D3.11

Profile transparency, Trade Secrets and Intellectual Property Rights in OSNs – v3

v 1.3 / 2016-10-25

Sari Depreeuw and Katja de Vries (ICIS-RU)

This document analyses whether the end users' right to profile transparency and the way in which DataBait supports this (see D3.10) 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 design implications for the DataBait tools. It also points to some issues that would benefit to be debated and studied in more detail by policy makers and legal researchers.

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	Eleftherios Spyromitros-Xioufis(CERTH) Rob Heyman (iMinds)
Version	1.3
Status	Final
Dissemination level	**PU: Public**
Due date	2015-09-30
Delivery date	2016-09-30

Version	**Changes**
1.0	Initial Release (Depreeuw, ICIS)
1.1	Adjusted version (De Vries, ICIS)
1.2	Review (Spyromitros-Xioufis, CERTH)
1.3	Review (Rob Heyman, iMinds)

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 generi*s) 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. With regard to the way the research in this deliverable could be integrated in the DataBait tool, we conclude that informing DataBait users about the possible tensions between third party transparency tools (such as DataBait) and OSNs is not necessary. However, on the DataBait developer website (mainly aimed at transparency tools developers but also at academics, policy makers and others interested in questions with regard to profile transparency) which will be launched shortly after the end of the USEMP project, it would be useful to give access to this report as it can provide guidance to anyone who wants to create an independent transparency tool.

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 *software 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. By contrast, the USEMP partners have used *data sets*, possibly protected under IPRs, compiled by OSNs or by other third parties. It will be verified to which extent the use of such data is protected, whether any exceptions apply or whether a licence is required.

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

1.1. Introduction

This deliverable[1] looks at how the intellectual property rights (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. EU data protection law (*Directive 95/46/EC* [DPD 95/46], which applies until 25 May 2018, and its successor, the recently adopted *General Data Protection Regulation 2016/679* [GDPR 2016/679]) 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:

Possible conflicts between profile transparency and profile protection through trade secrets and IPRs – some key points:

- The right to profile transparency: every data subject has the right to access her data and to know the logic of any profiling to which she is subjected.
- Profile transparency should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property of data controllers or third parties.
- However, when striking a balance between profile transparency and opposing rights, the result cannot be a refusal to provide *all* information to the data subject.
- The right to respect for private life and data protection carry a heavy weight and are not easily overruled by commercial interests.

> "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

[1] This deliverable builds on the research performed in task T3.7. (Relevant EU Intellectual property rights regarding databases and software employed by the OSN).

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*[2] (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 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 63 of the *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:

> "A data subject should have the right of access to personal data which have been collected concerning him or her, and to exercise that right easily and at reasonable intervals, in order to be aware of, and verify, the lawfulness of the processing. (…) Every data subject should therefore have the right to know and obtain communication in particular with regard to the purposes for which the personal data are processed, where possible the period for which the personal data are processed, the recipients of the personal data, the **logic involved in any automatic personal data processing** and, at least when based on profiling, the consequences of such processing. Where possible, the controller should be able to provide remote access to a secure system which would provide the data subject with direct access to his or her personal data. **That right should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software**. **However, the result of those considerations should not be a refusal to provide all information to the data subject**. Where the controller processes a large quantity of information concerning the data subject, the controller should be able to request that, before the information is delivered, the data subject

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

> specify the information or processing activities to which the request relates." (*bold ours*)

This leads to a situation where various profilers and the data subject have co-existing legal claims in the same profile or profiling process (see figure 1).

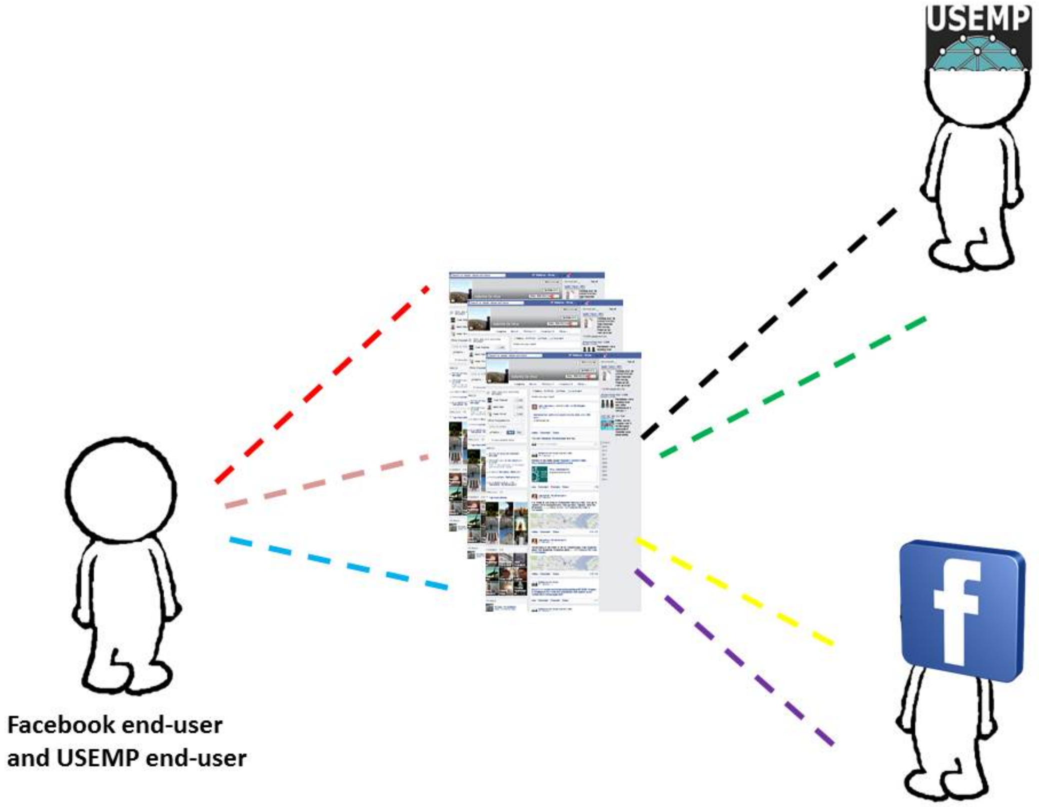


Figure 1: The "same" profile can be the object of various legal relations with multiple actors.

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[3]), up until now there is no case law where IP rights in profiling and data protection law are confronted with each other[4]. 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

[3] *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.

[4] 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.

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 D3.10, is defined in GDPR 2016/679 as:

> … any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements (Art. 4.1(4) GDPR 2016/679)

Which types of ‘profiles’ could potentially be subject to IPRs and trade secret protection?

[A] ‘Profiles’ relating to a particular individual:

- A set of volunteered (and/or behavioural) data
- A set of data inferred by using a trained machine learning algorithm (that is, ‘output data’)
- A set of data combining volunteered, behavioural and/or inferred data

[B] ‘Profiles’ *not* relating to a particular individual

- An untrained machine learning algorithm (i.e. a ‘recipe’ for building a predictive data model)
- A data set (relating to several individuals) used to train the untrained machine learning algorithm
- A trained machine learning algorithm, that is, an inferred predictive data model capturing the structure or underlying regularities in a data set

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”[5] is. However, there are two meanings that stand out:

> **an individual set of characteristics**, e.g., a Facebook profile consisting of volunteered data on the frontend, but including observed (behavioural) data at the backend.
>
> an **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[6], which implies that the holder of the IPR can exercise exclusive

[5] In the grey textbox on the right we list six types of "profiles". Under the fourth bullet point we qualify an "untrained machine learning algorithm" as a type of profile. Here we stretch the meaning of the word "profile". It would be better to say that an untrained algorithm is *a tool* that can be used *to create a user profile* (if trained with appropriate data). It is a general tool for learning from data and in this sense calling it a "profile" is non uncontroversial. However, for the sake of clarity (in opposition to the "trained algorithm") it seemed helpful to include the untrained algorithm in this list too.

[6] 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.

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 'recipes' how to capture data in a particular model), the trained algorithms (which one could also call the 'models' or *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 and trade secret protection.

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?

DataBait is an independent provider of transparency: a third actor with regard to the relationship between the 'OSN-user-as-data-subject' and the 'OSN-as-data-controller'

After all, DataBait is *an independent provider of transparency: 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.

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 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.

The DataBait architecture refrains from infringing on trade secrets and IPRs of OSNs

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 software code 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).

1.2. DataBait's profiling tool – independent creation

DataBait is a profiling tool for the user to get an idea which information the OSN has and can derive about her.

DataBait mimics the function of the profiling processes of the OSNs but is in fact an independent creation of the USEMP consortium. Its teams have invested their creative effort in the development of algorithms and software code to come as closely as possible to an accurate description of the profiled person, i.e. DataBait will derive certain information about the user on the basis of the data the user has put in the OSN. The purpose is not so much to approximate what exactly the OSN knows about the user.

Five 'objects' in the profiling process which can be relevant from the perspective of IPRs:

- the set of training and testing data,
- the untrained algorithm which still has to be 'trained',
- the output space,
- the resulting 'trained algorithm',
- the data analyzed by the trained algorithm.

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 by the OSN or not. DataBait shows what might be 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.

> ***DataBait shows what 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 by the OSN or not. DataBait shows what is possible.***

The information provided to the user by DataBait concerning what can be inferred from her digital trail, particularly the so-called 'disclosure score', is in some sense 'speculative'. The USEMP consortium does not have access to the technology used by OSNs, but 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.

It is assumed that the profiling techniques of USEMP and the main OSN are similar based on the following considerations. 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.

How does it work?

At least **five 'objects'** in a profiling process can be relevant from the perspective of IPRs: (i) the set of training and testing data, (ii) the algorithm which is 'trained', (iii) the output space, (iv) the resulting 'trained algorithm' (or: 'predictive data model' or 'classifier'[9]), and (v) the data analysed by the trained algorithm.

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 ***output space***: it is the set of possible outputs of the learner and thusdefines which outputs need to be considered. Now, let's assume that the OSN keeps its output 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, then 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

[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/
[9] 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.

inference is way more complex. How to explain to a computer what a cigarette looks like in a picture? And which words indicate that 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[10] data** come in. An **untrained machine learning algorithm** is a mathematical "recipe" for learning a function that maps inputs (e.g. pictures and posts) to outputs (e.g. smoker / non-smoker) based on labelled examples (i.e. input-output pairssuch as pictures and posts labelled by a human as representing a "smoker" or a "non-smoker"). The resulting "**trained algorithm"** or "classifier" is a predictive data model (that is, a "learned mapping function") which can then be used to label unlabeled examples. This is called **supervised learning[11]** (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 for classifying 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 without human supervision).

In short, 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 'creativity' that goes into each of the five named 'objects'?

- Making an output 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[12] but a particular problem (such as: distinguishing between smokers

[10]Training data are usually divided to a training set and a test set in order to tune the algorithm's parameters. Both the training and the test set are training data (i.e. labelled examples).

[11] The kind of machine learning algorithms used in the USEMP project are **supervised** machine learning algorithms, not unsupervised ones.

[12] Examples of such algorithms are *linear 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 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.

and non-smokers based on OSN pictures) will often require that such algorithms are tailored and/or combined with each other[13].

- 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).

The consortium does not have access to the internal data model or to the computer programs used by OSNs, and fully relies on its own data models. On top of that, the consortium only has access to very limited training data in comparison to large OSNs.

Both elements (own data model/output 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.

[13] 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 output 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.

1.3. IPR protected elements in the 'profiling' process

If the algorithms, the source code and the data sets used by DataBait are different from the OSN's, no identical copy of the profiling technology is made. Yet this finding does not suffice to rule out all infringement of IPRs.

Protected material posted on an OSN profile ordered according to its source:

- Copyright protected user generated content (e.g. images, photographs, texts, and videos that bear at least some trace of 'authorship') posted by users
- Material belonging to the OSN (e.g. the graphic interface in which content is presented)
- Material belonging to third parties (e.g. images, photographs, texts, and videos that bear at least some trace of 'authorship') created by others than the user who has posted it.

A user cannot license that which is not hers to give. Consequently, the DataBait *Data Licensing Agreement* (signed between the USEMP consortium and the DataBait user) does not cover third party material or material belonging to the OSN.

Two phases can be distinguished: the preparatory phase and the operational phase. The profiling tool consists of many components (algorithms, code, interfaces) that need to be developed and refined before being launched for public use. During this preparatory phase, certain acts are performed that can be protected under IPRs (e.g. transient copies of data set during the training of the algorithm). Similarly, when the profiling tool is launched, protected content may be used in a way that may be protected under IPR laws (e.g. photographs uploaded to Facebook are displayed in the DataBait environment).

The protected materials can have different sources:

Content posted by users. The content posted by the user can be protected under copyright (e.g. images, photographs, texts). These questions are analysed in D3.4, section 5 and Annex B, tables B.1 and B.2.

Material belonging to the OSN. The content extracted from the account the user has at the OSN may contain protected elements to which the OSN holds the rights (e.g. interfaces, content of protected databases). The OSN may also have assets protected under trade secrets.

Material belonging to third parties. DataBait uses data sets containing third party content (e.g. photographs or encyclopaedia posts) and users may post protected content (photographs, images, videos, texts), which is then processed by DataBait.

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 which can categorize new data (supervised learning) or discover interesting patterns in data (unsupervised learning). That means that neither output space, training and testing data, nor the algorithms, nor the trained algorithms are the same. The profiling processes of the OSN and DataBait consist of the same 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.

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.

> Protected subject matter from different sources is used while constructing and operating the DataBait profiling tool (content coming from the user, the OSN and third parties). The relevance of the source is that DataBait may need to clear licences from various parties.
>
> The protected material is used in different ways and in different phases of the profiling process (preparatory phase, operational phase). This may be relevant to assess whether potentially protected acts are covered by an exception in the applicable legislation.

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).

Two aspects of profiling matter with regard to IPRs: the profiling process (the profiling software may be protected under copyright or patent law, algorithms are typically protected as trade secrets) and the profiles (the profiles may be stacked in a database that may be protected under copyright or sui generis database right).

Two aspects of profiling to take into account with regard to potential IPR infringements:

- **The profiling 'tools':** What algorithms/software is used to profile with? Are these algorithms/software protected by IPRs? Are these algorithms/software created based on data or databases that are protected by IPRs?
- **The profiling 'objects':** Are the 'profiles' (that is, the data 'objects' analyzed by the profiling 'tools') protected by IPRs?

Both the OSN and DataBait hold certain rights to their profiles and profiling processes. In addition, DataBait must be aware of certain risks of infringement towards the OSN, where it reuses data coming from the OSN.

Here it will be examined which different types of rights that OSNs, browsers and third-party profilers might have in profiles (either the process or the protected subject matter used in the process). 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.

Legal rights which might be mobilized by an OSN to counteract third-party transparency tools (and thereby protect their economic, intellectual and creative efforts in user profiles):

- Trade secrets
- Patentable inventions
- Copyrighted 'expressions',
- IP protection of databases (through copyrights or sui geris rights)
- Trademarks

2.2. Trade secrets in profiling

Some of the biggest internet players keep distinctive features of their technology secret, while main principles of the functioning of the technology are well known. For example, it is understood how web search engines crawl and index the content published on the web or how the results to a query are drawn from the indexes, but how the search results are ranked is kept secret considering the competitive advantage this algorithm gives to one search engine operator over the other.

Such knowhow can be protected as a trade secret provided that the legal conditions are met. National laws differ currently very much in their definitions, in the type of legislation that affords protection and the scope of protection granted. The TRIPS Agreement obliges Member States to provide a minimum protection for undisclosed information, including trade secrets[14]. 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[15]. This will be the basis for the discussion.

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.

> ***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. […] One cannot claim protection for something that one has not tried to keep secret by taking "reasonable steps". Futile steps or mere pro forma measures are not sufficient.***

Protected subject-matter. The protection of trade secrets is provided at the national level but a European Directive has been adopted to harmonise the national protection rules. The Directive defines "trade secrets" as:

Information which meets all of the following requirements (Art. 39(2) TRIPS[16]; Art. 2(1) of the *Trade Secret Directive* - our emphasis):

[14] See Baker & McKenzie 2013.

[15] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, *OJ L* 157, 15.6.2016, p. 1–18.

[16] 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

(a) it 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) it has **commercial value** because it is secret;

(c) it has been subject to **reasonable steps** under the circumstances, by the person lawfully in control of the information, to **keep it secret**.

The legal definition of a trade secret in the EU Directive is very broad: a trade secret can be 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.

Four 'objects' in the profiling process which can be trade secrets:

- the set of training and testing data,
- the untrained algorithm which still has to be 'trained',
- the output space,
- the resulting 'trained algorithm'

In the context of profiling, the following elements could be trade secrets: a trained profiling algorithm (or 'predictive data model', which one can refer to as one type of "profile"), but also the "training set" as structured in a relational database on which an algorithm is trained (Ateniese, 2013), the output space (definition of the possible outputs, which is an essential part of the untrained algorithm) and the untrained machine learning algorithm (which is used to construct the trained algorithm).

While many people know what the approximate components of a profiling process are (an untrained or trained algorithm, an output space and a training set), the competitive advantage is exactly in the details (the data, their measurement and how the elements they interact). In this sense, the main ingredients of the Facebook news feed algorithm are well known, but the specifications can be trade secrets (provided that they remain secret and reasonable measures are taken to maintain the secret character).

Scope of protection. Such information is not protected under an intellectual property right but the Member States should provide for protection against the "unlawful acquisition, use or disclosure of their trade secret" (art. 4 Trade Secrets Directive)[17].

The **acquisition** of trade secrets is considered unlawful if it is carried out, without the consent of the trade secret holder by (a) unauthorised access to, appropriation of, or copying of any documents, objects, materials, substances or electronic files, lawfully

Convention), which makes it stand out. As such the TRIPS agreement is an extremely *comprehensive* legal IPR instrument.

[17] 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 secret information was used or misappropriated in an unlawful way. The specific conditions depend however on the legal instrument on which the trade secret owner relies, e.g. labour law against a (former) employee, contractual liability or unfair competition law against a competitor.

> under the control of the trade secret holder, containing the trade secret or from which the trade secret can be deduced or (b) any other conduct which, under the circumstances, is considered contrary to honest commercial practices. Preceding proposals of the Trade Secrets Directive mentioned examples such as theft, bribery, deception or breach of confidentiality obligations.
>
> 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 confidentiality agreement or any other duty to maintain secrecy of the trade secret; or (c) is in breach of a contractual or any other duty to limit the use of the trade secret.

The Directive thus marks two scenarios where the trade secret is brought outside the protected sphere either by a person who has unlawful access to the secret or by a person who has legal access but who disregards the secret character of the information she is supposed to respect.

The protection of the Directive has a potentially far-reaching scope. This is somewhat mitigated by the context provided in the considerations of the preamble, where it is provided that Directive should not create any exclusive right to know-how or information protected as trade secrets and that consequently, “the independent discovery of the same know-how or information should remain possible. Reverse engineering of a lawfully acquired product should be considered as a lawful means of acquiring information, except when otherwise contractually agreed” (cons. 16 preamble Trade Secrets Directive)[18].

> Where a profiler thus develops its own algorithms, on the basis of third party data sets, without having access to the profiling technology of existing OSNs, it makes no unlawful use of trade secrets belonging to OSNs. The same is true where it makes no attempt to reverse engineer the algorithms of other profiles (OSNs or other transparency tools) or to reconstruct their profiling tools.

Where a profiler develops its own algorithms, on the basis of third party data sets, without having access to the profiling technology of existing OSNs, it makes no unlawful use of trade secrets belonging to OSNs.

Beyond this first circle, the Directive extends the protection to a second circle, where a person acquires, uses or discloses information she has gotten from the person who initially violated the secret character of the information. The Directive provides that the acquisition, use or disclosure of a trade secret is thus considered unlawful when the person, at the time of acquisition, use or disclosure, knew or should, under the circumstances, have known that the trade secret had been obtained from another person who was using or disclosing the trade secret unlawfully (art. 4(4) Trade Secrets Directive).

[18] The European legislator also expects conflicts between the possibility to reverse engineer certain products and unfair competition laws. This issue is not addressed in the Trade Secrets Directive but may be tackled by the Commission at a later stage (cons. 17 preamble Trade Secrets Directive).

The Directive thus covers the acquisition and the use of information. It makes this protection more tangible by extending its coverage to the products that result from using the secret information without the holder's consent. Consequently, "unlawful use of a trade secret" is also the production, offering or placing on the market of infringing goods, or the importation, export or storage of infringing goods for those purposes, provided that the person carrying out such activities knew, or ought, under the circumstances, to have known that the trade secret was used unlawfully.

> ***The Trade Secrets Directive also extends its coverage to the products that result from using the secret information without the holder's consent.***

The Directive provides certain exceptions, where the use of trade secrets is not considered unlawful (despite the circumstance that the holder did not her prior consent). Where the acquisition, use or disclosure is done for (i) exercising the fundamental right of freedom of expression and information, (ii) revealing misconduct, wrongdoing or illegal activity in the general public interest, (iii) disclosure by workers to their representatives as part of the legitimate exercise by those representatives of their functions or (iv) for the purpose of protecting a legitimate interest recognised by Union or national law (art. 5 Trade Secrets Directive).

The case of DataBait. If we apply these rules to (i) the profiles (i.e., the 'objects' of the profiling) and (ii) the act of profiling (i.e., the profiling 'tools'), we come to the following conclusions (on the basis of the Trade Secrets Directive, without an examination of the national provisions that may apply).

> **Individual OSN profiles** as such are not considered trade secrets. The OSN user publishes her profile on the OSN and makes all information accessible to other OSN users (at least her connections or part of her connections). Moreover, taken on their own[19], these profiles are not likely to represent a commercial advantage because of their secret nature.

[19] So what about extracting *large amounts* of data (or profiles) from a browser or an online social network site? In contrast to individual profiles, such large amounts of data can be used to train a competitive algorithm, and, consequently, are likely to have commercial value, along with the precise training method and the analysis of the results. However, the fact that these data, taken together, have a commercial value is not enough to qualify them as trade secrets – they can only be that if they are actually kept secret. Thus, the crucial question is how information is « extracted » from a browser or an OSN. An OSN relying on the protection of its trade secrets should demonstrate that (i) the information is not generally known (which is, for example, not the case if it is publicly accessible information which is 'scraped') and the data are not "readily accessible to persons within the circles that normally deal with the kind of information" (cf. definition in the Trade Secrets 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. If the large amounts of data are extracted through a freely available API, which does not contractually stipulate nondisclosure, then the conclusion is no trade secret. In order to infringe on a trade secret the extraction either has to be is in some way illegal (e.g. hacking into a database) or use the data in ways not permitted by the OSN policy.

Aggregate OSN profiles and meta-data. OSNs process individual user profiles (e.g. by adding machine-readable behavioural data) and combine volunteered, behavioural and inferred data contained in the individual profile, resulting in valuable know-how that could qualify for protection as a trade secret. This is notably the case for Facebook, 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.

Is DataBait infringing on trade secrets? No.

- **Individual OSN profiles are no trade secrets.** They are too accessible to be considered secret and individual profiles provide no commercial advantage.
- **Aggregate OSN profiles and meta-data that can be accessed through a freely available API are no trade secrets,** at least as long as no nondisclosure clause is signed between OSN and an app developer.

DataBait does not have access to the predictive algorithms or inferred data that are trade secrets; nor does it reverse engineer or reconstruct them.

Facebook offers access to user data through a documented API that it controls[20]. These data have consciously (posts, uploads, profile information, etc.) or less consciously (date and place of Facebook login, device information, etc.) been generated by Facebook users. To the extent that the API and the access to these data is not subject to a confidentiality obligation, this information is not considered a trade secret.

Other information (predictive algorithms and inferred data). Next to the data that are accessible through the API, Facebook also possesses data which are fully kept a secret. For example, the Facebook API data do not include inferred (i.e., 'derived')

[20] If an API is freely available there are probably no "reasonable" steps to keep it secret. If an API is only made available after having accepted a secrecy obligation, this could be seen as 'reasonable' steps to keep the data secret. In the case of Facebook, use of the API has to be approved by Facebook. We didn't see any explicit contractual nondisclosure clause in Facebook's platform policy (https://developers.facebook.com/policy/ , last accessed 20 September 2016). However, it should be noted that there are some clauses that put restraints on what app developers can do with the information they receive through the API. For example, in Art. 3(1) it says: "*Protect the information you receive from us against unauthorized access, use, or disclosure*." Art. 3(7) says: "*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*". Art. 3(9) says: "*Don't sell, license, or purchase any data obtained from us or our services*". Art. 3(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*". Interpreting the cited clauses as contractual nondisclosure clauses seems a bit of a stretch to us. The limitations put to the proliferation of Facebook data by these clauses seem to aim to protect the privacy, security and confidentiality of user data (Arts. 3(1) and (7)) and to prevent others from making money by re-selling Facebook data (Arts. 3(9) and (10)). Thus, while the Facebook data provided through the Facebook API clearly have commercial value, it seems to us that it is a stretch to say that their commercial value is created because it is kept secret. However, whether such clauses should be interpreted as non-disclosure clauses something which can be debated. Moreover, if Facebook would add an explicit contractual nondisclosure clause to keep the data secret, these data would qualify as a trade secret. Therefore, we add a bit of nuance to our position that the data accessed through the freely available Facebook API are not trade secrets: after all, they could be protected trade secrets if Facebook would add a contractual confidentiality clause to its platform policy, or if the aforementioned clauses would be interpreted as such.

data: these data are used in Facebook's targeted advertising service. An OSN like Facebook might also possess algorithms that are kept fully secret, such as the algorithms applied to the data to derive more data. Considering the value of this information (the OSN's competitive advantage depends to a large extent on this information and its secret nature), such predictive algorithms and inferred data should be considered trade secrets.

The USEMP consortium does not violate any trade secrets protection: it does not have direct access to the OSNs' algorithms and it makes no attempt to reverse engineer or reconstruct the same algorithms on the basis of observations (which would not even be considered unlawful as such under the Trade Secret Directive). The inferred knowledge presented in the USEMP tools is based on untrained and trained machine learning algorithms developed within the USEMP consortium. 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"). The "proprietary" algorithms or data sets (which are not available through the APIs) belonging to the OSNs are not used in DataBait.

2.3. Profiles as patentable inventions

Any profiler should take into consideration that parts of a profiling process can be protected under patent law. On the one hand this means that it should tread carefully and – ideally – examine the relevant patents in its field (OSNs such as Facebook and Twitter may have protected certain processes). On the other hand, a profiler may want to try and get patent protection for its inventions in its own profiling process.

Two ‘objects’ in the profiling process which could sometimes be patentable inventions:

- the untrained algorithm which still has to be ‘trained’,
- the resulting ‘trained algorithm’

Subject-matter. To answer these pertinent questions we first have to look at the protected subject matter of a patent. A definition can be found in the European Patent Convention[21].

> 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.
>
> (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.

Four conditions should be met 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).

[21] The member states of the European Union are signatories of the European Patent Convention, which harmonises the conditions for protection and the assessment but does not result in a unitary title.

Invention. Whether algorithms (both ‘untrained’ algorithms, including the part of the algorithm defining a particular output 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[22].

Profiling algorithms or profiling data models can be seen as “mathematical methods” or “programs for computers” and consequently not considered an “invention” under the European Patent Convention.

However, it is observed that the European Patent Office does grant patents covering computer programs: 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:

Are profiling algorithms patentable? The object of the patent has to be:

(1) an **invention**, which is

(2) a **novel** and

(3) **inventive product, apparatus or a process** that

(4) has **industrial applicability**.

In the context of profiling algorithms it is particularly relevant that the following *cannot* be patented:

(1) **Computer programs (software) or mathematical models ‘as such’**; however if software or models are ‘functional’ to a technical solution to a technical problem they can be patentable as “**computer implemented inventions**”. This ‘loophole’ makes it sometimes possible for profiling algorithms to be patentable.

(2) **Methods that are described in scientific publications**.

> “…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.

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

[22] 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, while it is not patented in Europe. *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.

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 of software. The implementation in hardware ("system") is not decisive[25].

In recent case law[26] the European Patent Office (EPO) has adopted a flexible approach to the patent protection of software[27], if it has a technical character and can thus be qualified as 'computer implemented inventions'. A computer program is consequently 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].

> ***"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".***

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 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.

[23] https://www.epo.org/news-issues/issues/software.html

[24] Four examples of algorithm-related inventions patented by Facebook in this way :
(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] European Patent Office [EPO], Decision T208/84; OJ EPO 1/1987, 14, accessible via https://www.epo.org/law-practice/case-law-appeals/recent/t840208ep1.html (accessed 1 Nov 2015). In the *Vicom* case, 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".

[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*. The EPO uses a restrictive interpretation of the term 'as such'.

[28] http://www.epo.org/news-issues/issues/software.html.

[29] http://www.epo.org/news-issues/issues/software.html.

Large OSNs, like Facebook, have several patents and patent applications in Europe on various aspects of the complex functioning of the OSN. A patent may indeed be interesting to an OSN since it allows protecting a function or an outcome (not so much the code that executes the function) and can thus be used by the holder of a strategic patent to ward off its competitors (within the limits of competition law). Mostly a patent application will only concern a part of the technology used that can be presented as a computer implemented invention, while avoiding to reveal more information than required for acquiring the patent.

A patent may be interesting to an OSN since it allows protecting a function or an outcome (not so much the code that executes the function) and can thus be used by the holder of a strategic patent to ward off its competitors, while avoiding to reveal more information than required for acquiring the patent.

It is then not excluded that untrained algorithms (which can be described as a software expressing a mathematical 'model' for creating a trained algorithm) or trained algorithms (which can be described as software expressing a 'predictive data model') be patented within the EU.

Conditions for protection. Even if a computer program or an algorithm has a technical nature and can be considered an invention, in order to qualify for patent protection, it 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

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.

[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.

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.

We are not aware of any patented inventions that are reused in the DataBait tools. However, in practice, considering the important volume of patents applications and granted patents, it is (especially for a research project with limited means) almost impossible to check every patent application that could possibly overlap with DataBait.

Scope of protection. The scope of protection of a patent under the EPC is determined by the national patent laws. Considering that the protection could cover a product, an apparatus or a process (depending on the claims – art. 69 EPC), a profiler using a patented technology may infringe the exclusive rights protecting that invention (depending on the national law – art. 64 EPC).

While the rights protecting the patented invention may be broadly formulated ("production", "offer", "distribution", "use", "application")[31], many national laws also provide exceptions[32] covering non-commercial use in the private sphere or, more importantly for DataBait, acts regarding the patented invention for scientific purposes[33].

In the case of DataBait, the members of the USEMP consortium are not aware of any patented inventions that they would reuse in the DataBait tools. Furthermore, even if the DataBait tools were to use (parts of) a protected invention, such use could arguably not constitute an infringement to the extent that the DataBait tools serve a scientific purpose and most national legislations contain an exception for that purpose. Currently the USEMP consortium is looking into possibilities how the DataBait tool can be kept available online in a non-profit form after the USEMP project has ended, for example if it would be curated by a non-profit organisation,

Even if the DataBait tools were to use (parts of) a patent protected invention, such use could arguably not constitute an infringement to the extent that the DataBait tools serve a scientific purpose.

[31] E.g. art. XI.29 of the Belgian Code of Economic Law.

[32] « Since the socio-economic conditions and priorities of a country influence this balancing of interests, provisions in patent laws on exceptions and limitations vary from one country to another. Nevertheless, the SCP (*Standing Committee on the Law of Patents*) has identified that the legislation of many countries provides some or all of the following exceptions and limitations to patent rights: (i) private and/or non commercial use; (ii) experimental use and/or scientific research; (iii) extemporaneous preparation of medicines; (iv) prior use; (v) use of articles on foreign vessels, aircrafts and land vehicles; (vi) acts for obtaining regulatory approval from authorities; (vii) exhaustion of patent rights; (viii) compulsory licensing and/or government use; and (ix) certain use of patented inventions by farmers and breeders » http://www.wipo.int/patents/en/topics/exceptions_limitations.html

[33] See e.g. art. XI.34 of the Belgian Code of Economic Law.

research institute or civil rights organisation. Depending on the format this might mean that the purpose would no longer be scientific. However, given that most national jurisdictions have some exception for private and/or non commercial use[34] any use of a patent protected invention by DataBait (which in itself, as stated above, is not very likely) would not constitute an infringement.

The speculative nature of this statement immediately reveals one of the difficulties in current patent practice. Considering the obligation to publish the patents, developers are in theory supposed to be aware of the protected inventions being part of the state of the art. Considering the important volume of patents applications and granted patents and the limited means of research institutions to verify whether their inventions use patent protected, inventions, these measures are mostly theoretical.

[34] See above, footnote 28.

2.4. Profiling and copyright

The profiling process of DataBait may contain several acts with copyright relevance. The computer programs underlying DataBait and its user interfaces can be protected under copyright (held by the USEMP partners).

Inversely, the USEMP teams have used third party creations to come to the DataBait solution, in particular third party software and “data sets” containing protected images or text. These third right holders can be software developers, authors of graphic or literary works or derived right holders (such as OSNs). In this section it will only be examined whether the OSN or other parties can exercise their IP rights to prevent DataBait from functioning.

Whether the OSN user can invoke her IP rights against the OSN or profiler, as a complementary right to the data protection rights in order to resist the profiling, is examined elsewhere (in D3.12).

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

When examining the copyright constraints to the development and use of DataBait, two phases can be distinguished: (i) the preparatory phase where DataBait is being built and (ii) the exploitation phase where DataBait is up for use by its members.

Preparation. DataBait thus consists of software components (see D3.13) that each take up a function of this complex system, e.g. by analysing texts, links and image. However, it is not sufficient to develop the software for this system to function accurately: the computer code should also be “trained” to find the most relevant information. This training is done on the basis of sets of data[35] (images or text), which may contain works protected under copyright or which may be contained in protected databases (see next section, 2.5). An important question is consequently whether the content of these data sets is protected, whether the acts performed on the data sets are covered under copyright and whether a licence is required or an exception applies.

Copyright infringements can occur during:

- The preparatory phase (training the profiling algorithm)
- The exploitation phase (trained profiling algorithm is applied for classification purposes)

Exploitation. The DataBait system consists of several software components that fetch data from several sources to identify the data that the OSN user has submitted to the OSN in order to reconstruct which image the OSN has put together of its user, in particular by simulating which information can be inferred about the OSN user on the basis of the data she has actually shared with her friends via the OSN platform. The OSN operator 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), and could rely on these exclusive rights to prohibit transparency efforts.

[35] See D3.13 and D2.3 (p. 8-16)

Copyright. In order to address these 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).

Copyright is still a national matter, meaning that each country defines which rules apply under copyright as far as the protected subject matter, authorship and ownership, scope of protection (exclusive rights of economic and moral nature, exceptions) are concerned. However, important harmonisation efforts have been made at the international and European levels, in particular in the Berne Convention (BC), the 1996 WIPO Copyright Treaty (WCT) and, at EU level, the Directives in the field of copyright.

The **subject matter** protected under copyright is not explicitly defined but indications can be found in various legal instruments.

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

Five ‘objects’ in the profiling process which can be relevant from the perspective of copyrights:

- the set of training and testing data (either because they contain copyright protected content or because the database is copyright protected)
- source code or object code (the ‘expression’) of an untrained algorithm which still has to be ‘trained’ (i.e. this does not pertain to the ‘algorithm itself’ and copyright cannot prevent others from creating other expressions of the ‘same algorithm’)
- the output space (i.e. the set of possible ‘outputs’ from which an algorithm can ‘choose’) as such cannot be protected by copyright but as part of untrained (predefined set of outputs) or trained (outputs are defined through the learning process) algorithm (that is: its particular expression) it could.
- source code or object code of (the ‘expression’) the resulting ‘trained algorithm’ (i.e. this does not pertain to the ‘algorithm itself’ and copyright cannot prevent others from creating other expressions of the ‘same algorithm’)
- the data analyzed by the trained algorithm (either because they contain copyright protected content or because the database is copyright protected)

N.B. During the preparatory phase of the profiling also *other* (not specifically machine learning/profiling related) *protected software* might be reproduced. During the exploitation phase it should be checked that no *protected graphic interfaces* are reproduced when communicating to users through an interface.

(art. 4 WCT; art. 1 CPD[36]) and certain aspects of a database may also be protected under copyright[37].

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). In the words of the WIPO Copyright Treaty: copyright protection extends to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such.[38]

> ***Copyright protection extends to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such. 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" and reflects the author's personality.***

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"[39] and reflects the author's personality[40]. 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[41].

As far as computer programs are concerned, the Computer Programs Directive provides that protection shall only apply to the "expression in any form of a computer program" and that "ideas and principles which underlie any element of a computer program, including those which underlie its interfaces, are not protected by copyright" (art. 1(2) CPD). This is further explained in the preamble that "[in] accordance with this principle of copyright, to the extent that logic, algorithms and programming languages comprise ideas and principles, those ideas and principles are not protected under this Directive" (cons. 11). A computer program is only protected if it is original, i.e. it is the author's own intellectual creation (art. 1(3) CPD).

[36] 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).

[37] 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);*

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

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

[40] 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 ;*

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

To the extent that logic, algorithms and programming languages comprise ideas and principles, those ideas and principles are not protected. A computer program is only protected if it is original, i.e. it is the author's own intellectual creation.

The CJEU has interpreted these provisions of the Computer Programs Directive in the cases *BSA*[42] and *SAS Institute*[43]. As a summary, it was decided that the source code and the object code, which permits reproduction in different computer languages, are protected elements of a computer program but the graphic user interface is not (it can be protected as a work under the general copyright rules). Furthermore, it was stated that "the functionality of a computer program nor the programming language and the format of data files used in a computer program in order to exploit certain of its functions constitute a form of expression of that program for the purposes of Article 1(2) of [Computer Programs Directive]"[44]. Similarly, the "programming language and the format of data files used in a computer program to interpret and execute application programs written by users and to read and write data in a specific format of data files, these are elements of that program by means of which users exploit certain functions of that program", which are consequently not protected under copyright[45].

Profilers such as OSNs or operators of profile transparency tools base their profiling services on *computer programs*, which include elements protected under the specific *copyright regime for computer programs* (such as source and object code), elements protected under the *general copyright rules* (such as graphic user interfaces) and other elements that cannot be protected under copyright (functionality, programming language, format data files). It is however required that the profiler has acquired these rights from the author/natural person or the third party developer (under a licence).

Profilers such as OSNs or operators of profile transparency tools base their profiling services on computer programs, which include elements protected under the **specific copyright regime for computer programs** ***(such as source and object code), elements protected under the*** **general copyright rules** ***(such as graphic user interfaces) and other elements that cannot be protected under copyright (functionality, programming language, format data files).***

[42] Judgment in *Bezpečnostní softwarová asociace*, C-393/09, ECLI:EU:C:2010:816, par. 35.
[43] Judgment in *SAS Institute*, C-406/10, ECLI:EU:C:2012:259, par. 29 et s.
[44] Judgment in *SAS Institute*, par. 39.
[45] Judgment in *SAS Institute*, par. 42.

While operating their profiling programs, OSNs or other profilers also create metadata, information about a OSN user on the basis of the photos, images, texts, clicks and likes she submits through the OSN service. Such information is not likely to be protected under copyright.

While operating their profiling programs, OSNs or other profilers also create metadata, information about an OSN user on the basis of the photos, images, texts, clicks and likes she submits through the OSN service. Such information is not likely to be protected under copyright. Where an OSN shares such metadata through its API with other profilers in the form of a raw data stream, no works in a protected form of expression are exchanged.

2.4.2. Is ‘mining’ protected under copyright?

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” and “distribution” or “communication to the public” (art. 2 and 3 of Directive 2001/29; art. 4 CPD). This way copyright allows authors, or anyone who is licensed by the author, to exploit the fruits of copyright protected content.

> 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). The copy of a source code of software is a reproduction in the sense of the Computer Programs Directive.
>
> 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.

Other uses are not restricted under copyright (e.g., copyright does not prevent anyone from reading a copyright protected work) or are exempted under a legal exception.

Reproduction. 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 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).

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 made by “search engines, either for indexing, for the 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”.

From a technical perspective the "copies" made in these new fields of application (ephemeral as they may be) are copies and consequently "reproductions" (cf. CJEU decision in Infopaq I), even if they are technically different from the "pirated copies" of a novel[46] 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).

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).

It is safe to assume that profilers perform acts protected under copyright while developing and offering their services to their users.

During the preparatory phase they may use third party software for developing their profiling solution. The USEMP partners have integrated various technologies in the DataBait solution (see D3.13 for a detailed list), including computer programs coming from third parties and they have cleared the copyright holders' consent for this use (even if some software was available under "open" licences). The computer programs of the OSNs are not reproduced (no efforts are made to reverse engineer the code developed and used by the OSNs).

Similarly, the USEMP partners used existing data sets (not belonging to the OSNs) to train their algorithms. The images, photographs and texts contained in these data sets were reproduced many times (while loading and running the databases), in order to recognise shapes (e.g. cigarettes, bottles of alcoholic beverages, outdoors sports) or to interpret images and texts (e.g. emotion recognition in pictures), which will then be reused in the DataBait system to offer profile transparency services to the OSN/DataBait user at a later stage.

At this preparatory stage, no one but the USEMP team has access to the protected works. Their acts consequently do not qualify as communications to the public or making available to the public.

Once the DataBait system is up and running, the DataBait user has access to her DataBait account, some of the content she has added to her OSN profile will be displayed in the DataBait context. This part is dealt with in D3.12.

Exceptions. 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

[46] 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 versions of the original text does not exclude these copies from copyright law.

as parodies. New forms of use are not always easily squeezed in the existing list of exceptions.

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[47] this had also led to the intention “to adapt copyright rules to new technological realities so that the rules continue to meet their objectives”. The current revision of Directive 2001/29 (the so-called “Copyright in the Information Society” or “InfoSoc” Directive) also includes a revision of the list of exceptions in Art. 5. 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 analysing 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 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.

[47] ‘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.

A data miner or profiler should demonstrate that her practice falls within one of the exceptions provided in Directive 2001/29. 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. If it cannot be established that the 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.

In its Communication of December 2015 the Commission also acknowledges that the present situation (i.e. wide rights under copyright and an exhaustive list of exceptions none of which covers text and data mining processes in all member states with certainty) may create uncertainties in the research community. This may harm the "EU's competitiveness and scientific leadership at a time when research and innovation (R&I) activities within the EU must increasingly take place through cross-border and cross-discipline collaboration and on a larger scale, in response to the major societal challenges that R&I addresses". Consequently, in its recent proposal[48] to revise the Infosoc Directive 2001/29/EC the Commission introduced an exception allowing "for reproductions and extractions made by research organisations in order to carry out text and data mining of works or other subject matter to which they have lawful access for the purposes of scientific research" (Art. 3(1) proposed Directive). The Directive defines text and data mining under Article 2 sub (2) as "any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations". In the two following sections we take a closer look at the question whether (a) this proposed exception for text and data mining (TDM) and/or (b) the existing exception for temporary reproductions (Art. 5(1) Infosoc Directive 2001/29/EC) and (c) the existing exception for scientific research (Art. 5(3a) Infosoc Directive 2001/29/EC) are of any help for DataBait (and similar transparency tools/research projects).

[48] Brussels, 14.9.2016 COM(2016) 593 final 2016/0280 (COD). Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on copyright in the Digital Single Market. Online available : http://ec.europa.eu/transparency/regdoc/rep/1/2016/EN/1-2016-593-EN-F1-1.PDF

(a) Would the proposed exception for text and data mining be useful for DataBait?

If the purpose of the text and data mining (TDM) exception is indeed to facilitate the machine driven reading and analysing of vast amounts of data for research purposes by public interest research organisations (provided that they have, as Art. 3(1) of the proposed Directive requires, "lawful access" to the data), then it could be hoped that projects such as USEMP can benefit from such exception. The reproductions are part of a data mining process, carried out in the context of academic research (on the profiling practices of commercial OSNs) by academic research institutions.

The usefulness of this exception may however be limited by two considerations. Firstly, the research partners should have "**lawful access**" to the data. Already the research can be complicated because the researchers have no **access** to the data the OSNs process[49]. The problem might get exacerbated by the clause in Art.3(3) of the proposed Directive:

> "Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject-matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective."

The proposed copyright exception for text and data mining could be useful for projects like USEMP. However, its usefulness could become limited because:

(a) access needs to be lawful. This could be problematic if:

- researchers have no factual access to data;
- access that is given to data is subjected to conditions – transgression of which makes the access unlawful.

(b) the three-step test of art. 5(5) of the porposed Directive (the same as article 5(5) in current Directive 2001/29) would still need to be passed. This includes the condition that the use should "not conflict with a normal exploitation of the work". In some cases that will be the case and the exception will not apply.

This clause might allow "publishers to introduce random measures to protect the 'security and integrity' of their network"[50] thereby simply rendering the effective use of

[49] The problem gets exacerbated by the clause in Art.3(3) of the proposed Directive : « Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject-matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.possibility for rightholders to neutralise the exception in practice through so-called security & integrity measures creates a gaping loophole for abuses (Article 3 par 3 & Recital 12): by allowing publishers to introduce random measures to protect the 'security and integrity' of their network, the effective use of TDM could simply be rendered impossible, or the use of the publishers own platforms could become the only viable alternative for researchers.

[50] http://copyright4creativity.eu/2016/09/28/text-and-data-mining-how-the-future-tdm-workshop-highlighted-the-draft-exception-must-be-improved-for-tdm-to-have-a-future-in-europe/ (last accessed 25 September 2016).

text and data mining exception impossible. The lack of actual access confronts researchers with a practical problem. In the case of USEMP, the research partners had access to user data coming from the users (who agreed that their data be shared through a browser plugin with the USEMP teams) and data coming from Facebook (delivered through an API Facebook makes available). The latter creates a rather large dependency on Facebook's approval of an app like DataBait. During the run time of the USEMP project the DataBait app was in a testing phase (and thus not publicly accessible), for which no approval of Facebook was needed. However, in order for DataBait to be publicly accessible to everyone, Facebook has to approve the application and is thus able to decline access to the data. In the first round of the submission procedure the DataBait app was rejected (September 2016). This is in itself not unusual, and after some adjustments apps might be accepted after all. Nevertheless, as Rieder notes:

> "[E]mpirical research on large online platforms is getting more and more difficult. Last year, Facebook removed a number of functions from their API, and research applications like *Netvizz* lost a part of their capacity to produce transparency by giving researchers the means to do (a certain kind of data-driven) empirical research"[51].

It could also be imagined that an OSN decides not to share any data or that it only makes available data sets that are too limited to be academically relevant.

In addition, Art. 3(1) of the proposed Directive requires the access should be "**lawful**". This raises the question whether and to which extent the holder of the data can impose conditions upon the access to the data that may restrict the research (especially in those cases where the holder of the data is the object of the research). An important question is whether the access is no longer lawful if any of the OSN's conditions is not respected with regard to the further use of the data. If this is the case, the unlawfulness of the access will render the exception inapplicable and the copying of the protected creations in the data set a copyright infringement.

Secondly, a distinction should probably be made between the data sets that are created in order to provide a service (an online social network service) and the data sets that are put together for the sake of testing. This matters to the extent that the application of each exception is subject to the **three-step test** (which can be found in art. 5(5) Directive 2001/29, art. 10 WCT and art. 13 TRIPs). The test as included in Article 13 TRIPs reads:

> "Members shall confine limitations and exceptions to exclusive rights (1) **to certain special cases** which (2) **do not conflict with a normal exploitation of the work** and (3) **do not unreasonably prejudice the legitimate interests of the rights holder**." (*bold and numbering added by the authors of this Deliverable*)

[51] Bernhard Rieder (May 27th 2016), « Closing APIs and the public scrutiny of very large online platforms », blogpost on *The Politics of Systems Thoughts on Software, Power, and Digital Method*, available online at : http://thepoliticsofsystems.net/ (last accessed 20 September 2016).

An exception for text and data mining may apply to a data set that is protected under copyright or under database rights, but any such exception will be subject to the three-step test. This means that the three-step test may not be met when the text and data mining exception is applied to data sets that were conceived for testing and improving algorithms. The second condition of the three-step test imposes indeed that the exception may not conflict with the normal exploitation of the work. If the normal exploitation is making the data set available for testing, then the impact of the newly created exception for text and data mining for research institution should be assessed on this form of exploitation.

The **exception for temporary acts of reproduction** (art. 5(1) Directive 2001/29)

exempts such reproductions, which are

(1) **transient or incidental** [and]

(2) an **integral and essential part of a technological process** and

(3) whose **sole purpose** is to enable (a) a **transmission** in a network between third parties by an intermediary, **or** (b) **a lawful use** of a work or other subject-matter to be made, and

(4) which have **no independent economic significance**

To assess whether the exception is applicable to DataBait or other transparency tools, it is particularly relevant to etablish (a) if there is a factual possibility to get acces to the data, (b) if the acces is lawful (data used in accordance with conditions set by OSN), (c) if it is possible to qualify the use of the copies as ‘temporary’ and (d) how the notion ‘technological process’ should be interpreted (Court has not done this yet).

(b) Is the existing exception for temporary acts of reproduction useful for DataBait?

For the time being (as long as there is no TDM exception), the existing exception for temporary reproductions (Art. 5(1) Infosoc Directive 2001/29/EC) and the one for scientific research (Art. 5(3a) Infosoc Directive 2001/29/EC) can apply to the text and data mining performed by research institutions. We will return to the scientific exception below (under (c)).

The mandatory exception for **temporary** acts of reproduction exempts such reproductions, **which are transient or incidental** [and] an **integral and essential part of a technological process** and whose **sole purpose is** to enable (a) a **transmission** in a network between third parties by an intermediary, **or** (b) **a lawful use** of a work or other subject-matter to be made, and which have **no independent economic significance** (art. 5(1) Directive 2001/29)[52] Temporary reproductions covered by the exception are, for example, “acts which enable browsing as well as

[52] S. DEPREEUW, "De uitzondering voor "tijdelijke technische reproductiehandelingen" na Infopaq I en II en Premier League", A&M 2013, 76-85; S. DEPREEUW, "Internet browsing dan toch zonder toestemming van de auteur (noot onder HvJ 5 juni 2014, zaak nr. C-360/13, Public Relations Consultants Association Ltd t. Newspaper Licensing Agency)", A&M 2015, 2, 172-174.

acts of caching" (Recital 33 of the preamble of Directive 2001/29) which do not have an independent economic significance. The Court of Justice of the EU has issued several decisions on the exception for temporary copies (Infopaq I & II, FAPL, Meltwater). Let's take a closer look at the four constitutive requirements which are relevant in the context of profiling (e.g., we do not look at copies whose sole purpose is a transmission – because this is not the purpose in profiling):

(1) The process must enable a « lawful use », i.e. be authorised by the right holder or not be restricted by law (Recital 33 of the preamble of Directive 2001/29). It was considered that where the applicable national or European legislation does not require the author's consent for drafting summaries of an article (Infopaq II) or the mere reception of a satellite broadcast by means of a decoder (Premier League, par. 171) or the browsing of web pages (Meltwater), the use is « lawful ».

(2) Temporary and transient acts of reproduction are "intended to enable the completion of a technological process of which it forms an integral and essential part (…) those acts of reproduction must not exceed what is necessary for the proper completion of that technological process" (Infopaq I, par. 61). A copy is 'incidental' if it "neither exists independently of, nor has a purpose independent of, the technological process of which it forms part" (Meltwater, par. 43). A copy is 'transient' if the period during which the copies remain in existence is limited to what is necessary for the proper functioning of the technological process used for achieving its purpose (see Meltwater, par. 46). Thus, it should be noted that the definition of temporary and transient copies is functional one, not one bound to a particular time limit in itself.

(3) Furthermore, the acts of reproduction in question must be an "integral and essential part of a technological process". It is required that (i) the acts of reproduction be carried out entirely in the context of the implementation of a technological process and (ii) the completion of those acts of reproduction be necessary, in that the technological process could not function correctly and efficiently without those acts (Meltwater, par. 28).

(4) Finally, the temporary copies must not have an independent economic significance. Many technical copies have economic significance but this does not prevent the application of the exception as long as it is not "independent", i.e. it does not go beyond the economic advantage derived from the use pursued. In other words, the economic advantage derived from their implementation must not be either distinct or separable from the economic advantage derived from the lawful use of the work concerned and it must not generate an additional economic advantage going beyond that derived from that use of the protected work (Meltwater, par. 50). This is in particular the case when the temporary copies form "an inseparable and non-autonomous part of the process" (Premier League, par. 174-178).

Thus, when applied to DataBait, **the first question is whether the copies made by the USEMP consortium during the profiling process enable a "lawful use"**, i.e. are authorised by the right holder or not restricted by law. Because OSN profiles often contain third-party content (for which no authorisation for reproduction is given) it is particularly important to check the second route: if the copies made in the DataBait profiling process can be qualified as a use "not restricted by law". The answer to the question whether a particular use is "not restricted by law" depends largely on the purpose that is pursued by this process.

The USEMP partners make copies of the data sets, and the protected creations they contain, for several purposes.

First, during the preparatory stage ("training of algorithms"), the text and data mining process is carried out to improve the DataBait software ("algorithms") while it is being developed. If a computer program is designed on the basis of automatically rendered information, it cannot be protected by copyright. However, in the case of the training of the DataBait algorithms quite some human creative effort is used. Consequently, we argue that the training of the DataBait algorithms should be understood as the development or improvement of a new creation, which may actually contribute to its protection under copyright. As such it is likely that this is an act that is not restricted under copyright law.

Secondly, once the DataBait algorithms have been trained, and can be used to analyse DataBait user data, the purpose of the text and data mining process is:

i. training the algorithms of their computer programs so these are more adequate to identify the content of an image or interpret a text when applied to OSN users' profiles and,

ii. analysing the creations uploaded to the profile of an OSN user. The purpose of this analysis is to interpret the content shared on OSNs and extract additional information from submitted images or text. Considering that information is not protected under copyright, it could be argued that the description of information by automated extraction means is an act that is not restricted under copyright law. Moreover the purpose of, the analysis of the OSN data extracted from the profiles of DataBait users is to help the OSN user to understand more clearly in which way her data are being processed by the OSN, which may facilitate the exercise of her data protection rights.

    The aforementioned purposes of DataBait's text and datamining processes are all likely to be uses "not restricted by law" and therefore lawful uses. At the same time, the CJEU has not clarified how the purpose of a process should be understood (how each "process" should be delineated, how the purpose of each process should be determined), so legal uncertainty remains.

The **second question is whether the copies made by DataBait are transient and/or incidental.** Here it is important to note that, as far as the storage of the data is concerned, the treatment of the OSN user's data during the exploitation of the DataBait software and the data sets during the development of the DataBait software are not the same.

Where the USEMP partners have acquired a data set and the protected creations that are part of it for testing all kinds of algorithms (including but not limited to the DataBait algorithms), they are likely to keep a permanent copy of the data set so it can be reused for various purposes. The copies made during the training of one particular computer program could arguably be transient.

We would also argue that the copies made of the OSN user's data are *temporary*, possibly even *transient*. This might sound paradoxical, because the functioning of DataBait requires *continuous* use[53]: copies of user data are kept on DataBait's "historical data" server until the end of the continuous purpose, that is, until the DataBait user stops using the DataBait application. In this sense even such continuous use can be qualified as transient: they are deleted once they are no longer necessary for DataBait's purpose (or: the process realising the purpose). Keeping these data on the server is necessary for a proper functioning of DataBait. Automatically deleting the data after every analysis and downloading again at a later stage, when another analysis is requested (by the OSN user) would not be feasible given the time it takes to download an average amount of profile data and the throttling policy of Facebook.

Thirdly, do the copies made by DataBait have an independent economic significance? The temporary copies made of the data sets and their protected content have no independent economic significance, in the sense that the temporary copies are restricted to the text and data mining analysis: they merely serve to complete this analysis. The value of the text and data mining process resides in (i) the training of the DataBait algorithms and the improving of the DataBait software and (ii) the information that is extracted from the content shared on the OSN. The temporary copies add no value to these purposes.

Finally, in order to answer **the question if the acts of reproduction for the DataBait profiling process are an "integral and essential part of a technological process"**, one uncertainty remains, i.e. the definition of the technological process of which the temporary copies are part. The Court gives not guidance concerning this important element of the exception for temporary acts of reproduction.

(c) Is the existing exception for scientific research useful for DataBait?

Where the exception for temporary acts of reproduction does not apply, the exception for scientific research (Art. 5(3a) Infosoc Directive 2001/29/EC) can still exempt the reproduction from the obligation to obtain the right holders' consent. Directive 2001/29 allows member states to provide an exception for the «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" (art. 5(3)(a) Directive 2001/29)[54].

[53] Within three months after the end of the USEMP project (1 October 2016) all historical data (which are kept on the « historical data » server at HWC) will be deleted manually. DataBait cannot function without these data. This implies that the DataBait user will not be able to use DataBait after the end of the project. However, if the DataBait user was to be enabled to continue using DataBait, this would require that the historical data were to be kept. If funding is found to make DataBait online available again after the USEMP project has ended all user data will have to be downloaded again to the DataBait server.

[54] See on the difficulties with the territorial application of this exception : J.P. TRIAILLE, S. DUSOLLIER, S. DEPREEUW, J.-B. HUBIN, F. COPPENS, & A. DE FRANCQUEN, *Study on the application of Directive 2001/29/EC on Copyright and Related Rights in the Information Society*, Brussels, 2013, 585 http://ec.europa.eu/internal_market/copyright/docs/studies/131216_study_en.pdf, p. 245 et s.

> The DataBait software is developed in the context of scientific research, conducted by several academic and non-academic partners, without a commercial purpose at this stage.
>
> This exception can arguably justify the temporary reproductions made of data set and user profiles (and their contents) but not the permanent reproduction of entire data sets, with the purpose of training indistinct algorithms beyond the scope of this project (see supra on the three-step test).

Authorisation. Where the use of protected content cannot be justified on the basis of an exception, the user needs the prior consent of the right holder.

> The USEMP partners have cleared the right holders' consent on (i) computer programs (either through proprietary licences or through open source licences) and on (ii) data sets (either by acquiring a licence from the entity commercialising the data set or on the basis of an « open licence » of sorts). As far as the open licences are concerned, the USEMP consortium is aware that the use of the same creations in the context of a follow up project should be examined again. It is indeed not self-evident that possible requirements of a non-commercial purpose or research purpose are still met.

One observation can be made concerning the licence between Facebook and the Facebook users, where the latter give a wide licence on the content they share with their friends through Facebook. 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*[55] (see the Annex for the full text of these two articles).

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

2.5. Profiling and database rights

Up until now we have focused on creations protected under copyright, held by the OSN or other profilers (computer programs, other elements protected under general copyright). 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)[56].

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).

IP rights on databases :

- **Copyright on a database** because the selection or arrangement of the contents constitute the author's own intellectual creation (Art. 3 DBD);
- **Sui generis right on a database** because there has been qualitatively and/or quantitatively a substantial investment, either in the obtaining, or in the verification or the presentation of the contents. (Art. 7 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, as we already touched upon in the previous section (2.4), 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 ("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[57]. 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).

Two 'objects' in the profiling process which can be relevant from the perspective of database right:

- the set of training and testing data (either because they contain copyright protected content or because the database is copyright protected)
- the data analyzed by the trained algorithm (either because they contain copyright protected content or because the database is copyright protected)

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 DBD).

[56] 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.
[57] CJEU, Football Dataco Ltd and Others v Yahoo! UK Ltd and Others, C-604/10, ECLI:EU:C:2012:115.

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[58]. By contrast, the Court of Justice of the European Union decided that "although the search for data and the verification of their accuracy at the time a database is created do not require the maker of that database to use particular resources because the data are those he created and are available to him, the fact remains that the collection of those data, their systematic or methodical arrangement in the database, the organisation of their individual accessibility and the verification of their accuracy throughout the operation of the database may require substantial investment in quantitative and/or qualitative terms within the meaning of Article 7(1) of the directive"[59].

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)

The investment must be directed to the obtaining, verification or presentation of the content – but not the creation of the content – of the database[60].

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). The notion of "extraction" should be interpreted widely (Directmedia[61], par. 32). The purpose of the

[58] See *inter alia* judgment in 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.

[59] Judgment in The British Horseracing Board Ltd and Others v William Hill Organization Ltd, ECLI:EU:C:2004:695, par. 36.

[60] Judgment in The British Horseracing Board Ltd and Others v William Hill Organization Ltd, ECLI:EU:C:2004:695, par. 40.

[61] Judgment in Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg, Case C-304/07, ECLI:EU:C:2008:552.

database right being to protect the investment made in the creation of the database, “the concept of extraction (…) must be understood as referring to any unauthorised act of appropriation of the whole or a part of the contents of a database” (Directmedia, par. 34). The Court adds: “the decisive criterion in this respect is to be found in the existence of an act of ‘transfer’ of all or part of the contents of the database concerned to another medium, whether of the same nature as the medium of that database or of a different nature. Such a transfer implies that all or part of the contents of a database are to be found in a medium other than that of the original database” (Directmedia, par. 36). It matters not whether the purpose of the extraction is to constitute a competing database. Also, the fact that the content of the protected database is found in modified form in another database does not preclude the finding of an extraction (Apis[62], par. 48).

The database right is infringed if a substantial part of the database is extracted or re-utilised. The concept of “substantial part, evaluated quantitatively, of the contents of a protected database refers to the volume of materials extracted from the database and/or re-utilised, and must be assessed in relation to the volume of the contents of the whole of that database. If a user extracts and/or re-utilises a quantitatively significant part of the contents of a database whose creation required the deployment of substantial resources, the investment in the extracted or re-utilised part is, proportionately, equally substantial” (BHB, par. 70, Apis, par. 59). The size of the database to which the copied part of the initial database is transferred does not matter (Apis, par. 60).

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).

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

[62] Judgment in Apis-Hristovich EOOD v Lakorda AD, C-45/07, ECLI:EU:C:2009:132.

(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.

Data sets. A brief reference can be made to the data sets that the USEMP partners use during the development phase of the DataBait tool. It can be assumed that these data sets are protected under the sui generis right (considering the efforts to collect images that share certain qualities, to add meta-data, etc.). The USEMP partners copy the entire data set, with the purpose of training the algorithms and improving the DataBait tool (in particular the accuracy of concept detection in the text and image mining processes). The consortium used ImageNet (a manually curated database) as well as other academic datasets[63] but also the web.[64] Some privacy related concepts were not well covered, hence additional resources were found in Flickr sets and Wikipedia. The permanent storage and the temporary copies made while loading the data from the databases can be seen as extractions from the databases. For this reason the USEMP partners have secured licences covering their use, where no open licence was available[65].

DataBait collects user data in two ways: through a browser plug-in and a Facebook application ('app')[66].

Browser plug-in. The data collected through the browser plug-in (installed on the browser the OSN user uses to access her profile or to use the OSN services) does not constitute an extraction. Due to the plug-in, the data that the user submits to the OSN are also submitted to the DataBait tool. DataBait does not copy the data from a database but copies the data that will be included in the database at the same time they are transferred to the OSN.

The data collected through the DataBait browser plug-in does not constitute an extraction.

The browser plugin method has some drawbacks: collecting data through the Facebook application programming interface (API) is technologically a much easier and smoother way, and the data collected through the browser plug-in do not cover

[63] See D3.13.
[64] See D5.2, p. 9.
[65] See D3.13.
[66] 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.

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.

Facebook API. Next to the collection of data through the browser plug-in, DataBait also collects data through a Facebook app. The DataBait Facebook app is a computer program created by the USEMP consortium, which runs on the Facebook platform and which Facebook users can choose to add (or remove) to their account. This DataBait Facebook app is not hosted by Facebook, 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). Everyone who develops a Facebook app, has to submit the app for review[67] to Facebook before it goes 'live'[68]. The DataBait app was submitted for review in the course of 2016. In September 2016 the DataBait application was rejected on the ground that Facebook considered that part of the description of the functionality of the app ("shows you who tracks you on the internet") is misleading. Based on this comment the USEMP consortium has requested more detailed feedback on why the app was rejected, explaining that the ground on which the rejection was based was unsound. . No results of this request for additional have been received so far. In the meanwhile the USEMP consortium adjusted the DataBait video and explanatory text in the "DataBait: How, What, Why?"-tab by removing the sentence "*Ever wondered who is tracking you* ?" (because this refers to the browsing behavior of the DataBait user, not her Facebook activity) and only leaving the sentence "*Ever wondered what information can be extracted from your data* ?". At the time of writing these adjustments have just been realized and the DataBait app will be resubmitted shortly. "

An application can, however, also be rejected on other grounds. 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. An important review criterion of Facebook is 'utility'[69] 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.

[67] https://developers.facebook.com/docs/facebook-login/review

[68] https://developers.facebook.com/docs/facebook-login/review/what-is-login-review: "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 not approved or you don't submit for review, people will not be able to use Facebook Login in your app.".

[69] https://developers.facebook.com/docs/facebook-login/review/what-is-login-review

> 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 Facebook API 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.

Even if there were an extraction that was not covered by a licence from the OSN, considering the case law of the CJEU summarised above, it is unlikely that any use by DataBait of the user data submitted to the OSNs constitutes the entire database held by the OSN or a substantial part of its contents.

Even if there were an extraction that was not covered by a licence from the OSN, considering the case law of the CJEU summarised above, it is unlikely that any use by DataBait of the user data submitted to the OSNs constitutes the entire database held by the OSN or a substantial part of its contents.

The number of DataBait users is indeed fairly small compared to the billion of Facebook users, hence their data (including the meta data produced by Facebook) are relatively insignificant too. It can therefore be argued that no infringing extraction of the contents of any protected database occurs.

In addition, in the case of the USEMP project, both the regimes of copyright and sui generis right provide exceptions with regard to scientific research

In addition, 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[70] 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

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

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;

(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[71]. Secondly, tools similar to the ones developed by USEMP which are used outside a scientific context are more likely to infringe database rights.

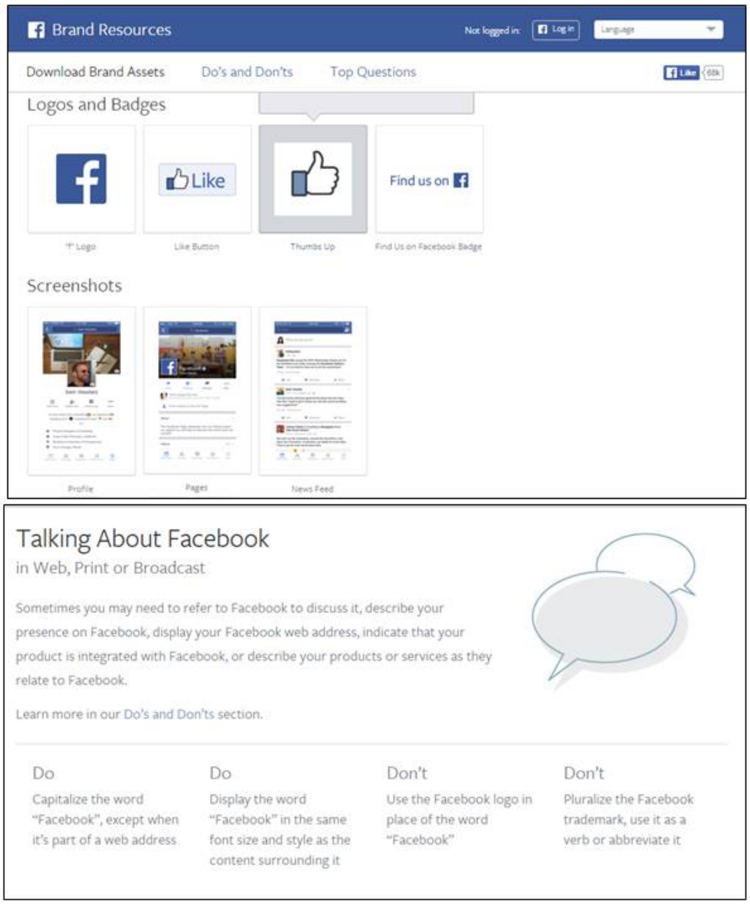
[71] Triaille e.a. (2014) studied the implementation of the scientific exceptions in the following member states : Netherlands, Germany, Poland, Luxembourg, Denmark, Hungary, Belgium, Spain, the UK and Italy. "The exception to copyright for scientific research in relation to databases contained in Article 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

In the case of DataBait, it can however be concluded that (i) no data are extracted from the OSN databases where the users volunteer their data to DataBait through a browser plug-in; (ii) no substantial part of the protected databases are extracted through the APIs and (iii) the extracted data from the training data sets and from the OSNs are covered by a licence. USEMP does therefore not infringe any database rights.

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. Profile transparency tools and trademark rights

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 adapted 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[72] 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.



Talking About Facebook

in Web, Print or Broadcast

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.

Learn more in our Do's and Don'ts section.

Do	Do	Don't	Don't
Capitalize the word "Facebook", except when it's part of a web address	Display the word "Facebook" in the same font size and style as the content surrounding it	Use the Facebook logo in place of the word "Facebook"	Pluralize the Facebook trademark, use it as a verb or abbreviate it

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

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

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/services or an identical or similar sign for identical or similar goods/service if that there exists a likelihood of confusion. A third type of infringement exists where the sign is identical with, or similar to, the EU trade mark and it is used in relation to goods or services which are identical with, similar to or not similar to those for which the EU trade mark is registered, where the latter has a reputation in the Union and where use of that sign without due cause takes unfair advantage of, or is detrimental to, the distinctive character or the repute of the EU 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. 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 USEMP consortium does not use the Facebook trademark to distinguish the goods and services it offers (it uses the name DataBait instead). The fundament right to freedom of expression should be considered in this latter case.

The USEMP consortium does not use the Facebook trademark to distinguish the goods and services it offers (it uses the name DataBait instead). 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.

One feature of the DataBait tool is "brands insights", which gives the user an overview of the brands she commonly engages with, e.g. because they are recognised in the pictures. Again, DataBait may display well known trademarks, but it does not use these trademarks to offer services or products of its own. It merely identifies the trademarks that can be recognised in the OSN/DataBait user. A priori this does not constitute any use that takes unfair advantage of the trademark or that harms the distinctive character or repute of the trade mark.

3.Conclusions

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

From the preceding analysis, it can be concluded that risk of DataBait infringing such IPRs or violating such trade secrets is quite low. The main reasons for this conclusion are that the USEMP consortium has not attempted to reverse engineer the profiling solutions or algorithms of the existing OSNs but has developed its own profiling tools instead.

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 is to 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 copyright protected software), this makes copyright or trade secret infringements on protected OSN software unlikely. It is unlikely that an OSN shares data through an API, which it wants to keep protected as a trade secret. The conditions of the reuse of the data can be explicitly set out in a licence agreement accompanying the API. The respect of such conditions will avoid any infringement of possible database rights.

The input for the analyses made by DataBait are user data collected through a browser plug-in and/or an OSN app. Considering that these data are not transferred to DataBait from an OSN owned database, there should be no risk that the database rights are infringed.

Trademarks owned by the OSNs and by third parties are displayed in DataBait. However, DataBait does not use these trademarks to indicate or promote goods or services it offers. It is rather a descriptive use relating to the profiling done by certain OSNs (which are identified by their trade mark) or relating to the brands that were recognised in the images the DataBait user has submitted to the OSN.

Finally, with regard to the way the research in this deliverable could be integrated in the DataBait tool, we concluded that informing DataBait users about the possible tensions between third party transparency tools (such as DataBait) and OSNs is possible but not necessary – there is a risk of information overload. However, on the DataBait developer website (mainly aimed at transparency tools developers but also at academics, policy makers and others interested in questions with regard to profile transparency) which will be launched shortly after the end of the USEMP project, it would be useful to give access to this report as it can provide guidance to anyone who wants to create an independent transparency tool.

This report, next to practical guidance it provides about IP issues to take into account when creating transparency tools, also gives some pointers to answer an academically and politically challenging question, namely: how should the copyright exception for text and data mining in the upcoming revision of the InfoSoc Directive look like? We conclude in this respect that such exception is very much needed. In the context of DataBait such exception would have been very helpful with regard to third-party content on OSN profile pages. Currently, the OSN user can license a tool like DataBait to reproduce user generated content of which she is the author. However, she cannot give such permission with regard to content of which she is not the author. The exception as proposed in the recent proposal could have been very useful in this respect, though we point to several aspects in its current formulation that might limit its protective scope and applicability.

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*[73] (Date of Last Revision: January 30, 2015):

Art. 8. Special Provisions Applicable to Developers/Operators of Applications and Websites

If you are a developer or operator of a Platform application or website or if you use Social Plugins, you must comply with the Facebook Platform Policy[74].

Art. 9. 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.

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

[74] Online available at: https://developers.facebook.com/policy. See below for some excerpts that seemed particularly relevant in the context of DataBait.

From the *Facebook Platform Policy*[75] (Date of Last Revision: May 26, 2016):

Art. 3. Protect data

1. Protect the information you receive from us against unauthorized access, use, or disclosure.
2. Only show data obtained from a user access token on the devices associated with that token.
3. Only use friend data (including friends list) in the person's experience in your app.
4. If you cache data you receive from us, use it to improve your app's user experience and keep it up to date.
5. Don't proxy, request or collect Facebook usernames or passwords.
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 ourmechanism.
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.
14. If you use friend data from Facebook to establish social connections in your app, only do so if each person in that connection has granted you access to that information.
15. Don't use data obtained from Facebook to make decisions about eligibility, including whether to approve or reject an application or how much interest to charge on a loan.

[75] Online available at: <https://developers.facebook.com/policy>

Art. 4. Encourage proper use

[...]

2. Respect the way Facebook looks and functions. Don't offer experiences that change it.

[...]

4. Respect the limits we've placed on Facebook functionality.

[…]

10. Don't sell, transfer or sublicense our code, APIs, or tools to anyone.

[…]

12. Don’t modify, translate, create derivative works of, or reverse engineer any SDK or its components.

13. Be honest about your relationship with Facebook when talking to the press or users. Comply with our Developer PR Guidelines and get approval from us before issuing any formal press release or blog post mentioning Facebook.

Art. 5. Follow the law

1. You are responsible for restricting access to your content in accordance with all applicable laws and regulations, including geo-filtering or age-gating access where required.

2. Don’t provide or promote content that infringes upon the rights of any third party.

3. Ensure that you own or secure all rights necessary to display, distribute and deliver all content in your app.

4. Satisfy all licensing, reporting and payout obligations to third parties in connection with your app.

5. If your app contains content submitted or provided by third parties:

a. In the United States, you must take all steps required to fall within the applicable safe harbors of the Digital Millennium Copyright Act including designating an agent to receive notices of claimed infringement, instituting a repeat infringer termination policy and implementing a notice and takedown process.

b. In other countries, you must comply with local copyright laws and implement an appropriate notice and takedown process for when you receive a notice of claimed infringement.

6. Don’t knowingly share information with us that you have collected from children under the age of 13.

7. Web sites or services directed to children under 13: If you use Social Plugins or our JavaScript SDK for Facebook on sites and services that are directed to children under 13, you are responsible for complying with all applicable laws. For example, if your web site or service is directed to children in the United States, or knowingly collects personal information

from children in the United States, you must comply with the U.S. Children's Online Privacy Protection Act. You must also adhere to our usage notes.

8. Comply with all applicable laws and regulations in the jurisdiction where your app is available. Do not expose Facebook or people who use Facebook to harm or legal liability as determined by us in our sole discretion.

[…]

10. You agree to indemnify and hold us harmless from and against all damages, losses, and expenses of any kind (including reasonable legal fees and costs) related to any claim against us related to your service, actions, content or information.

Art. 6. Things you should know

1. We can analyze your app, website, content, and data for any purpose, including commercial.

2. We can monitor or collect data related to your use of SDKs.

3. We will use information we receive from you or in connection with your Platform integration in accordance with our Data Policy.

4. You give us all rights necessary to enable your app to work with Facebook, including the right to incorporate information you provide to us into other parts of Facebook, and the right to attribute the source of information using your name or logos.

5. We may share your contact info with people who want to contact you.

6. We may use your name, logos, content, and information, including screenshots and video captures of your app, to demonstrate or feature your use of Facebook, worldwide and royalty-free.

7. You give us the right to link to or frame your app, and place content, including ads, around your app. If you use our social plugins, feed dialog or share button, you also give us permission to use and allow others to use such links and content on Facebook.

8. We can audit your app to ensure it is safe and does not violate our Terms. If requested, you must provide us with proof that your app complies with our terms.

9. We can create apps or products that offer features and services similar to your app.

10. We don't guarantee that Platform will always be free.

11. If you exceed 5M MAU, 100M API calls per day, or 50M impressions per day, you may be subject to additional terms.

12. Facebook and its licensors reserve all right, title and interest, including all intellectual property and other proprietary rights, in and to all SDKs.

13. Any SDKs you receive from us are provided to you on an "as is" basis, without warranty of any kind.

14. We can issue a press release describing our relationship with you.

15. We may enforce against your app or web site if we conclude that your app violates our terms or is negatively impacting the Platform. We may or may not notify you in advance.

16. Enforcement is both automated and manual, and can include disabling your app, restricting you and your app's access to platform functionality, requiring that you delete data, terminating our agreements with you or any other action that we deem appropriate.

17. We communicate with developers through Developer Alerts and email from the fb.com or facebookmail.com domain. Ensure that the email address associated with your Facebook account and the email address registered to the app are current and that you don't filter out these messages.

18. We may change these terms at any time without prior notice. Please check them regularly. Your continued use of Platform constitutes acceptance of those changes.

19. Your use of Facebook technology is subject to this Platform Policy, our Statement of Rights and Responsibilities and any other terms that apply to the applicable technology.

7. Login

[…]

4. Request only the data and publishing permissions your app needs.

[…]

Annex B: table summarizing the conclusions of chapter 2

IPRs 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?	*No.* 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. 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 contractual clauses (in case the extraction is enabled by the OSN/browser). 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 it may indeed be patentable.
	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 (such as DataBait)?	*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 cover only *commercial* use of the patented invention. Moreover, 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

		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) 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 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.
	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. 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.
Copyright and sui generis	Do profile transparency tools like the ones	*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

right in databases	created by USEMP reproduce (parts) of the overall structured way in which data are organized by, for example, Facebook?	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, 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.
	Is it possible that the DataBait profiling process infringes on any OSN database rights (the creator and/or 'maker-through-substantial-investment' of the database)?	*No.* 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 user provides her data, and these might contain some structure provided by the OSN), the OSN might invoke exclusive database rights. Both the regimes of copyright and sui generis right provide optional 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[76] to fall under the exceptions in Art 6(2b) and Art. 9(b) of the Database Directive – however, as these exceptions are optional they have not been transposed in every national jurisdiction. Another caveat is that tools similar to the ones developed by USEMP which are used outside a scientific context are more likely to infringe database rights. However, looking at the case law in this field the extraction made by USEMP will not qualify as substantial parts of the database.
Trademarks	Can the reproduction of an OSN logo within DataBait infringe on the trademark of that OSN?	*No.* 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.

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