Groningen Journal of International Law, vol 2(2): Privacy in International Law

Open Source Data and Criminal Investigations: Anything You Publish Can and Will Be Used Against You

Els De Busser*

Keywords
DATA PROTECTION; INFORMATION EXCHANGE; OPEN SOURCE DATA; RIGHT TO BE FORGOTTEN

Abstract
Various misconceptions exist regarding open source data, what is meant by this term and how these data can legally be used. This contribution focuses on developing a comprehensive definition of the term and highlights the differences with similar – often confusing – concepts. The fact that open source data are publicly available does not mean that they can be used and processed in any way or for any purpose. As far as open sources contain personal data, the general data protection legislation (national as well as EU and Council of Europe legislation) is applicable. Several difficulties however arise, especially when different types of data are mixed.

This can happen in the context of a criminal investigation. The use of personal data for the purpose of prevention, investigation or prosecution of criminal offences is protected by more specific legal provisions to protect the secrecy of the investigation as well as the fundamental rights of the suspect and the victim(s). The fair trial rights of article 6 ECHR should be respected once a criminal charge has been made.

Open source data are vulnerable for abuse by any individual. Additionally, they are widely available and distributable when the internet is used. In several instances open source data have been used for the purpose of vigilantism (individuals taking law enforcement into their own hands). It is important to draw the line between a legal use of open source data, including the use of open source data for the purpose of a criminal investigation and the illegal use of open source data.

This contribution combines the elements of open source data, personal data and criminal investigations. Answers to the following research questions are sought:

- What are open source data?
- How to protect personal data included in open source data?
- How to use open source data in criminal investigations while respecting data protection legislation?

I. Introduction

Various misconceptions exist regarding open source data, what is meant by this term and how these data can legally be used. This contribution focuses on developing a comprehensive definition of the term and highlights the differences with concepts that seem similar and therefore are often confused. The fact that open source data are publicly available does not mean that they can be used and processed in any way or for any

* Head of Section European Criminal Law, Max Planck Institute for Foreign and International Criminal Law, Freiburg, Germany. The author would like to thank Tania Boulot, Nora Römling and Kamand Gharun for their assistance and feedback while preparing this paper.

purpose. As far as open sources contain or consist of personal data, general data protection legislation (national as well as European Union and Council of Europe legislation) is applicable. Several difficulties, however, arise related to the nature of the data and their use or processing.

Besides data protection laws, the use of personal data for the purpose of prevention, investigation or prosecution of criminal offences is also protected by more specific legal provisions to protect the secrecy of the investigation as well as the fundamental rights of the suspect and the victim(s). The fair trial rights of Article 6 European Convention of Human Rights should be respected once a criminal charge has been made. This goes for personal data that are open source as well as closed source data. More and more open source data are used by law enforcement and intelligence services, especially where social media is concerned. In 2012 LexisNexis® Risk Solutions surveyed 1,200 United States federal, state, and local law enforcement professionals concluding that four out of five use various social media networks to assist in investigations with Facebook and YouTube ranking among the most used platforms. This use concerned identifying people and locations; discovering criminal activity and locations; and gathering evidence. Of all respondents, 67% reported believing that social media helps solving crime more quickly.[1] Even though this survey was conducted in the United States, it shows the rising importance of social media as an investigative tool for law enforcement.

Open source data are vulnerable for abuse by any individual. Additionally, they are widely available and distributable when the Internet is used. In several instances open source data have been used for the purpose of vigilantism (individuals taking law enforcement into their own hands). It is important to draw the line between a legal use of open source data, including the use of open source data for the purpose of a criminal investigation and the illegal use of open source data. Lastly, since the Court of Justice of the European Union (CJEU) ruled on a landmark case against Google in May 2014, it is equally relevant to discuss here the catchphrase "the right to be forgotten", the fact that it does not exist and what this debate is really about.

Referring to the so-called Miranda rights in the title—the rights that should be read by US law enforcement officers when taking an individual into custody—is not meant to sound harsh or depressing. It is rather intended to create awareness for Internet and social media users indiscriminately, publicly posting personal data identifying themselves or others. The consequences of this recent trend are not always directly perceived, which makes it all the more difficult to control. Besides raising awareness, this contribution focuses on identifying the precise problem(s) rather than offering concrete solutions.

Combining the elements of open source data, personal data and criminal investigations, this paper intends to offer an answer to questions such as what are open source data; how can personal data included in open source data be protected; and how can open source data be used in criminal investigations while respecting data protection legislation? The legal instruments that are used to answer these questions are the relevant legal instruments adopted by the Council of Europe (CoE) and by the European Union (EU). These include the European Convention on Human Rights (ECHR), the Convention on the processing of personal data by automated means (Data Protection

[1] LexisNexis Risk Solutions, *Role of Social Media in Law Enforcement Significant and Growing*, available online at <lexisnexis.com/en-us/about-us/media/press-release.page?id=134262308548118 1#sthash.pbREo4je.dpuf> (accessed 30 July 2014).

Convention),[2] Resolutions 73(22) and 74(29), and Recommendation 87(15). For the EU the most relevant legal instruments include Directive 95/46/EC 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 (Directive 95/46/EC), [3] the Framework Decision 2008/977/JHA of 27 November 2008 on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters (Framework Decision 2008), and the legislative proposals that are being negotiated at the present time to reform both the Directive[4] and the Framework Decision.[5] Using these legal instruments does not mean that the geographical scope of this paper is limited to the EU. Rather, all Member States of the CoE are bound by the same data protection standards as well as countries that are not Member States of the CoE.

II. Defining Open Source Data

In order to define what open source data are, it is necessary to first explain what they are not. Open source data are not identical to personal data but can contain or consist of personal data. Traditionally, personal and non-personal data are distinguished based on the characteristic of identifying an individual or enable to identify an individual. Personal data enable one to "single out" a person. The fact whether personal data are open source or not is not part of the definition. On the contrary, the definition of personal data includes any information, which can be open source or closed source. Open source data in their turn can be personal or non-personal.

II.1. Personal Data

The concept of personal data is frequently confused with the right to a private life or privacy. Both concepts overlap, but only to a certain extent. They are certainly not identical. Where personal data are those data that identify or enable to identify an individual, the private life of a person consists of personal as well as of non-personal data. As one of the most difficult concepts to explain—not in the least because of its evolvement in line with technological advancements—the best definition is still the traditional definition introduced by Warren and Brandeis in 1890 describing the right to a

---

2 Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, 1981, ETS No. 108 available online at <conventions.coe.int/Treaty/en/Treaties/Html/108.htm> (accessed 27 September 2014) (CoE Data Protection Convention).

3 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, available online at <eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=EN> (accessed 1 2014) (Data Protection Directive 95/46/EC).

4 Proposal for a Regulation on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 25 January 2012, COM (2012)11 final, available online at <ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf> (accessed 16 November 2014).

5 Proposal for a Directive on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data, 25 January 2012, COM (2012), 10, available online at <eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0010:FIN:EN:PDF> (accessed 3 November 2014).

private life as the right to be let alone.[6] Exercising the right to be let alone and not tolerate interference from private or public persons such as the government, involves more than only personal data.[7] One should thus be careful not to confuse both concepts. Nonetheless, in the jurisprudence of the European Court of Human Rights (ECtHR) the right to a private life has been used to include rulings on personal data. After all, a genuine right to data protection is so far only included in the EU Charter on Fundamental Rights and Freedoms, not in the ECHR.

The data protection standards applicable in the EU and the CoE Member States originate from the CoE Data Protection Convention and two preceding Resolutions.[8] In the Convention and in Directive 95/46/EC, personal data are defined as any information relating to an identified or identifiable individual.[9] Public sector information is also covered by this definition.[10] An identifiable person is a physical[11] person who can be easily identified, meaning not by using very sophisticated[12] methods that should be judged considering technological evolutions.[13]

"Any information" refers to any type of information, objective as well as subjective statements concerning objects, events or persons. Opinions, assessments or conclusions about objects or persons establish subjective information. The format in which the information is held or its carrier is not relevant. Information in any structured or

---

6 Often incorrectly quoted as "to be *left* alone". See Warren, S. D. and Brandeis, L. D., "The right to privacy", *Harvard Law Review*, vol. 4, 1890, 193–220. See also Council of Europe, Parliamentary Assembly, *Recommendation 509(1968) on human rights and modern scientific and technological developments*, 1968, available online at <assembly.coe.int/Main.asp?link=/Documents/AdoptedText/ta68/EREC509.htm> (accessed 30 July 2014). This Recommendation started the legislative development of data protection rules and guidelines.

7 See also De Busser, E., *Data Protection in EU and US Criminal Cooperation*, Maklu Publishers, Antwerp-Apeldoorn, 2009, 48–52.

8 The Convention's Explanatory Report explained that the terms and definitions generally follow those used in Resolutions (73) 22 and (74) 29. Some modifications and additions have been made in view of recent national legislation and having regard to the special problems called forth by transfrontier data flows.

9 Article 2(a) CoE Data Protection Convention; Article 2(a) Data Protection Directive 95/46/EC.

10 See Article 29 Data Protection Working Party, *Opinion 3/99 on public sector information and the protection of personal data*, WP 20, 3 May 1999, available online at <ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/1999/wp20_en.pdf> (accessed 30 July 2014).

11 In accordance with Article 3, paragraph 2(b) of the Convention, Member States have the opportunity to declare the provisions of the Convention applicable to legal persons. Declarations in that sense have been submitted by Albania, Austria, Italy, Liechtenstein and Switzerland.

12 The focus on 'very sophisticated methods' as is mentioned in the Explanatory Report to the Data Protection Convention can lead to confusion. One might think that the higher the level of sophistication in the method used in order to identify a person, the less likely it is for the personal information that is detected this way to fall within the scope of the Convention. However—and rightfully pointed out by Bygrave—the higher the level of sophistication is, the easier it is for a person to identify an individual and consequently have access to personal data. Bygrave, L.A., *Data protection law. Approaching its rationale, logic and limits*, Kluwer law International, The Hague, 2002, 43–44.

13 Council of Europe, Explanatory Report to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, ETS No. 108, 1981, available online at <conventions.coe.int/treaty/en/Reports/Html/108.htm> (accessed 30 July 2014). See also the remarks made by the Court in: *Klass and others v. Germany,* App no 5029/71), para. 4, and ECtHR, 16 February 2000, *Amman v. Switzerland,* App no 27798/95), Section 56. See also Concurring Opinion of Judge Pettiti in ECtHR, 2 August 1984, *Malone v. UK,* App no 8691/79.).

unstructured form (numerical, photographical, acoustic or stored in a computer file)[14] is covered by the definition, taking into consideration future technological developments.

The phrase "related to" would logically mean that the information is about a specific person.[15] However, the EU's Article 29 Data Protection Working Party[16] in a 2005 opinion on the application of Directive 95/46/EC on the practice of RFID-tags[17] stated that this phrase 'refers to the identity, characteristics or behaviour of an individual or if such information is used to determine or influence the way in which that person is treated or evaluated'.[18] In view of recent discussions on gathering data on people's online surfing behaviour and personalised online advertising, it is significant that also such data can qualify as personal data. Nevertheless, data gathered by RFID tags or surfing behaviour would not be open source data.

In 2007 the Data Protection Working Party divided the meaning of "relating to" in two parts. On the one hand, certain content is required to make information relate to a person, meaning it should provide in the person's identity, his or her characteristics or behaviour. No purpose or consequence on behalf of the handler of the data is necessary. On the other hand, the use that is made of the information is divided into demonstrating either an element of purpose to assess, treat in a different way or influence a person's status or behaviour or an element of result or impact. The latter refers to the impact on a person's rights and interests or the different treatment of a person as a result, independent of the question whether this result was achieved.[19]

Singling out an individual from the general population or a smaller group of persons by the use of information, or even the possibility of distinguishing an individual from a multitude or a category of persons, constitutes the determining factor in personal data. Identifying someone's unique behaviour can already be sufficient, for example by means of the aforementioned RFID-tags.[20] A person can be isolated directly by using identifying elements such as a name, provided that the name is sufficiently distinctive. Whether more identifiers (address, phone number, physical characteristics, employment information, etc.) are needed, depends on the context. The same piece of information can be personal data in one context and not be sufficient as an identifier in a different setting.[21]

Recital 26 of the Directive 95/46/EC preamble includes a reasonable means-test with regard to the means used for identifying a person. The Data Protection Working Party

---

[14] Article 29 Data Protection Working Party, *Opinion 4/2007 on the concept of personal data*, WP 136, 20 June 2007, available online at <ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf> (accessed 30 July 2014), 7.

[15] *Id.*, 9.

[16] The name "Article 29 Data Protection Working Party" is derived from the Article 29 of Directive 95/46/EC that set up this working party. It publishes opinions on specific issues related to the application of the Directive.

[17] Radio Frequency Identification Technology stands for a microchip storing data on certain behaviour, for example purchasing behaviour, by the person carrying the tag, which is read by the controller of the tag.

[18] Article 29 Data Protection Working Party, *Working Document on data protection issues related to RFID technology*, WP 105, 19 January 2005, 8.

[19] Data Protection Working Party Opinion 4/2007, *supra* nt. 14, 10–11.

[20] Data Protection Working Party Opinion 4/2007, *supra* nt. 14, 13–14; see also above nt. 17.

[21] See CJEU 6 November 2003, *Lindqvist*, C-101/0101, para. 27. In this case the Court decided on data on an Internet page referring to person's names in conjunction with their phone number or information concerning their working conditions and hobbies, to be personal data within the meaning of Directive 95/46/EC.

added the criteria of cost[22] of conducting the identification, the intended purpose, the way the processing is structured, the advantage expected by the controller, the interests at stake for the individuals, the risk of organisational dysfunctions and technical failures.[23] All means that are likely reasonably used by the handler of the data to identify the person concerned should be considered by the judge deciding upon a case-by-case-basis. The phrase "likely reasonably" causes confusion as to its exact meaning, in particular by joining such element of probability with the element of difficulty.[24] The fact that the handler of the information would be capable of identifying a person does not necessarily mean that he will in fact put this into practice. However, this would not cause the data to lose their quality of personal data.[25]

During the negotiations on the reform of the data protection legal framework in the EU, the concept of singling out was added to the text of the preamble in the amendments made by the European Parliament.[26] This was not a new notion as the Data Protection Working Party already used it in its 2007 opinion on the concept of personal data.[27] Data that can lead to the singling out of a person from a group of persons thus needs to be so specific—depending on the size of the group[28]—that only one individual can be isolated from the rest of the group.

II.2. Open Source Data

Open source data or open data do not have an official definition that is laid down in any legal instrument. Many documents use the term without defining it, yet limited sources have included their own definition.[29] The common characteristic of the definitions lies in the information being publicly available. When data are closed off from the general public, they can clearly not be considered open source data. When a fee is required to obtain the data, can they still be considered open source? And does it include information on social media profiles that are not public but still open to thousands of users? Where do we draw the line?

---

[22] The Article 29 Data Protection Working Party explicitly mentions the costs as a criterion for concluding on the identification (even though it states it is not the only factor). In 1997 the Council of Europe no longer included costs as a reliable criterion due to developments in computer technology. See Council of Europe, Committee of Ministers, *Recommendation No. R(97)5 on the protection of medical data*, 13 February 1997, available online at <wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=564487&SecMode=1&DocId=560582&Usage=2> (accessed 30 July 2014).

[23] Data Protection Working Party Opinion 4/2007, *supra* nt. 14, 15.

[24] Bygrave, L. A., *Data protection law. Approaching its rationale, logic and limits*, Kluwer law International, The Hague, 2002, 44.

[25] *Ibid*.

[26] European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)), available online at <europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2014-0212+0+DOC+XML+V0//EN> (accessed 3 November 2014), Amendment 66.

[27] Data Protection Working Party Opinion 4/2007, *supra* nt. 14, 13.

[28] For example when data refers to a dark haired woman in her thirties living in New York, the group of people will be too large to identify this individual. When the data are more specific and refer to a dark haired woman in her thirties living in New York and teaching English literature at University X in Manhattan, New York City, this would single out a specific individual.

[29] See also Eijkman, Q. and Weggemans, D., "Open source intelligence and privacy dilemmas", *Security and Human Rights*, No. 4, 2012, 286-287.

Open source data is not a new concept as such, demonstrated by the references in guidelines and manuals for intelligence services. Nevertheless, the boom of social media and other sources on the Internet have given it a new dimension by flooding the pool of existing open source data. That does not mean that open source data need to be digital. Even if the majority of open source data today will be found in digital form, observations, photographs or paper publications may just as well be open and publicly available data. The CoE Convention on Cybercrime uses the term open source data, but only indirectly refers to it as publicly available data without giving a definition.[30] With due care not to confuse information and intelligence notions, it is still useful to examine the definitions used in the area of (criminal as well as military) intelligence because open source data are also for intelligence services a necessary source, possibly even a starting point.

The United Nations Office on Drugs and Crime (UNODC) describes open source data as information that is publicly available and adds that one of the main difficulties in working with this type of source is evaluation, as information available in the public domain can frequently be biased, inaccurate or sensationalised.[31] This definition is clearly accommodated towards criminal intelligence analysts and is much wider than information containing personal data. In its Open Source Intelligence Handbook, North Atlantic Treaty Organization (NATO) first separates open source intelligence from academic, business or journalistic research by highlighting that ''it represents the application of the proven process of national intelligence to a global diversity of sources, with the intent of producing tailored intelligence for the commander'.[32] The proven process of national intelligence logically refers to the analysing of information for military purposes. Nonetheless, NATO's discerning definitions of four types of information and intelligence are relevant in this discussion due to the elements of restriction of information for a specific person or group of persons on the one hand and the element of verification or accuracy on the other hand. According to NATO, open source information means that a form of processing has taken place from the raw open source data.[33] It refers to those data that can be put together, generally by an editorial process that provides some filtering and validation as well as presentation management. Open source information is thus generic information that is usually widely disseminated and includes newspapers, books, broadcast, and general daily reports. Open source intelligence refers to information that has been deliberately discovered, discriminated, distilled, and disseminated to a select audience in order to address a specific question. This type of information applies the proven process of intelligence to the broad diversity of open sources of information, and creates intelligence. A more advanced type of information is the validated open source intelligence. This is defined as information to which a very high degree of certainty can be attributed. It can be produced by an all-source intelligence professional, with access to classified intelligence sources. It can also

---

[30] Council of Europe, *Explanatory Report to the Convention on Cybercrime*, ETS No. 185, 8 November 2001, available online at <conventions.coe.int/Treaty/en/Reports/Html/185.htm> (accessed 3 November 2014).

[31] United Nations Office on Drugs and Crime, *Criminal Intelligence - Manual for Analysts*, 2011, available online at <unodc.org/documents/organized-crime/Law-Enforcement/Criminal_Intelligence_for_Analysts.pdf> (accessed 20 July 2014), 12.

[32] NATO, *Open Source Intelligence Handbook*, 2001, available online at <oss.net/dynamaster/file_archive/030201/ca5fb66734f540fbb4f8f6ef759b258c/NATO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf> (accessed 20 July 2014), 1–3.

[33] Open source data is defined as the raw print, broadcast, oral debriefing or other form of information from a primary source.

come from an assured open source to which no question can be raised concerning its validity.[34] Open source data and open source information are thus in the NATO definitions meant for a wider audience and have been subject to a lower degree of scrutiny, while open source intelligence and validated open source intelligence are rather conclusions drawn from the data and information, the degree of accuracy and reliability is higher and it is meant for a restricted audience.

Open source data can be authored and developed by any person. In some cases the author or producer is unknown and the reliability or accuracy cannot possibly be verified, for example fake profiles on social media. In other cases, such as journalism, the author is known and the information has a high degree of reliability and accuracy. Still this is considered open source data.[35] It is thus not relevant for the description of open source data whether its reliability and accuracy has been checked.

The size of the audience to whom the data are available brings up the question of payment. Can data that is only available on payment be considered open source or not? It would not be realistic to limit the definition of open source data to freely available data as technically one would have to consider the cost of Internet connections even when newspapers or social media have freely accessible websites.[36] However, open source data can also exist in the offline world. For example, an expensive book or report can be publicly available, but due to its price, it is limited in accessibility. For this reason the element of payment should not be included in the definition of open source data, rather the aspect of availability to a wide or restricted public is significant. A restricted public is not the general population but a group of people that is separated from the general population based on one or more filtering conditions such as their professional occupation, their paid or unpaid subscription to a newspaper or their friendship with a person on a social media profile. The latter brings up a particular question regarding the threshold that is required. When the account holder of a Facebook profile that is not public posts information, one would tend to label this information as closed source data. However, if this Facebook user has over 5,000 friends, can we still rightfully speak of closed source data? In addition, every one of these friends can share the information with his or her friends creating a snowball effect and an uncontainable distribution of the information. The same goes for a newspaper that has thousands of paying subscribers who can spread information further. A solution could be to interpret the term "restricted public" as referring to the ability to specify the recipients of the data and to limit the dissemination of the information. This interpretation results in any information that is posted on a Facebook profile allowing the friends of the account holder to share, should be labelled as open source data. This does not mean that any person can do anything he or she wants with the data, for two reasons. First, the fact that such data are open source does not mean that they are reliable or accurate. Second, open source data can contain personal data. If this is the case they are protected by data protection regulations.

Developing a definition of open source data that is not exclusively meant for the field of criminal and military intelligence, it is clearer to describe what open source data are not rather than to describe what is covered by the term. Based on the analysis above,

---

[34] Explanatory Report to the Convention on Cybercrime, *supra* nt. 30.

[35] See above, nt. 29.

[36] Even in the description of personal data in 1997, the Council of Europe did not consider cost a reliable criterion due to the developments in computer technology. Council of Europe Committee of Ministers, *Recommendation No. R(97)5 on the protection of medical data*, 13 February 1997, available online at <wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=564487&SecMode=1&DocId=560582&Usage=2> (accessed 30 July 2014).

open source data can be described as any information that is not restricted to a specified public and that is not necessarily reliable or accurate. Whether or not the information identifies or enables to identify an individual is not part of the definition since open source data can include both personal data and non-personal data.

III. Personal Data Protection

When open source data contain personal data, they are protected by the traditional data protection standards. These are laid down in binding legal instruments. The data protection legal instrument that has the widest geographical scope and is also the oldest international convention on this matter is the 1981 CoE Data Protection Convention. Ratified by forty six states, the Convention has introduced the basic principles to be complied with when personal data are processed. Even though its scope is limited to automatic processing, many states have widened the scope of their implementing legislation to also include non-automatic data processing. In this part, the data protection standards are applied to the central theme of open source data including the particular challenges that this type of data can raise for data protection.

III.1. Data Protection Standards

As the basic binding[37] legal instrument, the Data Protection Convention sets out the five minimum requirements personal data should fulfil. Article 5 of the Convention was based on the text of two older CoE Resolutions[38] and distinguishes two groups of standards: quality standards for personal data on the one hand, and quality standards for the processing of personal data on the other hand. Both are divided into more detailed principles that will be dealt with here in line with the two fundamental legal standards presented by the CoE.[39] Besides the data subject giving his or her consent, derogations are allowed but only in accordance with Article 9 that is in turn based on the provisions of Article 8 ECHR. It should be pointed out that for the EU Member States, the standards of the Convention have been implemented and further specified in Directive 95/46/EC for commercial matters and in Framework Decision 2008 for criminal matters.

---

[37] As non-binding instruments, the OECD Guidelines Governing the Protection of Privacy and Trans-border Flows of Personal Data, C(80)58/Final, 23 September 1980 (OECD Guidelines) and the United Nations Guidelines concerning Computerized Personal Data Files, General Assembly, 14 December 1990, encompass the same basic principles, leaving room for national legislators to implement data protection rules based on these guidelines (UN Guidelines).

[38] Resolution (73) 22 of the Committee of Ministers of 26 September 1973 on the Protection of the Privacy of Individuals vis-à-vis Electronic Data Banks in the Private Sector, available online at <wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=589402&SecMode=1&DocId=646994&Usage=2> (accessed 21 September 2014); Resolution (74)29 of the Committee of Ministers of 20 September 1974 on the Protection of the Privacy of Individuals *vis-à-vis* Electronic Data Banks in the Public Sector, available online at <wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=590512&SecMode=1&DocId=649498&Usage=2> (accessed 21 September 2014). .

[39] Explanatory Report of the Council of Europe Convention for the Protection of Individuals with Regards to the Automatic Processing of Personal Data, ETS no. 108, Section 40, available online at <conventions.coe.int/Treaty/EN/Reports/Html/108.htm> (accessed 27 September 2014).

III.1.1. Quality Standards for Personal Data

III.1.1.1. Accuracy and Reliability

Ensuring the accuracy of personal data that are processed and updating them whenever necessary is the first standard of data protection. In other words, this standard assures the correspondence of the data to the reality they refer to, such as a person's name and address, employment status, health data, etc. The Data Protection Convention provides the data subject (the person who is identified by the data) with the right to have data corrected or erased if they do not comply with this standard. This implies notification to the data subject of the fact data were gathered and the purpose thereof, unless the individual already has this information or unless other exceptions apply such as the prevailing interests of an ongoing investigation.

As additional protection, Directive 95/46/EC assigns the data controller as the responsible party for ensuring the accuracy of the data as well as updates.[40] The data controller is the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data.[41] The frequency of updates is not regulated. Although United Nations (UN) Guidelines recommend updates to be held regularly or when the data contained in a file are used,[42] the Convention and the Directive limit updating data to "where necessary".[43] The Organisation for Economic Co-operation and Development Guidelines mention that data quality standards are not intended to be more far-reaching than is necessary for the purposes for which the data are used.[44] For example, data processed for historical or statistical purposes do not necessarily need updating.

In accordance with the definition of open source data developed in this contribution, they are not necessarily accurate or reliable. When open source data contain data that identify or enable to identify an individual however, they should also be updated or corrected when necessary. Considering the possibly wide and uncontainable distribution of open source data, updating and correcting can only be done at the source, whether this is an update on a social media page or a newspaper publishing an erratum. Logically, the data subject can enforce his or her right to correct or erase false personal data that are open source.

---

[40] Article 6, Section 1(d) and Section 2, 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, available online at <eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML> (accessed 27 September 2014).

[41] *Id.*, Article 2(d).

[42] Article A.2, GA Resolution 45/95 (68th plenary meeting) A/RES/45/95 14, December 1990.

[43] Article 5(d) CoE Data Protection Convention; Article 6(1)(d) Data Protection Directive 95/46/EC.

[44] Article 53, OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, available online at <oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm> (accessed 27 September 2014).

III.1.1.2. Adequate, Relevant and Proportionate Personal Data

Personal data should be adequate, relevant and not excessive in relation to the purposes they are gathered and processed for. The Data Protection Convention[45] and the Directive[46] provide for a qualitative and a quantitative condition;[47] no personal data should be collected and stored in view of a potential future use, without having an exact view on the purpose it would be used for.[48] This was one of the reasons why on 8 April 2014 the CJEU annulled the controversial Directive 2006/24/EC (Data Retention Directive) obliging telecommunication providers to store personal data for periods of time up to two years in case they may be needed in a future criminal investigation or prosecution.[49]

A qualitative connection should exist between the personal data and the purpose. If there is no direct nexus—for example the same result can be achieved by other less intrusive means[50]—the data are not adequate or relevant in relation to the purpose. No personal data can be processed for undefined purposes,[51] a specified purpose should be provided as well as a direct link between purpose and data. Respecting the proportionality rule means that the data controller should determine and distinguish the minimum amount of personal data needed in order to successfully accomplish a specific purpose and limit its processing to these data.[52] Blanket data collection or fishing expeditions[53] are not in line with the data protection standards.[54]

The purpose for the processing of personal data included in open source data could be journalistic purposes or academic research. Determining whether personal data are in such cases adequate, relevant and not excessive can be challenging. The recent case before the Court of Justice on the debated and often misunderstood catchphrase 'the right to be forgotten' demonstrates how difficult the adequacy and relevance of personal data in open source situations can be. For this reason a separate part of this contribution is dedicated to an analysis of the Court of Justice ruling of spring 2014.

---

45 Article 5(c), CoE Data Protection Convention.

46 Article 6, Section 1(c), Data Protection Directive 95/46/EC.

47 Also the non-binding UN Guidelines, Section A.3 and OECD Guidelines, para. 53 provide in this rule.

48 This should be distinguished from the case in which data are gathered and kept for a particular foreseeable emergency which may never occur, for example, where an employer holds details of blood groups of employees engaged in hazardous occupations. Information Commissioner, Data Protection Act 1998, Legal Guidance, 1998, 37.

49 European Court of Justice (ECJ), 13 May 2014, *Google Spain SL, Google Inc. v. Agencia Española de Protección de Datos (AEPD), Mario Costeja González*, C-131/12.

50 For example the Belgian Privacy Commission did not authorise the National Organization for the Identification and Registration of Dogs access to the national register of inhabitants based on the lack of proportionality, since in case a dog owner should be contacted, local or federal police authorities can be involved in order to find the owner's contact data. Belgian Privacy Commission, Advise no. 38/2001, 8 October 2001.

51 *Open Source Intelligence Handbook, supra* nt. 32, 41.

52 Information Commissioner, Data Protection Act 1998, Legal Guidance, 1998, 36.

53 Fishing expeditions refer to random and untargeted searches in a large collection of data in an attempt to find relevant information.

54 See Committee of Ministers, Resolution (1973) 22, Article 21, in which adopting a rule that would 'halt unbridled hoarding of data' is recommended.

III.1.1.3. No Such Thing as a Right to Be Forgotten

The so-called right to be forgotten became a catchphrase in 2012 with the launch of the EU data protection legal framework reform. The term however is fundamentally incorrect. There is no such thing as a right to be forgotten and there never will be as long as the individual human memory and the collective memory cannot be physically tampered with.[55] What exists in accordance with applicable data protection rules is a right to have personal data corrected, updated or deleted when necessary. This is nothing new as this right has been in existence since the aforementioned data quality standards were laid down in the 1981 Data Protection Convention.

On 13 May 2014 the Court of Justice ruled on a preliminary question brought before it by the Spanish Audiencia Nacional. The Court decided that the world's most popular search engine Google is responsible for removing links to personal data that are no longer relevant to the purpose they were processed for. Data subject in the case is Costeja González, a Spanish citizen who had social security debts in the late nineties. The recovery of these debts led to a real-estate auction that was in accordance with an order by the Ministry of Labour and Social Affairs announced in newspaper La Vanguardia with the purpose to give the auction maximum publicity and attract as many bidders as possible. In 1998, not every newspaper had an online version as is the case today. Also, La Vanguardia has in the meantime made its publication and archive available online, including the announcement mentioning Costeja González. When he realised the open source availability of this information after a Google search on his name, he submitted complaints with the Spanish data protection authority against the newspaper and against Google. According to the data protection authority, the publication by La Vanguardia was legally justified because of the order by the Ministry of Labour and Social Affairs. As a result, this complaint was rejected. The complaint against Google and the request that Google remove the links to the published personal data was brought before a national judge, who sent a request for a preliminary ruling to the Court of Justice. Contrary to what the Advocate General to the Court of Justice concluded, the Court first of all considered Google a data controller for the activity consisting in finding information published or placed on the Internet by third parties, indexing it automatically, storing it temporarily and, finally, making it available to Internet users according to a particular order of preference. Secondly, the Court considered the search engine also responsible for removing the links making the information concerning Costeja González available on the Internet.

The personal data as such are not contested in this case as they are not incorrect. Nevertheless, according to Costeja González an announcement for a real-estate auction held in 1998, published for the purpose of ensuring a higher amount of bidders has lost all relevance two decades later. Because personal data should be relevant for the purpose they were processed for, and should not be stored in a database longer than is necessary for that purpose, thus far the Court of Justice's ruling is acceptable. Holding Google responsible for the fact that this announcement is still available today however is focusing on the wrong target. Google only makes information that already exists searchable and creates an index of search results; it did not create the data, nor was Google the source of the information as such. Requiring Google to remove the links is not the correct issue to

[55] See also De Hert, P. and Papakonstantinou, V., "How the European Google Decision May Have Nothing To Do With a Right to Be Forgotten", *Privacy Perspectives*, International Association of Privacy Professionals, 19 June 2014, available online at <privacyassociation.org/news/a/how-the-european-google-decision-may-have-nothing-to-do-with-a-right-to-be> (accessed 27 September 2014).

address from a data protection point of view. Even after removal of the link, the information is still available on the La Vanguardia website. Many countries have legal provisions on the publication of certain announcements or judgments that entered into force before the development of the Internet. In the meantime, the concept of publication has evolved. It now also includes online publication, making newspapers available on a wider scale and for a possibly indefinite period of time. The real issue here is the fact that the personal data are still present on the La Vanguardia website two decades after it was legal and necessary to publish it for a particular real-estate auction. The correct target is therefore the national Spanish law and its data retention provisions, not Google. The Court of Justice was logically limited by the scope of the request for a preliminary ruling but could have at least ruled that Google was not responsible for removing the links in question.

Besides the described data protection issue in this case, it is also dangerous to put a private company in a position to decide whether or not the link to certain information is relevant. A search engine's interests are of a commercial nature and do not encompass the rights of the data subject. This is a task for a data protection authority or a judge, not a private company. Moreover, Google is now overwhelmed with over 90,000 requests for the removal of links since the Court of Justice ruling.[56] The company has even reinstalled links to newspaper articles from the Guardian after the British newspaper protested their removal.[57] This shows the difficulties for a private company to be in such a position and the inevitable tension with the freedom of information.

III.1.2. Quality Standards for the Processing of Personal Data

Personal data should be obtained and processed fairly and lawfully. This data protection rule means that gathering personal data, and as a possible result infringing upon a person's right to a private life, can only be done when this encompasses lawfully derogating from Article 8 ECHR. In other words, the gathering of personal data must be laid down in law, it should have a legitimate aim and it should be necessary in the interest of protecting state security, public safety, the monetary interests of the state or the suppression of criminal offences or in the interests of protecting the data subject or the rights and freedoms of others. Two principles complete the quality standards for data processing: the purpose limitation principle and the data retention principle.

---

[56] Schechner, S., "Google Grants Majority of 'Right to Be Forgotten' Requests", *The Wall Street Journal*, 24 July 2014, available online at <online.wsj.com/Articles/google-grants-more-than-half-of-right-to-be-forgotten-requests-processed-so-far-1406223241?mod=yahoo_hs> (accessed 27 September 2014).

[57] Lee, D., "Google reinstates 'forgotten' links after pressure", *BBC News*, 4 July 2014, available online at <bbc.com/news/technology-28157607> (accessed 27 September 2014).

III.1.2.1. Purpose Limitation

In the aforementioned 1973 Resolution of the CoE, purpose limitation first made its introduction when the need was felt to control the use made of information stored in electronic databanks.[58] The purpose limitation principle means that personal data should be stored for specified and legitimate purposes only and should not be used in a way that is incompatible with those purposes. In other words, the purpose for which personal data may be processed is either the original purpose they were collected for or a purpose that is compatible therewith. What exactly constitutes a compatible purpose is not defined by the Data Protection Convention or its explanatory report. It was not until 2013 that the EU Data Protection Working Party published an opinion on what should be understood by the term "compatible purpose".[59] Rather than offering a strict definition of compatibility, which would be too stringent, the Working Party listed key indicators to be considered when assessing compatibility. These are the relationship between the purposes for which the data have been collected and the purposes of further processing, the context in which the data have been collected and the reasonable expectations of the data subjects as to their further use, the nature of the data and the impact of the further processing on the data subjects and the safeguards applied by the controller to ensure fair processing and to prevent any undue impact on the data subjects. Since this is an opinion, it is not legally binding. Nevertheless, it offers guidance to data controllers, data protection authorities or judges deciding on the matter.

In this respect, Article 8, paragraph 2 ECHR should be referred to, since every interference with the right to privacy should be legal and necessary in the interests of a legitimate aim. Even with the difference between the right to privacy and the processing of personal data described above, derogating from both is governed by the same restrictions as Article 9 of the Data Protection Convention is modelled on the provisions of Article 8, paragraph 2 ECHR. Lawfully derogating from the data protection standards will be discussed in the next sub-section of this paper.

III.1.2.2. Purpose Limitation and Open Source Data

The significance of purpose limitation for open source data lies in the fact that the public availability of open source data raises the risk of processing for incompatible purposes. Any personal data that can be drawn from an open source, such as statements or pictures, posted on a public social media profile can be misused for other purposes.[60]

---

[58] Committee of Ministers, Resolution (73)22 on the Protection of the Privacy of Individuals *vis-à-vis* Electronic Data Banks in the Private Sector, 26 September 1973. In addition a similar Resolution was adopted on the protection of privacy of individuals *vis-à-vis* electronic data banks in the public sector, Committee of Ministers, Resolution (74)29 on the Protection of the Privacy of Individuals *vis-à-vis* Electronic Data Banks in the Public Sector, 20 September 1974. The Explanatory Report to Resolution (73)22 demonstrates the fear of data abuse: 'There is a certain risk that the user of a data bank, in order to pay off the cost of storing data, might try to find new applications for which the data in his possession could be used. If such applications were to go beyond the original purposes for which the information had been compiled, a violation of the right of the persons concerned to privacy might ensue.'

[59] Article 29 Data Protection Working Party, Opinion 03/2013 on purpose limitation, 2 April 2013, 23–27, available online at <ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf> (accessed 27 September 2014).

[60] This does not mean that a data protection infringement is the only possibly intrusion, (criminal) offences such as defamation and slander could occur or intellectual property rights could be disrespected. These, however, reach outside the scope of this contribution.

The first question is what the original purpose of the open source data is. In some cases such as academic research or journalism this can be clear, in the case of social media the purpose for publishing personal data can range from holiday pictures to wedding announcements, informing people of a new phone number or simply chatting. The bottom line is communication. Perhaps one could even consider social media as a purpose in itself combining the communication element with the element of spreading information on oneself to a limited group of persons or to the general public.[61] An example of misusing open source data could be the public posting of a birth announcement—including the address of the new parents—on a social media page that is followed by an insurance company sending the parents folders for life insurance.

This seems similar to behavioural advertising but the difference is that the latter uses cookies or similar devices installing software on Internet users' computers to track their surfing behaviour, enabling them to show users personalised ads on specific webpages. The EU Data Protection Working Party has argued that the use of such identifiers enabling the creation of very detailed user profiles can in most cases be considered personal data processing, so users' prior consent for installing cookies is required.[62] This does not concern open source data since the surfing behaviour can only be tracked by specific software that is connected to companies' websites; thus the data that are gathered are restricted to a specified public.

When personal data on social media are publicly available, often the perception is that these may be used for any other purpose by anyone. Nonetheless, traditional data protection laws still apply and besides the described compatible purposes, such personal data, may only be used when the legality and necessity requirements are fulfilled. A typical example is a criminal investigation. The riots in several London neighbourhoods in 2011 led not only to the arrests of those inciting the looters on Facebook and Twitter, but also those who had unwisely posted pictures of themselves on social media with stolen goods. In the next part of this contribution, the use of open source data for criminal investigations and prosecutions will be discussed further.

III.1.2.3. Retention of Personal Data

Even if personal data are adequate, relevant and not excessive at the moment of their collection, after a certain amount of time these data could be no longer adequate and relevant in relation to the purpose they were gathered for. This was the case in the recent ruling by the Court of Justice against Google (see above).

The longer personal data are stored for, the higher the risk of intentional or unintentional misuse becomes.[63] The data retention principle specifies that personal data can be saved in databases for as long as is required for the purpose they are stored for. After this period of time has passed, the data can still be retained but need to be separated from the identifying factor, removing the quality of personal data. This separation does not need to be permanent, it is sufficient that the identification of the person concerned cannot be done easily.[64]

---

[61] Oxford Dictionary defines social media as websites and applications that enable users to create and share content or to participate in social networking.

[62] Article 29 Data Protection Working Party, Opinion 2/2010 on online behavioural advertising, 22 June 2010, 9, available online at <ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp171_en.pdf>.

[63] Committee of Ministers, Resolution (73)22 on the Protection of the Privacy of Individuals *vis-à-vis* Electronic Data Banks in the Private Sector, 26 September 1973, paras. 23–25.

[64] *Open Source Intelligence Handbook, supra* nt. 32, 42.

Derogating from the data retention principle is lawful under the same conditions as explained above. In other words, personal data can be stored for longer than necessary, but this must be laid down in law and it needs to be necessary in the interests of protecting state security, public safety, the monetary interests of the state or the suppression of criminal offences or in the interest of protecting the data subject or the rights and freedoms of others. Before declaring the Data Retention Directive invalid, the Court of Justice confirmed that the fight against serious crime is indeed of the utmost importance in order to ensure public security. However, according to the Court such an objective of general interest cannot in itself justify a retention measure such as that established by the contested Directive being considered to be necessary for the purpose of that fight. In addition, the Court criticised the text of the Directive since it covers, in a generalised manner, all persons and all means of electronic communication as well as all traffic data without any differentiation, limitation or exception being made in the light of the objective of fighting against serious crime. The Directive applies even to persons for whom there is no evidence capable of suggesting that their conduct might have a direct or remote link with serious crime. It does not provide for any exception, meaning that it applies even to persons whose communications are subject, according to rules of national law, to the obligation of professional secrecy. Therefore, the Directive does not limit the processing of personal data to what is strictly necessary.[65]

III.1.2.4. Data Retention and Open Source Data

A significant issue with open source data and data retention is the lack of control. The availability of the data gives them a perceived outlaw-status. Without the data subject's knowledge, data identifying him or her can be copied and stored on servers, computers or portable devices where they may remain stored for very long periods of time and may or may not impact the data subject's life at a much later stage, as was the case in the aforementioned judgment against Google. Similar issues rise with public pictures or statements on social media that can be easily found online by potential employers negatively influencing their image of the data subject. For the reasons set out above, search engine operators should not be made responsible for providing links to open source data that are online.

In case a data subject would want to file a complaint against such retention and misuse of personal data, it is thus not the search engine but the website keeping the personal data in their databases that should be the target.

III.2. Derogating from Data Protection Standards

The lawful ways of derogating from the data protection standards have been briefly touched upon above. Not being able to derogate from these standards would hinder many forms of data processing that have legitimate aims and are necessary for the functioning of a democratic society, such as the prevention, investigation and prosecution of criminal offences. When open source data that contain or consist of personal data are processed outside the scope of the data protection standards, the processing should fulfil the requirements of legality and necessity.

Typically it is the necessity requirement that causes most difficulties in practice. The requirement of necessity was introduced in order not to give a state too much leeway and to identify a pressing social need. Still it encompasses a range of interests—fundamental

---

[65] ECJ, 8 April 2014, *Digital Rights Ireland and Seitlinger and Others*, C 293/12 and C 594/12, 51–58.

values in a democratic society—that could make derogating from the standards necessary.[66] The protection of state security refers to internal and external threats, making it legal to violate privacy rights—for example by conducting a telephone tap—in the context of an investigation against an attack on state institutions but also the gathering of intelligence. This derogation could therefore be used for allowing the use made of personal data by security services provided there is a nexus with a specific investigation. The monetary interests of the state refer to tax collection requirements and exchange control.[67] It does not entirely cover the economic well-being of the state, which was the wording used in Article 8, paragraph 2 ECHR. It covers however a specific part of it, namely all means of financing a state's policies. It is important—especially for the next part of this contribution—which the suppression of criminal offences does not require a criminal charge has been made against the individual involved. Where Article 8, paragraph 2 ECHR provides an exception for the prevention of disorder or crime, it encompasses a wider range of acts than merely investigation and prosecution of a criminal offence.

In its jurisprudence the European Court of Human Rights (ECtHR) has added three conditions: infringements of the right to a private life should also be precise, foreseeable and proportionate.[68] This means that every time an individual's right to a private life is restricted, the restriction should be counterbalanced by the assurance that it is legal and necessary for fulfilling a legitimate aim. Besides the fact that the legal provisions describing the allowed infringement should be precise enough, the individual should be able to predict from the relevant law in which cases his or her personal data could be collected and processed and these provisions should be precise and foreseeable in order for the individual to regulate his or her conduct accordingly.

Derogating from the right to a private life by processing personal data needs to be proportionate to the legitimate aim that is pursued. Proportionality is thus a requirement for the data itself as well as for the processing of the data. On the one hand, the personal data gathered by means of infringing upon an individual's privacy should not be excessive in quantity in relation to the objective to be served, for example the annulment of the Data Retention Directive was besides the potential use also based on the massive and undiscriminating retention of data. On the other hand, regardless the amount of data gathered, in cases where the same result could have been accomplished with actions that are less privacy-intrusive the proportionality requirement is not fulfilled.

The foreseeability aspect relates to the clarity of the legal provisions on processing of personal data as exceptions to the right to a private life. National data protection laws should be sufficiently clear in defining what constitutes a compatible purpose due to the interference with an individual's private life that the use of personal data entails. It is, however, not enough to simply provide in sufficiently clear laws. The EU Data Protection Working Party stated that in practice, laws should not only mention the final objectives of the legislative measure and designate the controller of the processing. They should also specifically describe the objectives of the relevant data processing, the

---

66 CoE Data Protection Convention, *supra* nt. 2, paras. 55–56.

67 *Id.*, para. 57.

68 European Data Protection Supervisor, Third Opinion 27 April 2007 on the Proposal for a Council Framework Decision on the Protection of Personal Data processed in the Framework of Police and Judicial Cooperation in Criminal Matters, *O.J.* C139, 23 June 2007, 5, available online at <secure.edps.europa.eu/EDPSWEB/webdav/shared/Documents/Consultation/Opinions/2007/07-04-27_3dpillar_3_EN.pdf> (accessed 27 September 2014); ECtHR, 2 August 1984, *Malone v. UK*, 8691/79, paras. 67–68; and ECtHR, 4 May 2000, *Rotaru v. Romania*, 28341/95, para. 55.

categories of personal data to be processed, the specific purposes and means of processing, the categories of persons authorised to process the data, the procedure to be followed for the processing, and the safeguards against any arbitrary interference by public authorities.[69]

Derogating from the rule of purpose limitation or from the data retention principle can only be foreseeable if it is formulated with sufficient precision to enable any individual – if need be with appropriate advice – to regulate his or her conduct. The individual should be able to predict the consequences of certain behaviour to a reasonable degree. However, the consequences should not be foreseeable with absolute certainty.[70] This requirement implies a responsibility on behalf of a state's legislator to design clear-cut and transparent provisions when enacting measures that interfere with individuals' right to a private life. In the judgment annulling the Data Retention Directive the Court of Justice criticised the EU legislator for not limiting data retention to what is strictly necessary. On EU level objective criteria should have been formulated, according to which the national legislators could limit the periods of data retention as well as the access rights to the databases.

IV. Evidence in Criminal Investigations

Open source data can and will be often used as evidence in criminal investigations. Based on their impact on the human rights of the individual(s) concerned—suspect, victims, witnesses, etc.—and on the society or community in which a criminal offence has been committed, criminal investigations and prosecutions are regulated by a special set of rights. The information that is used to investigate the offence and to establish the truth will also contain personal data, whose processing is regulated by the data protection standards discussed above. Open source data can equally be included in criminal investigations and prosecutions, triggering separate issues. In this section, these issues are identified after introducing the correct terminology and the rights to be considered.

IV.1. Information, Intelligence and Evidence

Before engaging in a discussion on investigations into criminal offences and the evidence used in criminal proceedings, it is important to understand the difference between the terms "information", "intelligence" and "evidence". Similar to NATO's explanation (see above), the UNODC explained the relevant terminology and stated that information is raw data of any type, whilst intelligence is data that has been worked on, given added value or significance. Information is evaluated through a process of considering it with regard to its context through its source and reliability.[71] This could, for example, include the combining of information with other information, the "connecting the dots" process.

Obviously information can consist of open source data. By interpreting open source data and giving them meaning, intelligence can be obtained. This is not yet evidence. Evidence is information and intelligence that is used to establish proof of one of more criminal offences. Which evidence is admissible and how evidence can be presented is

---

[69] See above nt. 22, 38.

[70] ECtHR, 16 February 2000, *Amman v. Switzerland*, 27798/95, para. 56; ECtHR, 26 April 1979, *The Sunday Times v. United Kingdom*, 6538/74, para. 49.

[71] United Nations Office on Drugs and Crime, *Criminal Intelligence – Manual for Analysts*, 2011, 1-2, available online at <unodc.org/documents/organized-crime/Law-Enforcement/Criminal_Intelligence_for_Analysts.pdf> (accessed 20 July 2014).

regulated by national laws. In principle there is no objection for open source data not to become evidence in criminal proceedings when they are relevant for the case and they are admissible as evidence. However, when open source data consist of personal data, then data protection standards should be complied with. Stating that police can use the information on a public social network profile without any restriction is thus incorrect.[72]

Besides the described data protection standards, another set of rights should be respected once a criminal charge is made: the so-called fair trial rights of Article 6 ECHR. In this context it is relevant to highlight the relationship between Article 6 and Article 8 ECHR. The latter describes the right to a private life, which is not identical to the right to data protection. At the present time only the EU has adopted a genuine right to data protection, in the Charter on Fundamental Rights and Freedoms. In the jurisprudence of the ECtHR however, the right to private life has been used to protect personal data as well. Therefore it is relevant to include the tension between Article 6 and Article 8 ECHR in this discussion.

IV.2. Fair Trial Rights

Article 6 ECHR is often referred to as the fair trial right, since it encompasses inter alia the requirement of an independent and impartial tribunal; the presumption of innocence and the right to a confrontation of witnesses. These rights should protect the defendant from arbitrariness or prejudgment in the course of the proceedings. Article 6 is applicable in civil as well as in criminal proceedings. However, when criminal proceedings are concerned, it is only applicable after a criminal charge has been made. In *Deweer v. Belgium*, the ECtHR determined this moment by means of the official notification of the allegation that the individual concerned has committed a criminal offence or an implication thereof has been given.[73] Whether or not a charge was criminal—and not administrative—was interpreted in further case-law. For a criminal charge it is necessary that the relevant national provisions belong to the criminal law of a state, disciplinary law or both, and when the nature of the offence and the severity of the penalty are considered to be criminal.[74]

Gathering information and intelligence is for a large part done before a criminal charge is made; usually it is needed in order to make a criminal charge. This would mean that the evidence derived from this information and intelligence would fall outside the scope of Article 6. With regard to the proactive use of special investigative techniques to collect information, the CoE has adopted specific recommendations.[75] Special

---

[72] Voigt, S., Jansen, N. and Hinz, O., "Law Enforcement 2.0 – The Potential and the (Legal) Restrictions of Facebook Data for Police Tracing and Investigation", *European Conference on Information Systems 2013 Completed Research,* 2013, available online at <staff.science.uu.nl/~Vlaan107/ecis/files/ECIS2013-0141-paper.pdf> (accessed on 25 July 2014).

[73] ECtHR, 27 February 1980, *Deweer v. Belgium*, 6903/75, para. 46.

[74] ECtHR, 8 June 1976, *Engel and others v. the Netherlands*, 5100/71, 5101/71, 5102/71, 5354/72, 5370/72, para. 82; ECtHR, 21 February 1984, *Öztürk v. Germany*, 8544/79, paras. 55–56.

[75] Council of Europe, *Guidelines on Human Rights and the Fight against Terrorism*, 11 July 2002, available online at <umn.edu/humanrts/instree/HR%20and%20the%20fight%20against%20terrorism.pdf> (accessed 1 October 2014); Council of Europe, *Recommendation Rec(2005)9 to member states on the protection of witnesses and collaborators of justice*, 20 April 2005, available online at <wcd.coe.int/ViewDoc.jsp?id=849237&Site=COE> (accessed 1 October 2014); Council of Europe, Recommendation Rec(2005)10 to Member States on special investigation techniques in relation to serious crimes including acts of terrorism, 20 April 2005, available online at <wcd.coe.int/ViewDoc.jsp?id=849269> (accessed 1 October 2014).

investigative techniques will not be needed when open source data are concerned. The question remains whether open source data that consist of personal data and have been collected before a criminal charge was made, can be used as evidence in criminal proceedings.

IV.3. Gap between Article 6 and Article 8 ECHR

Article 8 ECHR prohibits unnecessary interference with an individual's private life. It can be derogated from, provided that this is laid down in clear-cut and accessible legislation and provided it is necessary in the interests of preventing disorder or crime. Accurate information should be provided to the competent authorities that violation of a person's right to a private life is in fact genuinely preventing disorder or crime.[76] When it is clear to public authorities that there is little or no risk of disorder or crime occurring, they should refrain from interfering in a person's private life.[77]

Even though it was pointed out before that in the ECtHR jurisprudence Article 8 ECHR is used to protect personal data, it can still only serve as a basic rule and not as a detailed set of provisions for protecting personal data that are gathered for the purpose of prevention, investigation, prosecution and punishment of criminal offences. Article 6 ECHR in its turn protects the individual against whom a criminal charge was made but does not foresee in specific rights protecting the individual's private life or personal data.

The ECtHR has ruled more than once on the effect of a violation of Article 8 on the trial. In *Schenk v. Switzerland* and *Teixeira de Castro v. Portugal*, the Court considered it not necessary to discuss Article 8 after deciding on Article 6. In the first case no breach of Article 6 was detected due to the disputed recording of a private telephone conversation not being the only evidence.[78] In the Teixeira de Castro case the use of evidence as a result of incitement by undercover agents meant a clear breach of the right to a fair trial, so the Court did not see a need to consider the complaint on a breach of Article 8 separately.[79] In the Khan case, however, the Court took a stand on the relationship between Article 6 and Article 8. It ruled that a fair trial had been provided to the applicant who received due opportunities for challenging the evidence, even after confirming a breach of Article 8 based on the use of unlawfully installed listening devices.[80] With this judgment the Court cut the link between Article 6 and 8. The ruling is inspired by the established ECtHR case law stating that the right to a fair trial is based on all circumstances of the case. The proceedings as a whole, including appeal and cassation, should be part of the assessment whether a fair trial has taken place or not.[81] Rules on the admissibility of evidence as such are not within the ECtHR's competence. However, the Court concluded that, as long as the defendant has been given the

---

[76] ECtHR, 6 June 2006, *Segerstedt-Wiberg and others v. Sweden*, 62332/00, paras. 89 and 92. See also ECtHR, 6 September 1978, *Klass and others v. Germany*, 5029/71, para. 48; ECtHR, 25 December 2011, *P.G. and J.H. v. The United Kingdom*, 44787/98, paras. 50–51; ECtHR, 28 April 2003, *Peck v. the United Kingdom*, 44647/98, para. 67.

[77] For example, police officers should refrain from entering a person's private home in order to prevent crime or disorder when this is highly unlikely to occur due to the absence of the person who was considered to potentially cause a breach of the peace. See in that regard for example ECtHR, 23 September 1998, *McLeod v. United Kingdom*, 72/1997/856/1065, paras. 56–57.

[78] ECtHR, 12 July 1988, *Schenk v. Switzerland*, 10862/84, paras. 49 and 53.

[79] ECtHR, 9 June 1998, *Teixeira de Castro v. Portugal*, 44/1997/828/1034, paras. 39 and 43.

[80] ECtHR, 4 October 2000, *Khan v. United Kingdom*, 35394/97, paras. 38–40.

[81] ECtHR, 16 December 1992, *Edwards v. United Kingdom*, 13071/87, para. 39; ECtHR, 26 October 1984, *De Cubber v. Belgium*, 9186/80, para. 30.

opportunity to challenge the evidence brought against him and the evidence is reliable and not gathered by means of entrapment or inducement, encroaching on the right to a private life can still produce admissible evidence.[82]

Taking all circumstances of the case into account, three considerations should be made. First, the evidence resulting from the breach of privacy should not be the only evidence in the case. In practice, no prosecutor would take the risk basing a whole case on such evidence, especially if this evidence would be open source data. Open source data are not necessarily reliable or accurate and would therefore have to be accompanied by other evidence. Second, the nature of the violation of the right to a private life should be considered. Evidence resulting from entrapment or incitement cannot lead to a fair trial due to its effect on the reliability of the evidence. No entrapment or incitement will be needed to collect open source data. Uncertainty regarding their accuracy and reliability is inherently linked to the make-up of open source data. Third, the right to challenge the evidence means that the person concerned should be given the opportunity to object to the use of such data as evidence implying that he or she needs to be informed of the use of these data.

IV.4. Personal Open Source Data in Criminal Investigations

IV.4.1. Accuracy and Reliability

Since open source data can theoretically be produced and distributed by any person, their accuracy and reliability is difficult to verify. When using such data for an investigation into a criminal offence or an offender, sufficient care should be taken to check source and content of the data. Law enforcement and intelligence authorities have implemented systems for verifying such data. Already in 1987 the CoE recognised the importance of these issues for police authorities in Recommendation (87)15. [83] With this recommendation, a group of experts drafted a special set of data protection principles for the specific tasks of the police while at the same time adapting them to take account of particular requirements, notably in respect of the suppression of criminal offences.[84]

The explanatory text of the recommendation rightfully states that the retaining of personal data in a police file may lead to a permanent record and indiscriminate storage of data, which may prejudice the rights and freedoms of the individual. It is also in the interests of the police that it has only accurate and reliable data at its disposal for the performance of its tasks. For these reasons, these guidelines encourage the implementation of a system of data classification; suggest distinctions between corroborated data and uncorroborated data, including assessments of human behaviour; between facts and opinions; between reliable information (and the various shades thereof) and conjecture; between reasonable cause to believe that information is accurate,

---

[82] ECtHR, 4 October 2000, *Khan v. United Kingdom*, 35394/97, para. 36.

[83] Council of Europe, *Recommendation No.R(87) 15 regulating the use of personal data in the police sector*, 17 September 1987, available online at <wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=2196553&SecMode=1&DocId=694350&Usage=2> (accessed 1 October 2014).

[84] Even though this is not a legally binding instrument, it was explicitly endorsed in legally binding instruments covering police cooperation, such as the Europol Decision, the Schengen Implementation Convention as far as police cooperation is concerned, the Prüm Convention and the decision on the stepping-up of cross-border crime and the decision concerning access for consultation of the Visa Information System (VIS) by designated authorities of Member States and by Europol.

and a groundless belief in its accuracy.[85] For example Europol has not only included Recommendation (87)15 in their standard of data protection,[86] the EU's agency for police cooperation also has a system in place for distinguishing incoming information based on its reliability.

In the current debates on the revision of the EU's data protection legal framework, the proposed directive for data protection in criminal matters included an additional provision on distinguishing personal data in accordance with their degree of accuracy and reliability. Also, the distinction between personal data based on facts and personal data based on personal assessments has been introduced in the text of the proposed directive.[87] Upon adoption of the proposed directive, making such distinction would then be mandatory for all data controllers processing personal data within the scope of this legal instrument.

IV.4.2. Necessity

Information will be collected for the purpose of a criminal investigation for a large part before a criminal charge is made; often it will be collected in order to make a criminal charge in the first place. This means that the protection of Article 6 ECHR is not activated yet, but the collected information can include data on suspects, witnesses as well as victims, and it can range from hard facts to suspicion and mere speculation. These can contain personal data so the data protection standards should apply. In most cases this would mean derogating from the standards as the use of personal data for criminal investigations will be an incompatible purpose as well as a possible breach of the data retention principle. Since derogating from the data protection standards is only lawful when it is laid down in law and necessary in the interests of – in this case – the suppression of criminal offences, the precise meaning of necessity in this respect deserves a closer look.

In the above explanation, necessity was referred to as the link between personal data and the purpose for which they are processed, in this case an investigation into one or more criminal offences. It does not explicitly require a criminal charge to be made, allowing a wider form of information—including proactive—gathering of personal data. Defining this link however, remains a nearly impossible task. In its assessment of the necessity of the mass retention of data in accordance with the Data Retention Directive the EU Court of Justice stated that the fight against serious crime, in particular against organised crime and terrorism, is indeed of the utmost importance in order to ensure public security and its effectiveness may depend to a great extent on the use of modern investigative techniques. The Court continued nonetheless that such an objective of general interest, however fundamental it may be, does not, in itself, justify a retention

---

[85] Council of Europe, *Recommendation No.R(87) 15 regulating the use of personal data in the police sector – Explanatory Memorandum*, 17 September 1987, para. 52, available online at <wcd.coe.int/ViewDoc.jsp?id=704861&Site=CM> (accessed 28 November 2014).

[86] Council Decision of 6 April 2009 establishing the European Police Office (Europol) (2009/371/JHA), OJ L121/37, 15 May 2009, available online at <europol.europa.eu/sites/default/files/council_decision.pdf> (accessed 1 October 2014).

[87] European Parliament legislative resolution of 12 March 2014 on the proposal for a directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data (COM(2012)0010 – C7-0024/2012 – 2012/0010(COD)), 12 March 2014, available online at <europarl.europa.eu/sides/getDoc.do?type=TA&language=EN&reference=P7-TA-2014-0219> (accessed 1 October 2014).

measure such as that established by Directive 2006/24 being considered necessary for the purpose of that fight.[88] If the fight against serious crime is too wide to justify necessity, then a more specific link must exist. The 1987 CoE Recommendation regulating the use of personal data in the police sector gives further indications.[89] With regard to the collection of personal data, the recommendation defines the derogation regarding the suppression of criminal offences as the prevention of a real danger or the suppression of a specific criminal offence. "Real danger" should then be understood as not being restricted to a specific offence or offender but includes any circumstances where there is reasonable suspicion that serious criminal offences have been or might be committed to the exclusion of unsupported speculative possibilities.[90]

Translated into the issue of open source data, this means that it is a lawful exception to the data protection standards when open source data consisting of personal data are collected for the purpose of investigations into a specific criminal offence or offender, or in cases where a reasonable suspicion exists that one or more serious criminal offences have been or might be committed. Purely speculative collection of data—so-called fishing expeditions—does not concern necessary data collection and does not fall within the scope of the lawful derogation.

IV.4.3. Vigilantism

When discussing the topic of open source data and criminal investigations, the relatively recent trend of vigilantism using open sources on social media or the Internet should not be overlooked. What is meant by "vigilantism" or "vigilante justice" is a movement among citizens who take justice into their own hands and—often violently—react to alleged offenders out of discontent with law enforcement's action or lack thereof. Vigilantism in itself is not new, however it has been facilitated in recent years by the expansion of social media.

With estimates ranging from 80-90% of intelligence coming from open sources,[91] it is unsurprising that open sources are abused by persons outside the law enforcement and intelligence community. The fact that open source data are publicly available means that they are often viewed by the public as being used freely. This does not only have data protection violations as a consequence. Referring back to the aforementioned issue of

---

[88] Court of Justice of the European Union, 8 April 2014, *Digital Rights Ireland and Seitlinger and Others*, C 293/12 and C 594/12, para. 51.

[89] Council of Europe, *Recommendation No.R(87) 15 regulating the use of personal data in the police sector*, 17 September 1987, available online at <wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=2196553&SecMode=1&DocId=694350&Usage=2> (accessed 1 October 2014).

[90] In the text of the recommendation a helpful example is added: reasonable suspicion that unspecified drugs were being illegally brought into a country through a port by unidentified private yachts would justify the collection of data on all such yachts using that port, but not all yachts, their owners and passengers using every port in that country. See Council of Europe, *Recommendation No.R(87) 15 regulating the use of personal data in the police sector*, 17 September 1987, available online at <wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=2196553&SecMode=1&DocId=694350&Usage=2> (accessed 1 October 2014).

[91] Congressional Research Service, Best, R.A. and Cumming, A., *Open Source Intelligence (OSINT): Issues for Congress*, 5 December 2007, available online at <fas.org/sgp/crs/intel/RL34270.pdf> (accessed on 26 July 2014), 4.

accuracy and reliability of open source data, in cases of vigilantism, abusing such data can have fatal consequences.[92]

V. Reflections on Open Source Data

Open source data may appear, to the general public, as having an outlaw status and open to all kinds of use. This assumption is essentially incorrect. When open source data contain or consist of data that can identify or enable to identify an individual, they may not be used at free will. Even when the user is a law enforcement or intelligence officer doing his or her job to prevent or investigate a criminal offence, the data protection legislation should be complied with. Use of open source data for the suppression of criminal offences allows derogating from the personal data protection principles; nonetheless the following points deserve special attention.

Open source data are not necessarily verified, accurate or reliable. In comparison to already verified data, law enforcement and intelligence authorities have to invest more resources in organising, filtering and subsequently using open source data that are relevant for preventing and investigating criminal offences. The current revision of the EU's data protection legal framework makes the distinction of personal data based on different degrees of accuracy and reliability mandatory for data processing for the purpose of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties.

Collecting and processing open source data for targeted prevention and investigation into criminal offences constitutes a lawful derogation from the data protection standards. Mass or untargeted personal data collection, however, does not. Regardless of personal data being open source or closed source, the necessity and proportionality requirements apply.

A different light is shed on these data protection standards however by the ECtHR jurisprudence that has ruled on an independent relationship between the right to a private life and the right to a fair trial. When the person concerned had the opportunity over the course of the proceedings to challenge the evidence used against him, an interference with his privacy can still lead to a fair trial. Theoretically, this could endanger the necessity and proportionality requirements, also with regard to open source data. It is essential to closely monitor any future jurisprudence concerning this subject.

Open source data are available in vast amounts on account of the Internet and search engines such as Google, and they are tempting. In that sense, they are also unforgiving with regard to past mistakes and unfortunate life events. It may sound unfair to call this a "new reality", since the use of the Internet and social networks has increased for several decades already. However, the judicial and the legislative process are slow and cumbersome, or, to quote two privacy experts in a reaction to the judgment against Google: 'The CJEU decision is trying to balance things, perhaps assisting individuals a bit more than they deserve, until we all—Internet users, the Internet and Internet companies—get to better grips with the, still new, medium.'[93] On top of getting used to

---

[92] For example The Guardian, Morris, S., *Investigations opened into vigilante murder of man mistaken for paedophile*, 29 October 2013, available online at <theguardian.com/uk-news/2013/oct/29/vigilante-murder-paedophile-bristol-bijan-ebrahimi> (accessed on 26 July 2014).

[93] Privacy Perspectives, De Hert, P. and Papakonstantinou, V., *How the European Google Decision May Have Nothing To Do With a Right to Be Forgotten*, 19 June 2014, available online at <privacyassociation.org/news/a/how-the-european-google-decision-may-have-nothing-to-do-with-a-right-to-be/> (accessed 30 July 2014).

this relatively new reality, modernising the data protection legal framework is an undertaking with many stakeholders and diverging interests at stake. Legislators as well as judges are realising that the new questions that have surfaced need answers and by the time an answer has been found to one question, another issue will have appeared. The aftermath of the judgment against Google shows exactly how challenging this new reality is for those who perform a supervising role in it. It is of fundamental importance that in getting used to this new reality and adapting the existing legal framework to it, we do not lose touch with the data protection principles that have survived technological developments for several decades already.

The particular issues and questions that are triggered by the use of open source data warrant thorough and detailed reflection, although, this is not only the case for legislators and judges. Also the general public should reflect thoroughly on how to behave appropriately in this new reality. Prevention being the best cure, the simple awareness of what could happen once personal data are posted publicly can make a difference. This does not mean that the future of Internet and social media should come with a warning similar to the Miranda rights referred to in the title of this paper; it means that the debate on open source data should not only be held in parliaments and around congress tables but also in living rooms and around kitchen tables.

*