FP7-SCIENCE-IN-SOCIETY-2013-1

Grant Agreement Number 612345

Supporting Action

MAPPING

WP5 Privacy, Personality and Business Models

WP leader: LUH

D5.3
Round-Table on Future Scenarios of Personal Data Use in Commercial and Non-commercial Settings

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	**x**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document Version Control:

Version 0.1	Originated by:	CM Hawellek	On 23.02.2016 at 15h59
Version 0.2	Modified by:	F Knoke	On 24.02.2016 at 12h35
Version 0.3	Modified by:	J Stoklas	On 24.02.2016 at 16h30
Version 0.4	Modified by:	F Knoke	On 02.03.2016 at 14h55
Version 1.0	Revised by:	N Forgó	On 27.04.2016 at 16h51
Version 1.1	Modified by:	F Knoke	On 20.05.2016 at 19h10
Version 1.2	Revised by:	CM Hawellek	On 20.05.2016 at 20h00
Version 2.0	Revised by:	N Forgó	On 23.05.2016 at 16h07
Version 3.0	Revised by:	A.M Snijders	On 01.06.2016 at 17h11

TABLE OF CONTENTS

ACRONYMS

CoE	Council of Europe
CSA	Coordination and Support Action
DoW	Description of Work
EC	European Commission
EU	European Union
FP	Framework Programme
LUH	Leibniz Universität Hannover
MAPPING	Managing Alternatives for Privacy, Property and Internet Government
MML	Mobilisation and Mutual Learning
OSINT	Open-source intelligence
SiS	Science in Society
WP	Work Package

1. ABSTRACT

The MAPPING round table on *Future Scenarios of Personal Data Use in Commercial and Non-Commercial Settings*, deliverable D5.3 of the MAPPING Project, took place on 18 February 2016 in Malta. It was organised by the project's Work Package 5 (Privacy, Personality & Business Models) leader Leibniz Universität Hannover (DE).

The major theme framing the entire event was "predictive analytics" as a subject identified as one of the most controversial current and future scenarios with regard to personal data. The general methodological idea was to unite experts from different market segments of the private sector and from the public sector representing various backgrounds (health sector, law enforcement, intelligence services, internet economy, academia, civil society/NGOS, policy makers, etc.) and representing various disciplines (law, sociology, medicine, mathematics, information technology, etc.). The common denominator was the application of predictive analytics as such, while the field of application was widely varied (from predictive medicine over search engine optimisation to predictive policing). The objective was to foster cross-fertilisation, identify common issues and concerns, and to identify potentials for solutions practised in one sector to apply to other sectors as well.

The meeting involved 38 participants. It consisted of four 90-minute blocks, each dedicated to discussing one use case. This structure offered participants the opportunity to listen to expert speakers, before being encouraged to actively participate in debate and discussion, to exchange views based on their particular background and related expertise.

The four use cases were:

- **Predictive Analytics in the Health Sector,**

- **Auto-complete Function for Search Engines (Incremental UI Fields)**

- **Predictive Policing**, and

- **Software Solutions for Forensic and Analytic Applications.**

These blocks were separated by short breaks throughout the day that allowed for informal and productive conversation between attendees and encouraged further engagement with the MAPPING

project beyond the meeting itself. The MAPPING coordinating person and the coordinator of WP5 jointly wrapped up the entire event at the close of the meeting.

In order to disseminate the round table, the following news entry was published on the MAPPING website:

"Future Scenarios of Personal Data Use in Commercial and Non-Commercial Settings

The Round Table on Future Scenarios of Personal Data Use in Commercial and Non-Commercial Settings took place on 18 February 2016 in Malta. It was organised by the MAPPING's Work Package 5 (Privacy, Personality & Business Models) leader, the Institute for Legal Informatics, Leibniz Universität Hannover (DE). Full details of the Round Table are presented in deliverable D5.3 of the MAPPING Project.

The event consisted of four ninety minute discussions on predictive analytics Use Cases. Each Case was briefly presented in a two-fold view that offered participants the opportunity to think about benefits as well as privacy drawbacks and side-effects. The participants were encouraged to actively participate in an engaging debate and discussion.

Use Case I - Predictive Analytics in the Health Sector questioned the use of medical data for research. The debate tackled ethical issues, liability of medical staff, commercialisation, anonymization and security of data, and how to comply with and satisfy legal frameworks.

Use Case II – Auto-complete Function for Search Engines opened a debate regarding factors that influence searches, challenges of auto-complete functions that may cause negativity bias and self-referential loop, and finally included a debate about privacy related issues of search history.

The afternoon session was dedicated to the use of personal and statistical data in crime prevention and investigation. Both Use Cases showed different examples of pre-crime forecasting. Use Case III – Predictive Policing presented an example from Lower Saxony. The final Use Case, Software Solutions for Forensic and Analytic Applications, demonstrated results of testing of evidence-based policing in a hypothetical city by using a real data sample from South Wales."

The news entry is available at https://mappingtheinternet.eu/Future-Scenarios-of-Personal-Data-Use-in-Commercial-and-Non-Commercial-Settings .

The meeting was held under the Chatham House Rule plus. WP5 Stakeholder Meetings (as all other stakeholder meetings in the MAPPING framework) were held following the 'Chatham House Rule' to encourage a fruitful debate without fear of participants being misquoted outside of the debate. The rule states that "When a meeting, or part thereof, is held under the Chatham House Rule, participants are

free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed." - See more at: https://www.chathamhouse.org/about/chatham-house-rule#sthash.JI592zMY.dpuf".

The opinions and views expressed in this document are views of participants, not of LUH nor of the consortium.

1.1   PREDICTIVE ANALYTICS IN THE HEALTH SECTOR

The first speaker presented Use Case I, introducing the first session. He illustrated relevant developments and changes that have taken place in the health sector, which is, according to the speaker, evolving from treating symptoms with the aid of 1st generation diagnostics into personalized healthcare focusing on preventing diseases using automated systems. He pointed out that this required much data, and that it bore various challenges, such as legal and ethical issues, the need for logistics (e.g. for handling and curation of the data), rules and procedures for usage for research, and for clinical driven, evaluated, validated, standardised and certified tools and models.

The presentation ignited a lively debate that addressed the difficulties of enhancing privacy via technology in this field. As major obstacles, high costs and difficult implementation of privacy-enhancing technologies were mentioned. There were, according to discussants, many uncertainties, such as questionable IT security standards in medical facilities, which required professionalization of data centers in that respect. One participant pointed out that complete data security was an illusion. If, consequently, always a risk remains, it should be evaluated regarding which applications such a risk would be acceptable. Furthermore, as one participant commented, understanding the threat became more and more difficult with the capabilities being built with predictive technologies. Without being able to assess the impact of the threat, it became difficult to evaluate which safeguards were adequate. One participant stated that with regard to future uses one could only have either anonymisation – or consent. In the course of the discussion, the meaningfulness of consent in the times of big data was debated. It was expressed that consent appeared to be a concept of the 20th century, outdated by powerful methods of data correlation. One participant criticized that in practice it looks like the consent form serves as a tool limiting liability of physicians instead of protecting patients.

On top of that, the plenum tackled the ethical dimension of predictive analytics in the health sector, especially with respect to use of medical data for research. It was brought up that, given the need of personal data to foster research in fields such as oncology, the slogan "donate your data" was used[1], which raised the question whether one actually was able to assess what was the subject of donation really, in the meaning of fully understanding possible consequences.

In the course of the debate, participants also discussed possible solutions for the issues that were raised. For instance, one participant considered increasing transparency as a promising starting point, because much data is being collected via health apps on mobiles, but there is a lack of information what this data might be used for and with whom it is going to be shared. In general, improving patient-empowerment was deemed necessary by participants, while acknowledging that details admittedly are controversial, e.g. in respect of how to actually build systems depending on personal data from various sources in different states with all kinds of varying legal regimes. One participated pointed out that currently there is a lot of debate going on, but that in his opinion, on top of that, actually taking action is needed.

1.2 AUTO-COMPLETE FUNCTION FOR SEARCH ENGINES (INCREMENTAL UI FIELDS)

In the presentation of use case II, opening the second session, the functioning of auto-complete functions for search engines was explained. The speaker stated that users benefit of the technology by constant feedback (regarding spelling mistakes, narrowing of the search), saving time (because ideally there are fewer attempts needed to find what was desired and it is not necessary to type words completely (esp. on mobile devices)). As challenges the speaker presented the existence of a filter (which might result in a self-referential loop), the possibility of privacy concerns, especially regarding the search history, and the possible intensification of infringement of personality rights ("bad auto-completions").

In the discussion that followed the presentation, one participant raised the question whether, when the search engine was used for the first time, it is necessary to be asked whether the user would like to have his or her searches saved or not, instead of having a built-in technology, where it is not possible to

---

[1] See e.g. http://pharmaviews.eu/donate-your-data/, http://datadonors.org/mission/.

opt out of this service. It was stated that users could choose to delete their own searches, so that only terms based on searches of the entire user community would be proposed. One participant commented that all users were aware that they were being profiled and their data is used; since the privacy policy of the company offering the search engine changes so often, the vast majority is not conscious of what they consent to, which makes the meaningfulness of consent questionable in this respect.

Furthermore, several participants critically judged the influence of this technology on users' thinking, as well as, and consequently, on businesses' behaviour. Such technology e.g. might redirect the users' interest to products suggested by the auto-complete function. As a result, the technology holds a significant potential to take influence on markets. One participant pointed out that excluding certain words from auto-complete for legal reasons for e.g. risks to approximate a "privatized law enforcement".

The plenum also addressed possible impact of the auto-complete function on politics. It was stated that on the assumption that the technology actually allows to take influence on people's behaviour, there is a potential risk of influencing public opinion, e.g. before elections. This is true especially if certain search terms are suggested jointly (and thereby not only the users' focus is redirected, but also it is indicated that these subjects were what the majority was interested in), and if results are provided in a particular ranking (and thereby e.g. obfuscating certain accounts or opinions). On the other hand, the company itself is subject to attempted manipulations, and needs to spend large efforts and maintaining ability to detect and avert such attempted manipulation. As one participant stated, this puts the company into the powerful position to react or not react; including pulling-out of a country, if the government of this country tries to take influence by imposing conditions. The two sides of the coin are that this would not only allow to take positive influence in autocratic states, but also put a solitary private entity to largely influence states, which involves a dangerous momentum as well, if used less altruistically.

1.3   PREDICTIVE POLICING

After the lunch break, use case III – Predictive Policing – was presented. The technology allows the police to warn of particular dangers in certain areas according to experiences from existing data sets in order to prevent crimes in the context of certain events. The speaker reported on his experiences and presented open questions from a law enforcement perspective. According to him, this technology is still in the process of gaining acceptance among both police officers and citizens – a process thought of as

difficult since success of predictive policing is hard to measure, and, what is more, there is a lack of scientific studies addressing the topic. Additionally, he indicated it would be necessary to invest in technical and human resources (such as hiring additional staff and offering trainings).

After this introduction, the participants of the roundtable exchanged their views. One participant commented that prediction models have to be further developed and that in those models in particular such crime types as car thefts and burglary, where dark figures are relatively low, need to be included, because predictive policing heavily relies on sound databases. It was stated that open-source intelligence (OSINT) tools, collecting information from publicly available sources, could be used to gather information for predictive policing. The importance of democratic legitimation was pointed out, and that it could be helpful in this respect to establish a set of political rules for the use of technical equipment. As to that, according to one participant, it should be born in mind that the attitude towards police authorities differs from country to country in terms of trust and acceptance, which requires a legal debate on international level in preparation of legislative measures. In many cases, as one participant explained, information is already available; the challenge is to make it adequately available to police officers prior to taking action. Sometimes, another participant argued, reduced data input may result in higher accuracy of results. In a test run for the prediction of burglary, out of 60 factors only less than 5 factors seemed to be of relevance such as the number of people living in the respective area, the crime level of the last week, and the crime level of the last year.

Implications of the use of personal data for predictive policing were lively discussed. Using personal data would be quite helpful from a policing point of view. The distinction between personal and non-personal data, so one participant stated, might be also be difficult to make due to the fact that predictive policing could produce personal data by joining various data sets. Furthermore, targeting individuals with predictive policing was confirmed to be technically possible, if additional information on that individual is available, such as information on mental health issues. Legally, however, this was considered to be highly problematic, and that apart arguably ethically questionable.

1.4 SOFTWARE SOLUTIONS FOR FORENSIC AND ANALYTIC APPLICATIONS

Finally, use case IV, Software Solutions for Forensic and Analytic Applications, was presented. This use case comprised testing evidence-based policing in a hypothetical city by using a real data sample and a related demonstration. In this project, so it was explained, the work built upon statistics, adding known

human behaviour. The software aimed at directing the police's activities as a next generation of predictive policing by supporting the prediction and planning, and eventually resulting action. Consequently, the software was supposed to support more effective and efficient operations, by augmenting resource management in terms of deployment to the right place at the right time. Divided by themes (burglaries, violent crimes, road traffic etc.), the software provided a sort of crime/danger forecasting. According to the speakers, it could be developed towards helping presenting evidence in court by allowing illustrating complex evidence to the jury.

In the following discussion, the plenum acknowledged the technical excellence of the software, but pointed out the importance of understanding the privacy issues involved as well as possible links to Internet governance. One participant stated that in their opinion, trust in such software is the pre-condition for its acceptance; which means trust by all citizens, policy makers and law enforcement. It is therefore of elementary importance to look at the ethics and the risks of such a technology, and in particular to conduct a proper proportionality assessment. The participants voiced different criticisms: If running this software is offered as a service by a private entity, one participant claimed, this involves transferring data from police computer systems to that private entity, which results in various data protection risks. In addition, a possible negative effect on society was brought up, questioning the benefit for citizens. One participant said it would be interesting to know how open and transparent the district/community hosting the referenced trial of this software is towards citizens about this project.

Furthermore, it was stated that providing information created by computer systems using complex algorithms have a strong psychological effect on investigators, who could not evaluate the results. Difficulties in reasoning against the machine's suggestions, potential higher risks of having to justify if deciding against the suggestion of the machine and resulting limitation in distrusting the machine's suggestions might result in reduced critical questioning up to unreflectingly following the decision of a machine. One participant expressed the view that the software is simulation, not prediction, and that simulation is much stronger than prediction. With regard to the psychological effect, the need to study the impact of such systems from a cognitive sciences' point of view was uttered. A participant referred to Germany where there was a huge debate on automated decision-making, and pointed out that such a system must not replace an officer. Finally, the legal implications were discussed, such as whether or not it would be possible/thinkable to issue warrants based upon outputs of the software.

2. ANNEX: AGENDA

"ROUNDTABLE ON FUTURE SCENARIOS OF PERSONAL DATA USE IN COMMERCIAL AND NON-COMMERCIAL SETTINGS"

– PREDICTIVE ANALYTICS –

18 February 2016, Radisson Blu St. Julian's, Malta

From 08:30: *Arrival and Registration*

09:00: Opening / Welcome

09:15: Introduction

09:30: **Discussing Use Case I – Predictive Analytics in the Health Sector**

11:00: *Coffee Break*

11:30: **Discussing Use Case II – Auto-complete Function for Search Engines**

13:00: *Lunch Break*

14:00: **Discussing Use Case III – Predictive Policing**

15:30: *Coffee Break*

16:00: **Discussing Use Case IV – Software Solutions for Forensic and Analytic Applications**

17:30: Wrap-up and Closing Remarks

18:00: *End of Meeting*