Legal Feasibility of Automated Data Collection for Statistical Purposes in the EU

Dr Faye Fangfei Wang
Senior Lecturer in Law; Brunel Law School, Brunel University (West London), UK
faye.wang@brunel.ac.uk

Abstract. Internet-based measurement is using Internet as source of data gathering and it is a method of automated data collection. The three most common Internet-based measurement approaches are user-centric, network-centric and site-centric measurements. User-centric relies on in-depth analysis of behaviour of users by installing software and application; network-centric measures traffic flows between users and content throughout the network; and site-centric collects data from one to more websites. Internet as a source of data gathering could lower the costs and increase the speed of data collection for statistical purposes compared with traditional manual methods. Statistics analysis is important as it may contribute to value added service. On the other hand, data privacy rights may be at risk under such measurement approaches if technical measures for data security are not appropriate or users did not give prior consent to the use of such data. This paper discusses the current EU data privacy protection legislation and analyses the overall legal feasibility of the development of Internet-based measurements with regard to automated data collection for statistical purposes by looking into the detail of the reform of the new EC Directive on e-Privacy and the current review of the EC Directive on Data Protection.

Keywords: Data Privacy Protection, Automated Data Collection, Statistical Research

1 Current EU Legal Framework on Online Privacy

“Internet-based measurement is a set of methods that have been applied to quantitatively describe the structure, workload and use of the Internet. They provide a practical means of doing a kind of virtual ‘fieldwork’ on the Internet using online tools and network monitoring techniques to gather fine scale primary data, as opposed to relying on aggregate secondary data sources (such as government statistics).”

1 International Encyclopaedia of Human Geography, MS number: 457.
In other words, Internet-based measurement is using Internet as source of data gathering and it is a method of automated data collection. The three most common Internet-based measurement approaches are user-centric, network-centric and site-centric measurements. User-centric relies on in-depth analysis of behaviour of users by installing software and application; network-centric measures traffic flows between users and content throughout the network; and site-centric collects data from one to more websites.\(^2\) Internet as a source of data gathering could lower the costs and increase the speed of data collection for statistical purposes compared with traditional manual methods. Statistics analysis is important as it may contribute to value added service, which may, for example, “consist of advice on least expensive tariff packages, route guidance, traffic information, weather forecasts and tourist information”.\(^3\) On the other hand, data privacy rights may be infringed under such measurement approaches if technical measures for data security are not appropriate or users did not give prior consent to the use of such data.

Due to the ever fast-growing technology, legislation is always one step behind the latest invention of computing network services. This leads to a situation where computer scientists and entrepreneurs try to adjust or improve the application of products in order to comply with the existing law, or legislators try to amend the existing law in response to the new technology in order to protect the users’ rights and enhance the public safety without jeopardising technological innovation and market development. Currently, there are three main pieces of legislation concerning data and privacy protection in the European Union (EU): 1) EC Directive on Data Protection in 1995;\(^4\) 2) EC Directive on e-Privacy in 2002;\(^5\) and 3) EC Regulation on Personal Data Protection in 2001.\(^6\)

The EC Directive on Data Protection has been under the review of European Commission since 2009. In 2011, the Commission will propose a new general legal framework for the protection of personal data in the EU, covering data processing operations in all sectors and policies of the EU. This framework will then be negotiated and adopted by the European Parliament and the Council. The EC

---


\(^3\) Recital 18 of the EC Directive on e-Privacy.


\(^6\) Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data, OJ L 008, 12/01/2001 P. 0001 – 0022.
Directive on e-Privacy has been amended by the Directive 2009/136/EC. Member states are required to implement the new EC Directive on e-Privacy by May 25, 2011.7

To understand the legal feasibility of the adoption of the three common measurement approaches for statistical research, basic legal concepts that are arisen from the approaches shall be firstly identified and interpreted, for example, the definitions of personal data and sensitive personal data. Personal data is defined as “any information relating to an identified or identifiable natural person (‘data subject’); and identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, culture or social identity”;8 whilst sensitive personal data could be understood as “personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life”.9

The possibility of the implementation of automated data collection for statistical research in business organisations and statistical institutions depends on the feasibility of legal compliance. It is debatable whether automated data collection for statistical research is allowed under the current EU data privacy legislative framework. This paper intends to provide an overview of the EU data privacy protection legislation and discuss the overall legal feasibility of the development of Internet-based measurements with regard to automated data collection for statistical purposes by looking into the detail of the reform of the new EC Directive on e-Privacy and the review of the EC Directive on Data Protection.

2 Necessary Legal Compliance for Automated Data Collection for Statistical Purposes

According the current EU data privacy protection legal framework, there are four underlying steps in the EC directives that intend to ensure that privacy rights are put into action:

1) Member states should take appropriate technological and legislative measures to safeguard security and ensure the protection of personal data and privacy.

8 Article 2(a) of the EC Directive on Data Protection.
9 Article 8(1) of the EC Directive on Data Protection.
2) Service providers have a legal duty to inform users prior to obtaining their consent.

3) Service providers shall allow users to give and withdraw their consent freely as users have “the right to be forgotten”. It is debatable what constitutes a meaningful consent and whether “privacy by default” is sufficient.

4) Member States shall enhance enforcement of data privacy protection because any legislative and technological measures to protect users’ privacy can only be effective if they are properly implemented and enforced. EU citizens’ data privacy rights should be protected equally no matter where the service provider and data are situated. The service provider shall duly notify data breach and take appropriate measures to avoid escalation of the problem.

Among the above steps, the rightful implementation of consent determines the lawful processing of data. Member states may understand the valid form of consent differently, for example, the UK law interprets consent as ‘reasonable grounds for believing’ that consent to do, which do not comply with EU rules defining consent as “freely given specific and informed indication of a person’s wishes”. Under the new EC Directive on e-Privacy, the use of cookies requires users’ prior consent. Article 29 Working Group on Data Protection addressed that “currently three out of the four most widely used browsers have a default setting to accept all cookies. Not changing a default setting cannot be considered as a meaningful consent.” It is expected that there are various interpretations by member states when the new EC Directive on e-Privacy comes into force in May 2011.

With regard to data collection for statistical purposes, the exemption clause of prior consent remains unchanged in the new legislation. In general, data can be processed solely for the purpose of scientific research or kept in personal form for a period which does not exceed the period necessary for the sole purpose of creating statistics subject to the implementation of conditions:

1) adequate legal safeguards – the data are not used for taking measures or decisions regarding any particular individual;

2) clearly no risk of breaching the privacy of the data subject;

3) data kept only for necessary period and employ other appropriate safeguards provided by member states.

10 Recital 17 of the EC Directive on e-Privacy.
12 Recital 29, 39 & 40 and Article 11(2) & Article 13 of the EC Directive on Data Protection.
That is, automated data collection from the Internet for statistical purposes could be legitimately processed, provided that they fulfil the above three basic conditions except for the category of processing sensitive personal data that needs to meet the additional condition of public interest. There are four different layers of data collection that could possibly be used for statistical purposes: first, general personal data; second, further processing of personal data previously collected; third, data that are not obtained from the end users directly; and fourth sensitive personal data.

As to the last but not least important issue – enforcement of data privacy protection, service providers shall duly notify data breach to the competent national authorities and take appropriate measures to protect data privacy. In the author’s opinion, the interpretation of “without undue delay” for data breach notification under Article 4 of the EC Directive on e-Privacy is vital as the timing affects the certainty of data-privacy protection. The determination of the appropriation of time limit on notification and remedial action shall be taken into account of the speed, scope and capabilities of spreading personal data under the current and future development of technologies in particular automated information systems. In addition, the consideration of the time-limit issue for notification and remedial action can be learned from the interpretation of the time-limit requirement on the exercise of the right to access in Article 12(a) of the EC Directive on Data Protection regarding information storage and disclosure in the case of College van burgemeester en wethouders van Rotterdam v M.E.E. Rijkeboer Netherlands (judgement of 7 May 2009). The judgement provides that:

“Article 12(a) of Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data requires Member States to ensure a right of access to information on the recipients or categories of recipient of personal data and on the content of the data disclosed not only in respect of the present but also in respect of the past. It is for Member States to fix a time-limit for storage of that information and to provide for access to that information which constitutes a fair balance between, on the one hand, the interest of the data subject in protecting his privacy, in particular by way of his rights to object and to bring legal proceedings and, on the other, the burden which the obligation to store that information represents for the controller.

Rules limiting the storage of information on the recipients or categories of recipient of personal data and on the content of the data disclosed to a period of one year and correspondingly limiting access to that information, while basic data is stored for a much longer period, do not constitute a fair balance of the interest and obligation at issue, unless it can be shown that longer storage of that information would constitute an

---

13 Recital 34 and Article 8 of the EC Directive on Data Protection.
excessive burden on the controller. It is, however, for national courts to make the determinations necessary.”

Accordingly, it shall be for Member States to fix a time-limit for notification of the personal data breach and remedial action. Where the length of time for which a personal data breach is to be informed to the competent national authority or remedial action is to be taken is very long, the adverse effects of the breach of the personal data or privacy of a subscriber or individual may be higher as the implementation of appropriate technological protection measures may be delayed. The issue of a fixed time limit for notification and remedial action shall be further assessed when the Commission examines the modalities for the introduction in the general legal framework of a general personal data breach notification, including the addressees of such notifications and the criteria for triggering the obligation to notify according to the EU Comprehensive Approach 2010. The obligation of a time-limit for notification of data breach shall also be contained in the future EU standard forms of “privacy information notices”.

3 Conclusion and Recommendation

It is important to strike the balance between data privacy rights protection and the free movement of data within member states in order to build users’ trust on the Internet without jeopardizing technological innovation and market development. The recent European Commission review on the EC Directive on Data Protection has paid attention to that.14 Statistical methods on Internet as a source of data gathering could provide statistical outcomes faster than the traditional paper-based questionnaire methods. Using Internet as a source of data gathering could also collect data that is difficult or even impossible to be gathered in the offline world. The implementation of Internet-based measurements could bring us great added value to improve products and services and allow us to promptly respond to the market development. From a legal perspective, the successful implementation of the statistical methods on Internet as a source of data gathering depends on the appropriate use of the exemption clause, the correct operation of informing users and requesting consent where necessary and the strict compliance of lawful data storage and data breach notification system. The building of automated data collection systems for statistical purposes has to comply with appropriate legislative and technological measures.

REFERENCES


Case C-444/02, Fixtures Marketing Ltd v. Organismos prognostikon agonon podosfairou AE - “OPAP”, n. 20, 25.

Case C-553/07, College van burgemeester en wethouders van Rotterdam v M.E.E. Rijkeboer, European Court of Justice (Judgment of 7 May 2009).


International Encyclopedia of Human Geography, MS number: 457.

