
AUTOMATED JOURNALISM: ARTIFICIAL INTELLIGENCE TRANSFORMS DATA INTO STORIES

When data protection principles and privacy protect the right to
express opinions freely and to receive accurate information

Cédric GOBLET
Lawyer at the Brussels Bar

ABSTRACT: Narrative Science Inc., a company based in Chicago, has developed an artificial intelligence engine called Quill that *«transforms data into stories that are indistinguishable from those authored by people»*. This technology provides us a good opportunity to explore the complex relationship between privacy, data protection and freedom of expression, in the age of the Internet and of Big Data. Furthermore, Narrative Science's project raises questions on the relationship between human and the machine. Paradoxically, the use of a robot writer will not result in greater rationality, but will increase the tendency towards infotainment. More than ever before, the Internet has made the role of journalism a crucial one. Independence, the verification of sources and the search for truth, which are the foundation of journalism, are essential to make sense of the deluge of information we are now exposed to. Protecting journalists' right to freedom of speech, citizens' right to receive quality information, and the right to privacy of readers and users of social networks requires that a distinction be made between journalism and other types of activities. For each of these, specific data processing activities are performed with distinct purposes. The paper will focus on defining these purposes and on identifying the impact their pursuit entails on fundamental rights. The data protection regulation currently in force at both European Union and Council of Europe levels will be examined to determine whether it sufficiently protects these rights. This research will finally lead us to consider how this regulation may be improved.

KEYWORDS: Automated journalism; Artificial Intelligence; Freedom of expression; Right to information; Right to receive information; Privacy; Data protection; Article 9, Directive 95/46/EC; journalistic purposes.

1. AUTOMATED JOURNALISM: THE PROJECT

Narrative Science Inc., a company based in Chicago, has developed a platform called Quill that *«transforms data into stories that are indistinguishable from those authored by people»*¹. This innovative technology incorporates the latest advances in Ar-

1 Narrative Science's website. Retrieved January, 29th, 2013 from <http://www.narrativescience.com/technology> (The content of this page has been updated).

tificial Intelligence and Big Data analytics. Initially, this robot writer was only capable to generate content in specific domains where the vocabulary is limited and the stories follow a predictable pattern, such as sport and real estate. Now, it is used by the business magazine *Forbes* to produce financial reports². During the American presidential election, Quill has analyzed twitter traffic related to the Republican primary candidates to output daily articles about the campaign³. Progressively, computer intelligence is conquering the domain of political journalism. The massive amount of data available in social media constitutes a particularly interesting information source to achieve this goal.

This powerful technology provides us a good opportunity to explore the complex relationship between privacy, data protection and freedom of expression, in the age of the Internet and of Big Data. Furthermore, Narrative Science's project raises questions on the relationship between human and the machine. Paradoxically, the use of a robot writer will not result in greater rationality, but will increase the tendency towards infotainment. More than ever before, the Internet has made the role of journalism a crucial one. Independence, the verification of sources and the search for truth, which are the foundation of journalism, are essential to make sense of the deluge of information we are now exposed to.

Protecting journalists' right to freedom of speech, citizens' right to receive quality information, and the right to privacy of readers and users of social networks requires that a distinction be made between journalism and other types of activities. For each of these, specific data processing activities are performed with distinct purposes. The paper will focus on defining these purposes and on identifying the impact their pursuit entails on fundamental rights. The data protection regulation currently in force at both European Union and Council of Europe levels will be examined to determine whether it sufficiently protects these rights. This research will finally lead us to consider how this regulation may be improved.

-
- 2 Bell, E. (2012). The robot journalist: an apocalypse for the news industry? *The Guardian* (May 13). Retrieved January, 29th, 2013 from <http://www.guardian.co.uk/media/2012/may/13/robot-journalist-apocalypse-news-industry/print>. See also Morozov, E. (2012). A robot Stole My Pulitzer! How automated journalism and loss of reading privacy hurt civil discourse. *Slate Magazine* (March 19). Retrieved February, 11th, 2013 from http://www.slate.com/Articles/technology/future_tense/2012/03/narrat...ists_customized_news_and_the_danger_to_civil_discourse_.single.html.
 - 3 Templon, J. (2012). *Quill Analyzes Presidential Campaign Funding*. Retrieved January, 29th, 2013 from <http://www.narrativescience.com/blog/quill-analyzes-presidential-campaign-funding>. See also Hammond, K. (2012). *Just to Clarify - Generating stories from social media: Getting to the meat of the tweets*. Retrieved January, 28th, 2013 from <http://khammond.blogspot.be/2012/02/generating-stories-from-social-media.html>.

2. RELATIONSHIP BETWEEN FREEDOM OF EXPRESSION AND PRIVACY IN THE AGE OF THE INTERNET AND BIG DATA

2.1. Freedom of expression

Ensuring access to a maximum of information to the largest number of people has long been a priority and a condition of democratic development. Technology has made a great contribution to achieve this goal. It is not mere chance that the 19th century represented the golden age of the press. Techniques for the communication and reproduction of information have undergone great developments during this era. The electric telegraph (1837), photography (1839), the telephone (1871) have brought profound changes to social relations and have provided a wider circulation of information⁴. In addition, developments in means of transportation, such as the railways, have played a role in speeding up the distribution of newspapers and information.

Yet this development would have been impossible under Western democracies without a supporting legal framework; and more specifically without enshrining freedom of expression and of press as a fundamental right. Sweden is believed to be the first country to have adopted a law protecting the freedom of the press in 1766⁵.

In the Member States of the Council of Europe, article 10 of the European Convention on Human Rights⁶ (hereinafter «ECHR»), has played a decisive role in protecting this right for over fifty years. According to this provision, «*everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas*».

The wording taken from article 10 of the ECHR mentions the right to communicate and receive information as two indissociable facets of the same freedom. Many times over, the European Court of Human Rights tells us that «*not only does the press have the task of imparting such information and ideas: the public also has a right to receive them*»⁷. The reference to this part of freedom of expression denotes a greater inclusion of the role of the recipient of the information. This trend continues with the Internet, where the recipient now acts as an information provider.

4 Albert, P. (1970). *Histoire de la presse* (Collection «Que sais-je ?» n°414). Paris: Presses Universitaires de France, pp. 34-35.

5 OECD (2010). *News in the Internet Age: New Trends in News Publishing*. Paris: OECD Publishing, p. 26.

6 Signed in Rome by the Member States of the Council of Europe on 4 November 1950.

7 ECHR, *Sunday Times (No. 1) v. the United Kingdom*, judgment of 26 April 1979, §65. See also case of *Lingens v. Austria*, judgement of 8 July 1986, §41 (All the case-law is available at the Court website, at <http://cmiskp.echr.coe.int>).

The right to receive information essentially places a negative obligation on the State not to infringe on the freedom of receiving information⁸. It is presented as a particular aspect of the right to information, or to be more accurate, as one of the rights to information. The Court does not recognise the guarantee to absolute right to information in article 10. Such an interpretation would enshrine the existence of a general right of access to administrative data and documents⁹, and even the unlimited right to know, as is so often claimed by the tabloid press. To avoid any inaccuracy or confusion, I will exclusively use the concept of the right to receive information in the continuation of this study.

The Court has had the chance to apply the principles it has set out in terms of freedom of expression in cases related to Internet. Among these rulings, we can pinpoint the case of *Times Newspapers Ltd v. the United Kingdom*. It stated that: «*In light of its accessibility and its capacity to store and communicate vast amounts of information, the Internet plays an important role in enhancing the public's access to news and facilitating the dissemination of information generally. The maintenance of Internet archives is a critical aspect of this role and the Court therefore considers that such archives fall within the ambit of the protection afforded by Article 10*»¹⁰.

2.2. Privacy and data protection

In the second half of the 19th century, legal constructs gradually appeared in Western countries to accompany the swift rise in communication techniques. This progress enabled the distribution of information to such a degree that the intimacy, honour and reputation of individuals could be affected¹¹. Excesses of the press at the time were taken into account in the drafting of these new rights.

In Germany, the legal doctrine developed the right of personality (Persönlichkeitsrecht) which can be defined as «*the right of the individual to be an end in itself, to assert itself and to flourish as an end in itself*»¹² and which protects many aspects of human be-

8 Council of Europe, ECHR, Research Division (2011). *Internet: Case-law of the European Court of Human Rights*, 2011, p.20-23. Retrieved February, 5th, 2012 from http://www.echr.coe.int/NR/rdonlyres/E3B11782-7E42-418B-AC04-A29BEDC0400F/0/RAPPORT_RECHERCHE_Internet_Freedom_Expression_EN.pdf.

9 *Ibid.*

10 ECHR, *Times Newspapers Ltd v. the United Kingdom (nos. 1 & 2)*, Judgement of 10 March 2009, §27.

11 Rigaux, F. (2004). Protection de la vie privée. In *Répertoire pratique de droit belge* (Tome IX complément). Bruxelles: Bruylant, p.825.

12 Neuner, C. (1866). *Wesen und Arten der Privatrechtsverhältnisse*. Kiel, Schwers'sche Buchhandlung, p.16.

ings, such as image, reputation, honour, health and family issues¹³. In the United States, S. Warren and L. D. Brandeis¹⁴ evoked the concept of *privacy* through not only American, but also English case law. They defined this new right as «*the right to be let alone*». At the same time, French and Belgian case law witnessed decisions which show certain close similarities with the rulings commented on by Warren and Brandeis¹⁵. We already see the expression private life (*vie privée*) in a lecture given by B. Constant before the Athénée Royal in 1819. He said that «*our liberty has to consist of the peaceful enjoyment of private independence*»¹⁶.

The right to intimacy shines through these legal constructs as a common feature, as the heart of Privacy. It is also this aspect that comes from the wording of article 8 of the ECHR, according to which «*everyone has the right to respect for his private and family life, his home and his correspondence*». For over fifty years, the European Court of Human Rights has played a decisive role in the development of the concept of privacy. It has adopted an extensive and dynamic interpretation¹⁷ of this provision, which has enabled it to move far beyond the strict framework of the right to intimacy. The Court appears to have definitively moved beyond this threshold in the case of *Niemietz v. Germany*, by enshrining «*the right for individuals to establish and develop relationships with other human beings*»¹⁸.

Over the course of cases submitted before the Court, the direction of the right to (informational) self-determination¹⁹, or in other words the right to control over information, is gradually taking shape. The emergence of this new facet of privacy is directly linked to the emergence of information and communications technologies.

Developments in ICT gradually have led to reflection from the 70s onwards²⁰, and then to the adoption of regulations with regard to the processing of personal data. *Convention n° 108 of 28 January 1981 of the Council of Europe* (hereinafter «Convention n°

13 See references cited by Rigaux, F., *op. cit.*, p. 825.

14 Warren, S.D., Brandeis, L.D. (1890). The Right to Privacy. *Harvard Law Review*, pp. 193-220.

15 See references cited by Rigaux, F., *op. cit.*, p. 824;

16 Constant, B. (1819). *De la liberté des anciens comparée à celle des Modernes, lecture to the Athénée Royal of Paris in 1819* (Paris: éd. Mille et une nuits - 2010).

17 Sudre, F. (2005). La construction par le juge européen du droit au respect de la vie privée (Rapport introductif). In Sudre, F. (dir.) *Le droit au respect de la vie privée au sens de la Convention européenne des droits de l'Homme* (Collection «Droit et Justice» n° 63). Bruxelles: Bruylant - Nemesis, p.11.

18 ECHR, *Niemietz v. Germany*, judgement of 16 December 1992, §29.

19 ECHR, *Pretty v. United Kingdom*, judgement of 29 April 2002, § 61.

20 Council of Europe, Committee of Ministers. *Resolution 74(29) of 30 September 1974 on the protection of the privacy of individuals vis-à-vis electronic data banks in the public sector*. Retrieved

108»)²¹ is the first notable text, as it lays out all the basic principles applicable in this area. These principles would later resurface in *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* (hereinafter «Directive 95/46/EC»). This text was transposed to all European Union Member States, ensuring a high standard of protection.

The appearance of the Internet and the mass processing of personal data that we see today confirm the need for a specific body of rules enabling individuals to control personal data which is held by companies and public authorities. The European Court of Human Rights itself considers that Article 8 ECHR applies to the processing of personal data²².

2.3. Reconciling the irreconcilable?

Freedom of expression and privacy are generally presented as only conflicting rights. This is surely the case in most disputes over media and press law, where the appellants invoke image rights, complain of the infringement of their honour, or claim to be the victims of defamatory statements. According to a group of experts for the Council of Europe over data protection, «*the potential for conflict is rendered more acute with the increasing recourse to automation by the various organs of the media*».²³

The conflicting aspect of this relationship is highly evident if we put the freedom of the individual to express an opinion on one side, and the right to intimacy on the other. However, the nature of the relationship is much less clearly identified if we include the right to receive information and the right to (informational) self-determination. As already seen by the members of the *Article 29 Data Protection Working Party* in 1997, these «*two fundamental rights must not be seen as inherently conflicting. In the absence of adequate safeguards for privacy individuals may be reluctant to freely express their ideas. Similarly identification and profiling of readers and users of information services is likely to reduce the willingness of individuals to receive and impart information*»²⁴.

from <https://wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=590512&SecMode=1&DocId=649498&Usage=2>.

21 Convention n° 108 of 28 January 1981 for the Protection of Individuals with regard to Automatic Processing of Personal Data.

22 ECHR, *Rotaru v. Romania*, judgement of 4 May 2000, §43.

23 Council of Europe (1990). *Data Protection and Media, Study prepared by the Committee of experts on data protection (CJ-PD) under the authority of the European Committee on Legal Co-operation (CDCJ)*. Retrieved from: http://www.coe.int/t/dghl/standardsetting/dataprotection/Reports/Media_1990.pdf.

24 Article 29 Working Party 1/97 of 25 February 1997. Data protection and the media. Retrieved from: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/1997/wp1_en.pdf.

All of these fundamental rights form part of a single train of logic which aims to enable individuals to be free, to build their own identity and at the same time interact freely with other people in society. All things considered, the social and individual dimensions of these freedoms seem to be intertwined.

Studying the link between identity and privacy, A. Roosendaal²⁵ writes: «*Making choices and defining wishes and desires is closely related to identity. Identity is who you are as an individual and how you want to be seen by others, so it has an internal and an external element. The internal element can be described as how human beings acquire a sense of self*⁶. *The external element relates to social interaction with others*». These two elements appear to correspond to the two parts of privacy we looked at above: the internal element coincides with the right to intimacy; and the external element, with the right to self-determination.

The aim of the concept of right to self-determination is to provide a response to the following issue: ICT has improved our ability to process data which may relate to identified or identifiable people. This greater control over information considerably increases the powers of control held by public authorities and private entities over the people whose data they are processing. This development brings with it a risk of subservience of individuals to omniscient –and therefore all-powerful– authorities and companies. It has also awakened fears that these individuals may be subject to decisions taken against them based on entirely automated processing. In such a situation, the right to self-determination regarding information enables individuals to maintain control over their data and therefore over their individual destiny, while upholding a relationship with companies, the public authorities and other citizens.

Let us now return to freedom of expression. Case law from the European Court of Human Rights has precisely defined the role of this right in terms of both the personal development of individuals and in the community at large. The participation of citizens in public affairs –which translates as their ability to form an opinion, criticise the powers-that-be and to enter into debate– presuppose access to information. The Court regularly reaffirms that «*freedom of expression is one of the basic conditions for the progress of democratic societies and for the development of each individual*»²⁷. It continues to insist on the importance of this freedom in terms of the press: «*Freedom of the press affords the public one of the best means of discovering and forming an opinion of the ideas and attitudes*

25 Roosendaal, A. (2012). We Are All Connected to Facebook...by Facebook! In Gutwirth, S., Leenes, R., De Hert, P., Poulet, Y. (ed.), *European Data Protection: In Good Health?* Dordrecht, Heidelberg, London, New-York: Springer, p.11.

26 Hekman, S. J. (2004). *Private selves, public identities: Reconsidering identity politics*. University Park: The Pennsylvania State Univ. Press, p22.

27 ECHR, *Handyside v. the United Kingdom*, Judgement of 7 December 1976, § 49.

of their political leaders. In particular, it gives politicians the opportunity to reflect and comment on the preoccupations of public opinion; it thus enables everyone to participate in the free political debate which is at the very core of the concept of a democratic society»²⁸.

In the age of the Internet and of Big Data, it is interesting to note that freedom of expression and privacy have an impact on the quantity and on the nature of information circulating in society. The issue is indeed specifically how to manage what has become a limitless amount of information.

Some may say «*too much information kills information*»²⁹. This statement posits that there is a threshold, beyond which the amount of data is so large that it is no longer possible to process it and produce news articles and information of quality. Beyond this threshold, the objectives of quantity and quality would separate. With this in mind, the Internet and new technology may be a source of hardship...

In this study, we attempt to put some distance from these pessimistic ideas. Access to a growing source of information is incredible good fortune. It helps to free us as individuals and assists with democratic development, under two conditions. The first is that new technology must work for Man and not be used as a tool for enslavement. New technologies have to provide us with tools enabling us to cope with this flood of information. Partially, we already have this. Think of search engines, for example. The second is to consider that this growth in the amount of information available must bring with it an increase in work, which consists of checking that the processed information is true and reliable, and also of putting this information into context. This is precisely the job of journalists. Privacy and data protection rules can enable the fulfilment of these two conditions.

3. CAN DEMOCRACY SURVIVE WITHOUT JOURNALISM? CAN A ROBOT REPLACE A JOURNALIST?

The access to reliable and accurate news is a prerequisite to the empowerment of individuals and for citizen participation in public affairs. This is why journalism plays a central role in a democracy. Independence³⁰, the search for truth and the verification of sources lie at the base of this profession and guarantee access to reliable information.

28 ECHR, *Lingens v. Austria*, Judgement of 8 July 1986, Series A no. 103, p. 26, §42.

29 This fairly well known expression is attributed to a French politician Mamère, N. (1988) *La Dictature de l'audimat*. Paris: La découverte.

30 Independence is not a synonym for neutrality. See Kovach, B. and Rosenstiel, T. (2001). *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect*. New-York: Three Rivers Press (experts on <http://www.nieman.harvard.edu/reports/article/102544/Journalists-Must-Maintain-an-Independence-From-Those-They-Cover.aspx>).

The Internet has revolutionized the manner in which information is produced and processed. A first factor of change concerns the relationship between the public and journalists. With traditional mass media, the information receiver played a more passive role in the communication process. On today's Internet, the information receiver acts properly as an information provider. Blogs, social media and chats have emerged as dynamic tools to publish fresh contents and to express comments about a variety of topics and events. Cyberspace more widely reflects the large spectrum of opinions and views that characterize our society.

However, this very proximity of journalists with the public may undermine the principle of independence. It entails the risk of seeing journalists adapt their positions to what their readers think, or with the dominating perceptions and norms of society. This lack of distance may in turn entail greater conformism.

With the Internet, the daily newspaper is replaced by continually renewed information, available around the clock. A first glance might lead one to think this information flow contributes to the production of a varied range of opinions and contents. Unfortunately, such is not the case. For many journalists, this rhythm is a source of pressure that results in a race to produce contents, sacrificing the time dedicated to on-site research, to background work, and to reflection. Many specialists have noted a trend on the web towards the homogenization of contents and stressed the dangers of churnalism, a practice which consists on producing articles on the basis of previously available content without any information check and without the least critical look³¹.

Speed and interactivity are inherent characteristics of the Internet. Considerable progress can be made for democracy with this powerful tool. However, it all depends on how it is used. **Now, automatic content production is a response to a demand by companies, including press publishers, who must labour with ever-increasing speed to create contents that will bring traffic to their website.** The generation of advertising income depends on the audience reached. However, website traffic depends largely on search engines such as Google, which select pages on the basis of keywords corresponding to web users' searches.

From the rise of internet, we assist to the decline of this traditional medium, of the paper press. This downward trend has been amplified with the economic crisis³². The newspaper industry has to face an increased competition, a drop in advertising revenues and a decrease of the readership³³.

31 Katrandjian, O. (2012). *Churnalism and Its Discontents*. Retrieved February, 11th, 2013 from <http://www.policymic.com/articles/1400/churnalism-and-its-discontents>.

32 See OECD (2010), *op. cit.*, p. 3.

33 OECD (2010), *op. cit.*, pp. 17, 36, 60.

Narrative Science's project flourishes in this context characterized by cuts in editorial resources and budgets for investigation journalism, overworked journalists, understaffed newsrooms and closure of newspapers³⁴. A client of Narrative Science, declared that he was «*impressed by the cost*» and explained that he pays «*less than \$10 for each article of about 500 words –and the price will very likely decline over time*»³⁵. Kris Hammond, a cofounder of Narrative Science, predict that «*in five years, a computer program will win a Pulitzer Prize*». ³⁶ He estimates that 90 percent of news would be written by computers in 15 years³⁷.

Nevertheless, the contents generated by a platform like Quill cannot be classified as journalistic for a number of reasons. Getting rid of the journalist in the process of content creation leads to the loss of all editorial autonomy whose purpose it is to shield the decision-making process from economic pressures when it comes to which topics are to be covered, and how they are to be treated. The result is the risk that a goodly number of news themes that are of no interest to the masses yet are significant in terms of democracy might no longer be covered. The journalist's role as the watchdog of democracy seems to be in jeopardy.

A robot –however efficient it might be– does not allow sources to be verified, nor does it deal with information with a critical eye and with the required remoteness. Indeed, the journalist's task is not limited to gathering, connecting and correlating data. One might be lead to believe that the intervention of a machine would allow the transmission of information which would be neutral, perfectly objective, and would no longer depend on the subjective interpretations and impressions of those who produce it. But that would lose sight of the fact that machines process information according to parameters and algorithms previously defined by their creators. Furthermore, one may seriously question whether there is not in fact a risk of arbitrariness that ensues from the analysis of a variable and changing reality through a predetermined interpretation chart and predefined criteria.

Furthermore, whereas journalists draw most of their information from the real world, a Quill-like robot is fed solely with data drawn from the Internet, *i.e.* mostly user-produced contents found in social networks, in the chat rooms and forums of newspaper

34 OECD (2010), *op. cit.*, pp. 10, 18, 120.

35 Lohr, S. (2011). In Case You Wondered, a Real Human Wrote This Column. *New York Times* (September 10). Retrieved February, 11th, 2013 from http://www.nytimes.com/2011/09/11/business/computer-generated-articles-are-gaining-traction.html?_r=0&pagewanted=print.

36 *Ibid.*

37 Levy, S. (2012). Can an Algorithm Write a Better News Story Than a Human Reporter? *Wired* (April 4). Retrieved February, 11th, 2013 from <http://www.wired.com/gadgetlab/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter>.

sites. But most of the information found through these means represent the expression of comments, impressions, feelings on sundry topics. This therefore leads to a paradox: the use of a machine will not result in greater rationality; on the contrary, it will increase the tendency towards infotainment throughout an Internet which already devotes considerable space to comments and the expression of personal impressions and feelings.

The production of contents from previously published data presents a danger E. Morozov described very well: «*some people might get stuck in a vicious news circle, consuming nothing but information junk food and having little clue that there is a different, more intelligent world out there*». ³⁸

There are those who predict that the Internet will lead to the disappearance of traditional journalism in favour of citizen journalism. On the contrary, the Internet has made the role of journalism a crucial one.

«*The main belief is that better technology equals better communication, and that's not true*»³⁹, explains D. Wolton who underlines the need to take into account the human dimension of communication⁴⁰. He said that «*the problem is not to send information quickly but to have common understanding. The challenge of democracy is to help people live together in peace, and communication isn't always successful...If you put 500,000 computers between Israël and Palestine, you won't get peace*».

4. DATA PROTECTION RULES TO JOURNALISM'S RESCUE

The protection of quality journalism, and more particularly journalists' right to freedom of speech, citizens' right to receive quality information, and the right to privacy of readers and users of social networks requires that a distinction be made between three types of activities: 1. journalism; 2. the study for statistical purposes of data drawn from net user-produced contents; 3. the production of personalized contents according to profiles assigned to readers.

For each of these activities, specific data processing (mostly of personal data) is performed in the context of distinct purposes. Undoubtedly, the purpose principle constitutes one of the most crucial data protection mechanisms. Directive 95/46/EC puts it as follows: «*personal data must be collected for specified, explicit and legitimate purposes*»⁴¹. The corollary to this provision is that personal data cannot be processed later on a man-

38 Morozov, E. (2012), *op. cit., supra*.

39 Public lecture of D. Wolton in the Alliance française du Macao, 28th September 2010. Retrieved from <http://www.alliancefrancaise.org.mo/spip.php?article222&lang=en>.

40 See Wolton, D. (2009). *Informer n'est pas communiquer*. Paris: CNRS éditions.

41 Article 6, 1°, b).

ner which is incompatible with the purpose for which they were collected. The purpose principle permits us to define –and thereby delimit– the power of the person responsible for the processing: the various operations performed on the data must fit within the framework of the purposes defined.

In the following points, we will focus on defining these purposes, identifying the impact their pursuit entails on fundamental rights, and understanding how data protection rules may be applied to each one of them.

Let us remember that, for the most part, the obligations that ensue from Directive 95/46/EC fall on the controller, *i.e.* on «*the natural or legal person which alone or jointly with others determines the purposes and means of the processing of personal data*»⁴². Journalists and press publishers must be deemed jointly controller for the processing performed for journalistic purposes. Generally, press publishers will be held controller for the processing corresponding to the purposes mentioned in the points 2 and 3.

Note that the purpose principle and the notion of controller are to be found in Convention n° 108.⁴³

4.1. Journalistic purposes

Both Convention n° 108 and Directive 95/46/EC contain a provision that manage the conflicts which might arise between the data protection rules they contain and freedom of speech.

Under article 9, 2° of Convention n° 108, «*derogation from the provisions of Articles 5, 6 and 8 of this convention shall be allowed when such derogation is provided for by the law of the Party and constitutes a necessary measure in a democratic society in the interests of (...) b. protecting (...) the rights and freedoms of others*». As part of the current modernization of the Convention, the Convention's Consultative Committee recommends adding, at the end of the previous text, «*notably freedom of expression*».⁴⁴

Article 9 of Directive 95/46/EC states that: «*Member States shall provide for exemptions or derogations from the provisions of this Chapter, Chapter IV and Chapter VI for the processing of personal data carried out solely for journalistic purposes or the purpose of artistic or literary expression only if they are necessary to reconcile the right to privacy with the rules governing freedom of expression*». Recital 17 of the Directive permits a better interpretation of this provision by specifying that: «*the processing of personal data for purposes of*

42 Article 2, d).

43 Articles 2, d) and 5, b).

44 Modernisation proposals adopted by the 29th Plenary meeting, 27-30 November 2012 (T-PD(2012)4Rev3). Retrieved February, 4th, 2013 from http://www.coe.int/t/dghl/standardsetting/dataprotection/modernisation_en.asp.

journalism or for purposes of literary or artistic expression, in particular in the audiovisual field, should qualify for exemption from the requirements of certain provisions of this Directive in so far as this is necessary to reconcile the fundamental rights of individuals with freedom of information and notably the right to receive and impart information, as guaranteed in particular in Article 10 ECHR».

On 25 January 2012, the European Commission established rule proposal 2012/11⁴⁵, the purpose of which is to reform Directive 95/46/EC. This text, currently under discussion in the European Parliament, contains an article 80 which repeats the statement of the aforementioned article 9.

The aforementioned provisions of the Directive and Convention n° 108 both permit a near-complete derogation of data protection rules. When freedom of speech and data protection clash, the interests at stake must be weighed and a balance found on the basis of the proportionality criterion.⁴⁶

Both texts do differ in scope, however. Article 9 of Convention n° 108 addresses, in general terms, all situations where freedom of speech and data protection clash. The text of the Directive, however, concerns «*journalistic purposes or the purpose of artistic or literary expression*». The use of the term «*solely*» demonstrates the restricted nature of this exceptional regime.⁴⁷ Therefore, all the data processing the purpose of which might jeopardize freedom of speech is not necessarily covered by the hypothesis of article 9 of the Directive. Nevertheless, even where the data processing is not protected by the regime of this provision, specific derogations to the Directive's obligations may be allowed if they are necessary to reconcile the opposing fundamental rights.

Whether on the basis of the text of the Directive or of that of Convention n° 108, it is accepted that data processing for journalistic purposes (editorial including electronic publishing) almost requires a derogation to nearly all applicable rules concerning data protection. The journalist's task would become difficult –if not impossible– to perform if the consent of individuals whose data was being processed had to be obtained every time, or if information, especially regarding processing purposes, had to be provided to them.

In order to differentiate which types of data processing do qualify under the Directive's derogation system from those which do not, we should define what must be understood by «*journalistic purposes*».

45 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 (General Data Protection Regulation). Retrieved February, 1st, 2012 from http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm.

46 See ECJ, Case C-101/01, *Bodil Lindqvist*, judgment of 6 November 2003 (All the case law is available at the Court website, at <http://curia.europa.eu>).

47 Article 29 Working Party 1/97, *op. cit.*, p. 8.

The Court of Justice of the European Union (hereinafter «ECJ») has considered this concept in a judgment dated 16 December 2008. It decided that activities «*relating to data from documents which are in the public domain under national legislation, may be classified as 'journalistic activities' if their object is the disclosure to the public of information, opinions or ideas, irrespective of the medium which is used to transmit them. They are not limited to media undertakings and may be undertaken for profit-making purposes*»⁴⁸. It determined that any activity had to be considered as a journalistic one, if it was one «*in which data relating to the earned and unearned income and assets of natural persons are: collected from documents in the public domain held by the tax authorities and processed for publication; published alphabetically in printed form by income bracket and municipality in the form of comprehensive lists; transferred onward on CD-ROM to be used for commercial purposes; and processed for the purposes of a text-messaging service whereby mobile telephone users can, by sending a text message containing details of an individual's name and municipality of residence to a given number, receive in reply information concerning the earned and unearned income and assets of that person*»⁴⁹.

This decision warrants a number of criticisms. The interpretation given by the ECJ to journalistic activities far exceeds the manner in which they are usually understood. The ECJ even includes in this notion those activities which are intended solely to market the data, without any journalistic treatment of the information. Such activities often include significant risks of invasion of privacy and are unrelated to the mission conferred unto journalism in a democracy. While the practice of journalism may also include the search for financial returns, the danger lies in confusing data processing for journalistic purposes with data processing for marketing purposes. And that is all the more worrisome now that, with the Internet, the lines between marketing and journalistic contents tend to get blurred, «*causing problems for people in search of objective information*». ⁵⁰

It is essential to incorporate into the notion of journalistic purposes the ethical norms and the values that are the foundation of journalism. Working Party experts were already moving in this direction in a recommendation adopted in 1997. Furthermore, the role of the media in a democratic society «*to impart information and ideas on all matters of public interest*» –as stressed in the jurisprudence of the European Court of Human Rights⁵¹– should be taken into consideration. Advocate General Kokott had adopted this position in her opinion in the aforementioned case. However, the ECJ did not agree.

48 ECJ (Grand Chamber), *Tietosuoja-valtuutettu v. Satakunnan Markkinapörssi Oy, Satamedia*, judgment of 16 December 2008.

49 *Ibid.*, §34.

50 Custom Content Blog (2012). *Is Content Marketing Invading Traditional Journalism's Turf?* Retrieved January, 29th, 2013 from <http://blog.customcontentcouncil.com/?p=1898>.

51 ECHR, *Lingens v. Austria*, *op. cit.*, § 44.

Therefore, I propose journalistic purposes be defined as: the production of information on topics of general interest, in accordance with ethical and professional obligations that are the foundation of journalism, to wit independence, the verification of sources and the search for truth.

The distribution of information without any commentary falls within the scope of this purpose, whether in publishing a photograph or simply reproducing a document, figures or a set of raw data. Indeed, the distribution of such information presupposes a decision on the part of the journalist and efforts to contextualize the information: he will have come to the conclusion that this information was relevant as part of a debate on a topic of general interest. Let us consider, for example, the publication of a photograph. Its very contents may bring to the public's attention facts of great importance. Yet no journalist who exercises his profession correctly would publish it without first verifying its authenticity, duly stepping back and expressing, if necessary, reservations.

Furthermore, the reuse of personal data previously made available to the public comes under journalistic purposes so long as the aforementioned ethical rules have been respected, which requires the intervention of a flesh-and-blood journalist. The same applies to data obtained from social networks or those published over the Internet. *A contrario*, merely providing personal data out of context, or subject to criteria or purposes which are left entirely to the recipient of the information, does not come under this hypothesis. Thus, the fact of granting access to the fiscal data of various people to users who may consult them out of idle curiosity does not fall within the scope of journalistic activities. Which is not to say that a person's fiscal data cannot be published by the press in certain cases. An excellent illustration of this is the *Fressoz and Roire c. France* case. In that instance, «*the article was published during an industrial dispute at one of the major French car manufacturers. The workers were seeking a pay rise which the management were refusing. The article showed that the company chairman had received large pay increases during the period under consideration while at the same time opposing his employees' claims for a rise*»⁵².

4.2. Identification of the purposes of data processing performed as part of automated «journalism»

The automatic production of contents through the reuse of data drawn from contents uploaded by Internet users has serious consequences regarding fundamental rights. As previously demonstrated, the right of citizens to receive quality information is seriously jeopardized. The same applies to the right to privacy and data protection of users of social networks, chat rooms and forums. Data published on line by a user, to which

52 ECHR, *Fressoz and Roire c. France*, judgement 21 January 1999, §50.

must be added the data resulting from interactions with other people, reveal many personality aspects, ranging from political opinions to consumption habits, and including health, philosophical convictions, sexual life, profession.... Such information allows the creation of highly detailed profiles of individuals.

Here, rules applicable to data protection must be applied fully, as this does not remotely enter into our previous definition of journalistic purposes. Concretely, users of social networks, chat rooms and forums communicated their data to express an opinion, contact other users, and share experiences and contents with them. The robot journalist reuses the data for a purpose which may be defined as follows: to perform statistical studies and research on opinion trends on the web and in social networks. The published results are no longer personal data, but general data that provide a global overview.

The robot's processing is incompatible with the purpose of the initial collection. Therefore, reuse of the data implies in this case the birth of a new process, one which must meet all the requirements of Directive 95/46/EC. It bears noting that the Directive allows Member States to provide for a favourable regime where the data are reused for historical, statistical or scientific purposes, so long as appropriate safeguards are included. In most Member States, these guarantees include the anonymization of the data. This statistical-research purpose, as we have defined it, excludes any profiling of individuals and any marketing purpose. Traffic data may be processed in this case, only if the sole objective is to understand the audience in a global manner. Under no circumstance may such a process exercise any influence over the contents of the articles made available to readers. That would then constitute a marketing purpose.

Statistical purposes must be carefully distinguished from profiling ones. «*Statistics aim at analysing mass phenomena. Statistics allow (...) the drawing of a general affirmation from a series of systematic individual observations (...) In this way, although statistics are based on individual observations, their objective is not to acquire knowledge of the individuals as such (...) Statistical activities (...) are not directed at taking decisions or individual measures, but rather (...) collective judgments or decisions*»⁵³.

Directive 95/46/EC calls for other safeguards, some of which are: the protection of people requires compliance with article 6, 1°, (e), which stipulates that the data must be «*kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed*». Readership studies that are conducted do not require the use of data which permits those involved to be identified. Such data should therefore always be anonym-

53 Council of Europe (1997). *Explanatory Memorandum of Recommendation No.R (97) 18 of the Committee of Ministers to Member States concerning the protection of personal data collected and processed for statistical purposes*. Retrieved from [http://www.coe.int/t/dghl/standardsetting/dataprotection/EM/EM_R\(97\)18_EN.pdf](http://www.coe.int/t/dghl/standardsetting/dataprotection/EM/EM_R(97)18_EN.pdf).

ized, keeping in mind the fact that the simple interconnection of data may easily allow the identification of the individuals concerned. This is especially so given the quantity and nature of the data collected.

The processing of sensitive data, such as medical data or data revealing ethnic origin, political opinions, philosophical beliefs, requires obtaining the prior consent of those persons involved. They must be provided with complete information on the intended purposes, the identity of whoever will receive the data. The persons involved must also be able to exercise effectively their rights, including those to access and correction.

It bears noting that publication of the result of the research is protected by freedom of speech.

4.3. From targeted advertisement to customized «journalistic» contents

*«Imagine creating multiple versions of the same story, with each story's content customized for different audiences and tailored to fit a particular voice, style and tone.»*⁵⁴ This is what we could read until very recently on Narrative Science's website in a section regarding the services Quill could provide in the field of *«Publishing and Media»*.

Data processing as it is viewed here is not intended to consider the contents of blogs and social networks as a source of information that allows us to gain better knowledge of the world that surrounds us and the debates on ideas that are going on in our societies. The intent here is purely a commercial one. Data on newspaper readers are analyzed to produce non-objective but personalized information. Topics of general interest are treated in a non-journalistic manner.

This type of approach is not specific to the Narrative Science project. A. Altert warns us, in an article published in the Wall Street Journal, that *«your E-book is reading you»*⁵⁵. *«The major new players in e-book publishing - Amazon, Apple and Google - can easily track how far readers are getting in books, how long they spend reading them and which search terms they use to find books. Book apps for tablets like the iPad, Kindle Fire and Nook record how many times readers open the app and how much time they spend reading. Retailers and some publishers are beginning to sift through the data, gaining unprecedented insight into how people engage with books»*.

This type of treatment constitutes a serious attack on the right to receive objective information, as well as on the right to privacy of the readers and Internet users. Let us apply the provisions of Directive 95/46/EC. First of all, the consent of those people

54 Retrieved January, 29th, 2013 from <http://www.narrativescience.com/services>.

55 Alter, A. (2012). Your E-Book Is Reading You. *Wall Street Journal* (July 19). Retrieved January, 29th, 2013 from <http://online.wsj.com/article/SB10001424052702304870304577490950051438304.html>.

involved should always be obtained prior to this type of processing. As the purpose is a marketing one, those involved have the right to oppose, without cost, such a process. Furthermore, the processing of sensitive data should be forbidden even should the consent of those persons involved be obtained. The processing of such data is disproportionate in the context of marketing purposes, which constitutes a violation of Article 6, c) of Directive 95/46/EC, according to which the processed data must be «*adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed*».

5. CONCLUSION

The context applicable to data protection is currently being reviewed both by the Council of Europe and the European Union. This therefore provides an opportunity to add to the current text of Directive 95/46/EC a more general provision, such as Article 9 of Convention n° 108, to the derogation applicable to «*journalistic purposes or the purpose of artistic or literary expression*». This also would be the right time to insert in Convention n° 108 a derogation which would be applicable specifically to these purposes, such as the one mentioned in the Directive. The purpose of journalism should clearly be defined in each of these instruments. Moreover, the exact object of the derogations allowed in this matter should be clearly specified. The same needs be done as regards artistic and literary expression. Generally speaking, the derogations applicable to all of these activities should be allowed only if the data processing is performed by human beings, not solely by machines.

BIBLIOGRAPHY

- ALBERT, P. (1970). *Histoire de la presse* (Collection «Que sais-je ?» n° 414). Paris: Presses Universitaires de France, pp. 34-35
- ALTER, A. (2012). Your E-Book Is Reading You, *Wall Street Journal* (July 19). Retrieved January, 29th, 2013 from <http://online.wsj.com/article/SB10001424052702304870304577490950051438304.html>
- BELL, E. (2012). The robot journalist: an apocalypse for the news industry? *The Guardian* (May 13). Retrieved January, 29th, 2013 from <http://www.guardian.co.uk/media/2012/may/13/robot-journalist-apocalypse-news-industry/print>
- CONSTANT, B. (1819). *De la liberté des anciens comparée à celle des Modernes, lecture to the Athénée Royal of Paris in 1819* (Paris: Mille et une nuits – 2010)
- COUNCIL OF EUROPE, COMMITTEE OF MINISTERS. *Resolution 74(29) of 30 September 1974 on the protection of the privacy of individuals vis-à-vis electronic data*

- banks in the public sector*. Retrieved from <https://wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=590512&SecMode=1&DocId=649498&Usage=2>
- COUNCIL OF EUROPE (1990). *Data Protection and Media, Study prepared by the Committee of experts on data protection (CJ-PD) under the authority of the European Committee on Legal Co-operation (CDCJ)*. Retrieved from: http://www.coe.int/t/dghl/standardsetting/dataprotection/Reports/Media_1990.pdf
- COUNCIL OF EUROPE, ECHR, Research Division (2011). *Internet: Case-law of the European Court of Human Rights*, 2011. Retrieved February, 5th, 2012 from http://www.echr.coe.int/NR/rdonlyres/E3B11782-7E42-418B-AC04-A29BEDC0400F/0/RAPPORT_RECHERCHE_Internet_Freedom_Expression_EN.pdf
- CUSTOM CONTENT BLOG (2012). *Is Content Marketing Invading Traditional Journalism's Turf?* Retrieved January, 29th, 2013 from <http://blog.customcontentcouncil.com/?p=1898>
- HAMMOND, K. (2012) *Just to Clarify - Generating stories from social media: Getting to the meat of the tweets*. Retrieved January, 28th, 2013 from <http://khammond.blogspot.be/2012/02/generating-stories-from-social-media.html>
- HEKMAN, S. J. (2004). *Private selves, public identities: Reconsidering identity politics*. University Park: The Pennsylvania State Univ. Press, p.22
- KATRANJIAN, O. (2012). *Churnalism and Its Discontents*. Retrieved February, 11th, 2013 from <http://www.policymic.com/articles/1400/churnalism-and-its-discontents>
- KOVACH, B. & ROSENSTIEL, T.(2001). *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect*. New-York: Three Rivers Press
- LEVY, S. (2012). Can an Algorithm Write a Better News Story Than a Human Reporter? *Wired*, (April 4). Retrieved February, 11th, 2013 from <http://www.wired.com/gadgetlab/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter>
- LOHR, S. (2011). In Case You Wondered, a Real Human Wrote This Column. *New York Times* (September 10). Retrieved February, 11th, 2013 from http://www.nytimes.com/2011/09/11/business/computer-generated-articles-are-gaining-traction.html?_r=0&pagewanted=print
- MOROZOV, E. (2012). A robot Stole My Pulitzer! How automated journalism and loss of reading privacy hurt civil discourse. *Slate Magazine* (March 19). Retrieved February, 11th, 2013 from http://www.slate.com/Articles/technology/future_tense/2012/03/narrat...ists_customized_news_and_the_danger_to_civil_discourse_.single.html
- NEUNER, C. (1866). *Wesen und Arten der Privatrechtsverhältnisse*. Kiel, Schwers'sche Buchhandlung, p.16

- OECD (2010). *News in the Internet Age: New Trends in News Publishing*. Paris: OECD Publishing
- RIGAUX, F. (2004), Protection de la vie privée. In *Répertoire pratique de droit belge* (Tome IX complément). Bruxelles: Bruylant, p.825
- ROOSENDAAL, A. (2012). We Are All Connected to Facebook...by Facebook! In Gutwirth, S., Leenes, R., De Hert, P., Pouillet, Y. (ed.), *European Data Protection: In Good Health?* Dordrecht, Heidelberg, London, New-York: Springer, p.11
- SUDRE, F. (2005). La construction par le juge européen du droit au respect de la vie privée (Rapport introductif). In Sudre, F. (dir.) *Le droit au respect de la vie privée au sens de la Convention européenne des droits de l'Homme* (Collection «Droit et Justice» n°63). Bruxelles : Bruylant - Nemesis, p.11
- TEMPLON, J. (2012). *Quill Analyzes Presidential Campaign Funding*. Retrieved January, 29th, 2013 from <http://www.narrativescience.com/blog/quill-analyzes-presidential-campaign-funding>
- WARREN, S.D., BRANDEIS, L.D. (1890). The Right to Privacy. *Harvard Law Review*, pp. 193-220
- WOLTON, D. (2009). *Informer n'est pas communiquer*. Paris: CNRS éditions