases contained in Article 6(2)(b) of the Database Directive". (p. 68); "The exception to the sui generis right for scientific research contained in Article 9(b) of the Database Directive has been implemented in nine countries among those considered in this study: Belgium, Spain, the UK, the Netherlands, France, Germany, Poland, Luxembourg, and Hungary." (p. 80); "Except for Spain and the Netherlands, the exception for scientific research contained in article 5(3) a) of the Infosoc Directive has been transposed in all the Member States that are analyzed by the Study" (p.53). This study did not concern Swedish law.

2.6. Can a profile transparency tool infringe on trademarks?

A profile transparency tool, like DataBait, always provides a certain kind of profile transparency (namely transparency about the content of the user's digital trail, what can be extracted from it, trackers and the 'audience' of the user) with regard to some *particular* other internet service. In developing the tool the USEMP consortium has focused on providing transparency with regard to a large OSN, such as Facebook, Twitter or Instagram. We have used Facebook as an exemplary case, but the tool could be adjusted to other OSNs. Because the object of a profile transparency tool is another internet service, it is unavoidable to mention or refer to this service within the tool – however, it is important to ensure that the way of 'mentioning' or 'referring' to this other service (e.g., Facebook, Google plus or Twitter) does not infringe on trademark rights. Each of these services have protected trademarks and are quite serious about the protection of their brand, providing extensive guidelines[64] on how to correctly refer to their brand (see e.g. Figure 5). The most direct and established form of trademark infringement is to offer services or products under another's protected brand. This is obviously not the type of infringement that the USEMP consortium would risk committing.



[64] For example, Facebook : https://www.facebookbrand.com/ ; Google : https://www.google.com/permissions/trademark/ ; Twitter : https://about.twitter.com/company/brand-assets

Figure 5: Facebook's policy on using their brand (excerpts from https://www.facebookbrand.com/, last accessed 1 Nov 2015)

However, Facebook's rules ('Do's and Dont's') with regard to the use of their brand also cover how Facebook should be referred to in situations which are not about unfair competition or trademark confusion: "Sometimes you may need to refer to Facebook to discuss it, describe your presence on Facebook, display your Facebook web address, indicate that your product is integrated with Facebook, or describe your products or services as they relate to Facebook" (see Figure 5).

The USEMP consortium does refer to the brand names of OSNs (protected as trademarks), for example, in the explanatory DataBait animation (https://www.youtube.com/watch?v=dJinztt5PrA).

Generally, the holder of a registered trademark is entitled to prevent all third parties, who do not have his consent, from using in the course of trade an identical sign for identical goods or an identical or similar sign for identical or similar goods if the "use of that sign without due cause takes unfair advantage of, or is detrimental to, the distinctive character or the repute of the trade mark" (art. 5 Directive 2008/95, art. 9 Regulation 207/2009). For trademarks with a reputation, protection extends even to non-similar goods and services.

The USEMP consoritum does not use the Facebook trademark to distinguish the goods and services it offers (it uses the name DataBait instead). However, some Member States also restrict the use of a trademark for other purposes than distinguishing goods and services (so not "as a trademark"). In that case, it should be avoided that the "use of that sign without due cause takes unfair advantage of, or is detrimental to, the distinctive character or the repute of the trade mark" (art. 5(5) Directive 2008/5). The fundament right to freedom of expression should be considered in this latter case.

An infringement of the Facebook trademarks is unlikely in the case of how DataBait refers to OSNs like Facebook. No goods or services are offered under a sign even remotely similar to the protected trademarks. Instead, the trademark is used to explain its functioning and its impact on its users and to explain the use and functioning of the DataBait tools. In conclusion we think that the likelihood that DataBait infringes on the trademark of an OSN is low. A priori we do not see it necessary to examine this matter further.

3.Conclusion and next steps

IPRs and trade secrets of OSNs are sometimes presented as being a possible obstruction to profile transparency.

Firstly, trade secrets could be used to obstruct transparency. A data subject who wants access to the data that are processed about her and to the logic of the profiling to which she is subjected, might be faced with a data controller who is reluctant to disclose the 'secret recipe' (trade secret) used to profile users of a service.

The empowering 'profile transparency' offered by DataBait about the digital trail of a user on an OSN does not suffer from this problem. DataBait is an independent actor giving insight in the overall way in which profiling functions – which makes the data derivatives presented in DataBait 'speculative' (DataBait does not claim that an OSN like Facebook infers and/or uses the same derived data as the one generated by DataBait; nor does DataBait claim to use exactly or nearly the same methods/algorithms). DataBait shows what is technologically possible considering the SotA in machine learning and conceivable considering the business models of OSNs and their expertise. This will be explained in a disclaimer within DataBait. Thus, a first legal requirement following from this deliverable is a disclaimer in DataBait.

Because DataBait functions in an independent way, using methods and algorithms developed by the USEMP consortium itself (not trying to acquire trade secrets, nor reproducing any patented or copyrighted software), this makes copyright or trade secret infringements on protected OSN software unlikely. Whether OSN data, which can be accessed through an API, are protected by trade secret will depend on whether any contractual non-disclosure clauses exist that could be interpreted as constituting reasonable steps to keep these data secret. If the latter is the case, an infringement would be possible if the data were to be used to create a commercial advantage. This is an issue we will continue to monitor to avoid any infringement.

The input for the analyses made by DataBait are user data collected through a browser plug-in and/or an OSN app. Copyright issues are not to be expected because the DataBait user, who is the author of her copyright works, licenses DataBait in the Data Licensing Agreement. However, more work is needed with regard to the exact formulation of the IP clause in the Data Licensing Agreement. The licensing conditions are drafted keeping in mind the function and implications for the copyright owner of the 'act of reproduction' for the profiling in DataBait (no commercial exploitation, reproduction is functional to the profiling process and serves no purpose of the exploitation of the works belonging to the user). Consequently, the second legal requirement following from this deliverable is the formulation of the IP clause in the DLA.

While the risks are not very high, it is not excluded that DataBait's functioning would infringe copyrights and/or sui generis rights on databases owned by the OSN, copyrights on the graphic user interfaces of the OSN. This is something we will continue to monitor in the last year of the project.

During the last year of the USEMP project we continue to keep track of the possible IP issues that we discuss in this deliverable and which could obstruct DataBait's functioning as a transparency tool. Once the DataBait app has been reviewed by Facebook (in the course

of 2016), and the division of which user data will be read through this app and which ones will be collected through the browser plug-in, implications for DataBait in terms of IP law will be considered again based on the latest information about the DataBait architecture.

With regard to sui generis database right, it will be considered whether a substantial part of a protected database is used. It may also be useful to analyse the exception for temporary acts of reproduction in more detail. We will also extend the research into an analysis of the legal compatibility with IP rights of profilers of the DataBait tools in a commercial market, as examples of Data Protection by Design, developed and provided by commercial or non-profit data controllers. Since the exceptions for scientific research may not apply in that case, a further analysis is required into the extent to which such tools may violate copyright, sui generis database rights or trade secrets. Another question we will continue to explore is what the best way is to handle the issue that some of the content on the wall of an OSN user, which is collected through DataBait, might contain copyright works of third parties which have not licensed the USEMP consortium. We might also investigate further whether individuals could base deployment of transparency tools on the need to provide effective tools to exercise freedom of information. Finally, we will study the newly adopted Regulation with regard to data protection and see if this changes anything with regard to the balance between profile transparency and IP rights.

Bibliography

Ateniese, G., Felici, G., Mancini, L. V., Spognardi, A., Villani, A., & Vitali, D. (2013). Hacking smart machines with smarter ones: how to extract meaningful data from machine learning classifiers. arXiv preprint arXiv:1306.4447.

Baker & MacKenzie, Study on trade secrets and confidential business information in the internal market, Study prepared for the European Commission by Baker & McKenzie, 2013, available at http://ec.europa.eu/internal_market/iprenforcement/docs/trade-secrets/130711_final-study_en.pdf.

Commission (2015), 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Towards a modern, more European copyright framework' Brussels, 9.12.2015, COM(2015) 626 final.

Custers, B. (2009). Profiling in Financial Institutions. FIDIS (The Future of Identity in the Information Society) deliverable 7.16 (pp. 57-67). Brussels: EU.

Ferraris, V., Bosco, F., Cafiero, G., D'Angelo, E., & Suloyeva, Y. (2013). Defining Profiling. Working paper on definition and domain of application of profiling. Profiling. Protecting Citizens' Rights Fighting Illicit Profiling: Research Project funded by the European Commission, DG Justice, under the Fundamental Rights and Citizens programme.

Hargreaves, I., Guibault, L., Handke, C., Valcke, P., & Martens, B. (2014). Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining: report from the expert group. Report commissioned by the European Commission, Directorate-General for Research and Innovation.

Hildebrandt, M. (2008). Defining profiling: a new type of knowledge? In M. Hildebrandt & S. Gutwirth (Eds.), Profiling and the Identity of the European Citizen (pp. 39-50): Springer.

Holbrook, T.R. (2007). Extraterritoriality in US Patent Law, (4) William & Mary Law Review 6, 2119-2192.

Janssens, M.-C., "Bescherming van computerprogramma's: oude wijn in nieuwe vaten?", *DAOR* 2011, 98, 205-221

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, 110(15), 5802-5805. doi:10.1073/pnas.1218772110

Kosinski, M., Matz, S., Gosling, S., Popov, V. & Stillwell, D. (2015) Facebook as a Social Science Research Tool: Opportunities, Challenges, Ethical Considerations and Practical Guidelines. American Psychologist. Dataset available at: mypersonality.org

Van Der Noll, R., Van Gompel, S., Guibault, L., Weda, J., Poort, J., Akker, I., & Breemen, K. (2012). Flexible copyright: the law and economics of introducing an open norm in the Netherlands. Report in the SEO Economic Research-Series, Amsterdam, 2012.

Roosendaal, A. (2013). Digital Personae and Profiles in Law. Protecting Individuals' Rights in Online Contexts [Ph.D. thesis, Tilburg University]. Oisterwijk: Wolf Legal Publishers.

Triaille, J.-P., de Meeûs d'Argenteuil, J., & de Francquen, A. (2014). Study on the legal framework of text and data mining (TDM). Brussels: European Union.

Van Dijk, N. (2009). Intellectual Rights as Obstacles for the Transparency of Profiling Processes. In A. Deuker (Ed.), From Mobile Marketing in the Perspective of Identity, Privacy and Transparency, FIDIS deliverable 11.12 (pp. 57-67).

van Dijk, N. (2010a). Auteursrecht in profielen. Computerrecht, 35(2), 53 - 61.

Van Dijk, N. (2010b). Property, Privacy & Personhood in a World of Ambient Intelligence. Ethics and Information Technology, 12(1), 57 - 69.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036-1040. Online available at : http://www.pnas.org/content/112/4/1036.full

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

Case Law

Ashby Donald and others v. France, Appl. nr. 36769/08, ECtHR (5th section), Strasbourg 10 January 2013 ;

Fixtures Marketing Ltd v Oy Veikkaus Ab, C-46/02, ECLI:EU:C:2004:694 ;

The British Horseracing Board Ltd and Others v William Hill Organization Ltd, C-203/02, ECLI:EU:C:2004:695.

Infopaq International A/S v Danske Dagblades Forening, C-5/08, ECLI:EU:C:2009:465, para. 37.

Eva-Maria Painer v Standard VerlagsGmbH and Others, C-145/10, ECLI:EU:C:2011:798

Football Dataco Ltd and Others v Yahoo! UK Ltd and Others, C-604/10, ECLI:EU:C:2012:115.

Deckmyn v. Vandersteen, C-201/13, EU:C:2014:2132.

Annex A: Excerpt from "Facebook Statement of Rights and Responsibilities"

From the *Facebook Statement of Rights and Responsibilities*[65] (version of November 15, 2013):

Art. 9. Special Provisions Applicable to Developers/Operators of Applications and Websites

If you are a developer or operator of a Platform application or website, the following additional terms apply to you:

1. You are responsible for your application and its content and all uses you make of Platform. This includes ensuring your application or use of Platform meets our Facebook Platform Policies and our Advertising Guidelines.
2. Your access to and use of data you receive from Facebook, will be limited as follows:
    1. You will only request data you need to operate your application.
    2. You will have a privacy policy that tells users what user data you are going to use and how you will use, display, share, or transfer that data and you will include your privacy policy URL in the Developer Application.
    3. You will not use, display, share, or transfer a user's data in a manner inconsistent with your privacy policy.
    4. You will delete all data you receive from us concerning a user if the user asks you to do so, and will provide a mechanism for users to make such a request.
    5. You will not include data you receive from us concerning a user in any advertising creative.
    6. You will not directly or indirectly transfer any data you receive from us to (or use such data in connection with) any ad network, ad exchange, data broker, or other advertising related toolset, even if a user consents to that transfer or use.
    7. You will not sell user data. If you are acquired by or merge with a third party, you can continue to use user data within your application, but you cannot transfer user data outside of your application.
    8. We can require you to delete user data if you use it in a way that we determine is inconsistent with users' expectations.
    9. We can limit your access to data.
    10. You will comply with all other restrictions contained in our Facebook Platform Policies.

[65] Online available at: <https://www.facebook.com/legal/terms%20>

3. You will not give us information that you independently collect from a user or a user's content without that user's consent.
4. You will make it easy for users to remove or disconnect from your application.
5. You will make it easy for users to contact you. We can also share your email address with users and others claiming that you have infringed or otherwise violated their rights.
6. You will provide customer support for your application.
7. You will not show third party ads or web search boxes on www.facebook.com.
8. We give you all rights necessary to use the code, APIs, data, and tools you receive from us.
9. You will not sell, transfer, or sublicense our code, APIs, or tools to anyone.
10. You will not misrepresent your relationship with Facebook to others.
11. You may use the logos we make available to developers or issue a press release or other public statement so long as you follow our Facebook Platform Policies.
12. We can issue a press release describing our relationship with you.
13. You will comply with all applicable laws. In particular you will (if applicable):
    1. have a policy for removing infringing content and terminating repeat infringers that complies with the Digital Millennium Copyright Act.
    2. comply with the Video Privacy Protection Act (VPPA), and obtain any opt-in consent necessary from users so that user data subject to the VPPA may be shared on Facebook. You represent that any disclosure to us will not be incidental to the ordinary course of your business.
14. We do not guarantee that Platform will always be free.
15. You give us all rights necessary to enable your application to work with Facebook, including the right to incorporate content and information you provide to us into streams, timelines, and user action stories.
16. You give us the right to link to or frame your application, and place content, including ads, around your application.
17. We can analyze your application, content, and data for any purpose, including commercial (such as for targeting the delivery of advertisements and indexing content for search).
18. To ensure your application is safe for users, we can audit it.
19. We can create applications that offer similar features and services to, or otherwise compete with, your application.

Art. 10. About Advertisements and Other Commercial Content Served or Enhanced by Facebook

Our goal is to deliver advertising and other commercial or sponsored content that is valuable to our users and advertisers. In order to help us do that, you agree to the following:

1. You give us permission to use your name, profile picture, content, and information in connection with commercial, sponsored, or related content (such as a brand you like) served or enhanced by us. This means, for example, that you permit a business or other entity to pay us to display your name and/or profile picture with your content or information, without any compensation to you. If you have selected a specific audience for your content or information, we will respect your choice when we use it.
2. We do not give your content or information to advertisers without your consent.
3. You understand that we may not always identify paid services and communications as such.

Annex B: table summarizing the conclusions of chapter 2

Rights which can be relevant for DataBait	Questions or tensions with regard to the functioning of DataBait	Conclusions/answers/course of action with regard to these questions or tensions
Trade secrets	Can DataBait's trained and/or untrained profiling algorithms infringe on a trade secret of an OSN?	*No.* DataBait's trained and/or untrained profiling algorithms are independently developed within the USEMP consortium and are not obtained in an illegal way (e.g. by hacking into protected information or by manipulating employees or service providers to gain access to such information).
	Can DataBait's extraction of data from an OSN infringe on a trade secret of the OSN?	*Unlikely.* Whether Facebook or another OSN or browser could claim infringement of trade secret, depends on the steps these actors undertake to keep these data secret and/or if they have any effective contractual clauses (in case the extraction is enabled by the OSN/browser). *In the case of Facebook this is not very likely, because the extraction of data is enabled through Facebook's API. However, one could argue that some of the clauses in Facebook's policy for app developers (https://developers.facebook.com/policy/#data; clauses 3.6-13) are nondisclosure clauses which indicate that the data should be treated as trade secrets.* In case an OSN or browser would claim the infringement of trade secrets, the legitimate exercise of the right to freedom of information and expression could be invoked in defence (art. 4(2) proposed Trade Secret Directive, art. 4(a) amended proposal).
Patents	Can trained and/or untrained profiling algorithms be patented?	*Maybe.* Software and mathematical models 'as such' cannot be patented. However, if software or a mathematical method solves a technical (and not a purely administrative) problem as a "computer implemented invention", it may indeed be patentable (if it meets the other conditions as well).
	Could DataBait could be patented?	*Probably not.* One of the requirements for a patent is the novelty of the invention. Given the fact that DataBait is already used online, it is unlikely that the USEMP consortium (or somebody else) could patent DataBait.
	Can an OSN, if it holds any relevant patents rights, use these to oppose the development, offer and use of transparency tools	*The chances that an OSN could effectively oppose DataBait based on patent protection are limited.* Many European national patent legislations contain a research exception, which entails that the patent holder cannot prevent the use of the invention when the use is for scientific purposes. However, if a profile transparency was to be exploited

	(such as DataBait)?	by a commercial party, the OSN could probably use its relevant patent rights to oppose the functioning of the transparency tool.
Copyrights	Can copyrights in content created by an OSN user be opposed to a profiler (an OSN or a profile transparency tool provider like DataBait)?	*If there is no applicable exception or user consent: yes.* Profiling requires that data are copied ('reproduced'). In as far as these data are copyright protected content (pictures, videos or text with some element –however minimal - of 'authorship' and thus of 'originality'), making copies is a protected act under copyright law. Consequently, any profiler who wants to profile based on such content needs either to demonstrate that her practice falls within one of the exceptions provided in Directive 2001/29 (for exceptions for temporary acts of reproduction and for scientific reproduction as implemented in the applicable national law) or obtain the right holders' prior consent (a license). If no exception applies and the profiler has not obtained a license there will be a copyright infringement. USEMP is unlikely to infringe on the copyright on user generated content, considering firstly the applicable exceptions for temporary acts of reproduction and use for scientific research and secondly the licence granted by DataBait users in the DLA.
	Can an OSN provider who holds copyright on elements constituting the OSN (e.g., the graphic user interfaces, computer programs, databases and user generated content which has been licensed to the OSN) rely on these exclusive rights to prohibit transparency efforts?	*It is unlikely that the USEMP consortium infringes any rights to the computer programs developed by the OSNs.* The USEMP consortium has developed its own computer programs in an independent way. It has not had access to the OSN computer programs and has not attempted to reverse engineer their computer programs, hence no infringements of copyright on OSN software are to be expected. Based on our consultation with the technical partners in the USEMP project, it is unlikely that any parts of an OSN's graphic user interfaces will be reproduced. However, given that the fact that the final DataBait visualizations are still under development, we will continue to closely monitor that no elements of the graphic user interface of OSNs are reproduced. Considering that GUIs are not protected under the Computer Programs Directive but as other copyright works (cf. CJEU's decision in *BSA*), it should be verified (at a later stage) whether any exception provided in the InfoSoc Directive can apply. At this stage of the USEMP project, it is likely that the exception for scientific purpose can apply.
Copyright and sui generis right in databases	Do profile transparency tools like the ones created by USEMP reproduce (protected parts) of the copyright protected database (if any) of the	*Maybe.* The data which the consortium gets through the Facebook app, in contrast to those collected through the browser plug-in, are in some way structured by Facebook (the OSN) and could thus be protected by both the copyright in the database structure or sui generis right of the OSN. It can be argued that the structure of the OSNs database is not original and therefore not protected under copyright. The selection of the contents is arguably done by the

	OSN?	OSN users. The arrangement may be done by the OSN, but it should be verified whether the arrangement is taken over in DataBait. In any case, it should be verified whether any exception for scientific research may apply.
	Is it possible that the DataBait profiling process infringes on any OSN database rights (substantial-investment)?	*It depends.* In the first place it should be verified whether a substantial part of the contents of the database are extracted and/or re-utilised. The extraction and/or re-utilisation of non-substantial parts is not restricted. In as far as the data one gets through the API Facebook are based on the explicit permissions to access certain data (and the structure in which they are offered), the USEMP consortium cannot be said to infringe on either the copyright in the database structure or sui generis right of the OSN. However, if USEMP would not get permission to obtain data through the Facebook API and a work-around was to be used, the OSN might invoke exclusive database rights. The sui generis right provides an optional exception with regard to scientific research for the extraction or re-utilization of substantial parts of a database (sui generis right) for the sole purpose of scientific research[66] (art. 9(b) of the Database Directive, provided that this exception is implemented in the applicable national law). This exception may not apply to other tools developed outside a scientific context are more likely to infringe database rights. The merit of any hypothetical infringement claims will be examined in more detail in the remainder of the project, should the OSNs not consent to USEMP's use of the data.
Trademarks	Can the display of an OSN logo within DataBait infringe on the trademark of that OSN?	*The likelihood that DataBait infringes on the trademark of an OSN is low.* An infringement of the Facebook trademarks is unlikely in the case of how DataBait refers to OSNs like Facebook. No goods or services are offered under a sign even remotely similar to the protected trademarks. Instead, the trademark is used to explain DataBait's functioning and its impact on its users and to explain the use and functioning of the DataBait tools.

[66] See for a more nuanced and detailed discussion: Traille et al., 2014.