IF PERSONAL INFORMATION IS PRIVACY'S GATEKEEPER, THEN RISK OF HARM IS THE KEY: A PROPOSED METHOD FOR DETERMINING WHAT COUNTS AS PERSONAL INFORMATION

*Éloïse Gratton**

ABSTRACT

In the late sixties and early seventies, with the development of automated data banks and the growing use of computers in the private and public sector, privacy was conceptualized as having individuals "in control over their personal information." The principles of *Fair Information Practices* were elaborated during this period and have been incorporated in data protection laws ("DPLs") adopted in various jurisdictions around the world ever since.

These DPLs protect *personal information,* which is defined similarly in various DPLs (such as in Europe and Canada) as "information relating to an identifiable individual." In the U.S., information is accorded special recognition through a series of sectoral privacy statutes focused on protecting "personally identifiable information" (or "PII"), a notion close to personal information. Going back in time, we can note that identical or at least similar definitions of *personal information* were in fact used in the DPLs dating back to the early seventies, which illustrates that a similar definition of *personal information* was already elaborated at that time, and has not been modified since.

In recent days, with the Internet and the circulation of new

* Éloïse Gratton, LL.M, LL.D, Partner and National Co-Chair of the Privacy Practice Group, McMillan LLP. The author would like to thank Vincent Gautrais, Ian Kerr, Mark MacCarthy and Avner Levin, as well as all of the participants of the PLSC 2013 who provided valuable comments on the earlier draft of this article.

types of information, the efficiency of this definition may be challenged. Recent technological developments are triggering the emergence of new identification tools allowing for easier identification of individuals. Data-mining techniques and capabilities are reaching new levels of sophistication. Because it is now possible to interpret almost any data as *personal information* (any data can in one way or another be related to some individual) the question arises as to how much data should be considered as *personal information*.

When using a literal interpretation of the definition of *personal information*, many negative outcomes may occur. First, DPLs may be protecting all *personal information*, regardless of whether the information is worthy of protection, encouraging a potentially over-inclusive and burdensome framework. This definition may also prove to be under-inclusive as it may not govern certain profiles (falling outside of the scope of the definition), even if these profiles, although they may not "identify" an individual by name, may still be used against the individuals behind them. A literal interpretation of this definition may also create various uncertainties, especially in light of new types of data and collection tools which may identify a device or an object which may be used by one or more individuals.

In light of these issues, various authors have recently proposed potential guidance, mostly on the issue of what "identifiability" actually means. For example, the work of Bercic and George is examining how knowledge of relational database design principles can greatly help to understand what is and what is not *personal data*. Lundevall-Unger and Tranvik propose a different and practical method for deciding the legal status of IP addresses (with regard to the concept of *personal data*) which consist of a "likely reasonable" test, resolved by assessing the costs (in terms of time, money, expertise, etc.) associated with employing legal methods of identification. Schwartz and Solove also argue that the current approaches to PII are flawed and propose a new approach called "PII 2.0," which accounts for PII's malleability. Based upon a standard rather than a rule, PII 2.0 would be based upon a continuum of "risk of identification" and would regulate information that relates to either an "identified" or "identifiable" individual (making a distinction between the two categories), and establishing different requirements for each category.

My contribution in providing guidance on this notion of "identifiability" has to do with using a new method for

interpreting the notion of *personal information*, taking into account the ultimate purpose behind the adoption of DPLs, in order to ensure that only data that were meant to be covered by DPLs will in fact be covered. In the context of proposing such interpretation, the idea is to aim for a level of generality, which corresponds with the highest level goal that the lawmakers wished to achieve. I will demonstrate how the ultimate purpose of DPLs is broader than protecting the privacy rights of individuals, as it is to protect individuals against the *risk of harm* that may result from the collection, use or disclosure of their information. Likewise, with the proposed approach, only data that may present such *risk of harm* to individuals would be protected.

I argue that in certain cases, the *harm* will take place at the point of *collection* while in other cases, at the point where the data will be *used* or even *disclosed*. Instead of trying to determine exactly what "identifiable" individual means, I maintain that a method of interpretation, which is consistent with the original goals of DPLs, should be favoured. Relying and building on Calo's theory and others, I will elaborate a taxonomy of criteria in the form of a decision tree which takes into account the fact that while the *collection* or *disclosure* of information may trigger a more subjective kind of harm (the collection, a feeling of being observed and the disclosure, embarrassment and humiliation), the *use* of information will trigger a more objective kind of harm (financial, physical, discrimination, etc.). The *risk of harm* approach, which I propose, applied to the definition, will reflect this and protect data only at the time that it presents such risk, or in light of the importance or extent of such risk of objective or subjective harm. Accordingly, interpreting the notion of "identifiability" will vary in light of the data handling activity at stake. For instance, while I maintain that the notion of "identifiability" should be interpreted in light of the overall sensitivity of the information being *disclosed* (taking into account other criteria which are relevant in evaluating the risk of subjective harm), I am also of the view that this notion is irrelevant when evaluating information being *used* (only the presence of an objective harm being relevant).

In Part II, I will elaborate on how a literal interpretation of the definition of *personal information* is no longer workable. In light of this, I will be presenting the proposed approach to interpreting the definition of *personal information*, under which the ultimate

purpose behind DPLs should be taken into account. I will then demonstrate that the ultimate purpose of DPLs was to protect individuals against a *risk of harm* triggered by organizations collecting, using and disclosing their information. In Part III, I will demonstrate how this *risk of harm* can be subjective or objective, depending on the data handling activity at stake. I will offer a way forward, proposing a decision-tree test useful when deciding whether certain information should qualify as *personal information*. I will also demonstrate how the proposed test would work in practice, using practical business cases as examples.

The objective of my work is to come to a common understanding of the notion of *personal information*, the situations in which DPLs should be applied, and the way they should be applied. A corollary of this work is to provide guidance to lawmakers, policymakers, privacy commissioners, courts, organizations handling personal information and individuals assessing whether certain information are or should be governed by the relevant DPLs, depending on whether the data handling activity at stake creates a *risk of harm* for an individual. The approach is meant to provide for a useful framework under which DPLs remain efficient in light of modern Internet technologies.

TABLE OF CONTENTS

I. INTRODUCTION

Legal philosopher Jeroen van den Hoven asks:

> Personal data are and will remain a valuable asset, but what counts as personal data? If one wants to protect X, one needs to know what X is.[1]

With the development of automated data banks and the growing use of computers in the private and public sector, privacy was, in the late 1960s and early 1970s, conceptualized as having individuals "in control over their personal information."[2] The principles of Fair Information Practices ("FIPs") were elaborated during this period and have been incorporated in data protection laws ("DPLs") adopted in various jurisdictions around the world ever since.[3]

What is *personal information*? The answer to this question is crucial because DPLs govern only information that qualifies as *personal*.[4] The fact that certain information is *personal* triggers certain rights for individuals with respect to the processing or handling of information relating to them under DPLs: right to be informed of the collection, use and disclosure of their personal information and right to consent to it, access rights to the information to ensure the accuracy of the information, etc.[5] Organizations handling personal information also have certain duties, for instance the obligation to only use accurate information and to ensure the confidentiality and security of this information.[6]

---

[1] Jeroen van den Hoven, *Information Technology, Privacy, and the Protection of Personal Data*, *in* INFORMATION TECHNOLOGY AND MORAL PHILOSOPHY 301, 307 (Jeroen van den Hoven & John Weckert eds., 4th ed. 2010).

[2] *See* ARTHUR R. MILLER, THE ASSAULT ON PRIVACY: COMPUTERS, DATA BANKS, AND DOSSIERS 25 (1971) ("[T]he basic attribute of an effective right of privacy is the individual's ability to control the circulation of information relating to him . . . ."); ALAN F. WESTIN, PRIVACY AND FREEDOM 7 (1967) ("Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."); Charles Fried, *Privacy*, 77 YALE L.J. 475, 482 (1968) ("Privacy is not simply an absence of information about us in the minds of others; rather it is the *control* we have over information about ourselves.").

[3] *See generally* Jean S. Stratford & Juri Stratford, *Data Protection and Privacy in the United States and Europe*, 1998 IASSIST Q. 17, 17 ("Where the U.S. approach has been to provide specific and narrowly applicable legislation, in Europe there are unified supra-national policies for the region. Most countries have implemented these policies with omnibus legislation.").

[4] *See id.* at 17–18.

[5] *See id.* at 18.

[6] *See id.* at 18–19.

Personal information is defined similarly in various national DPLs (such as in Europe and Canada) as information relating to an identified or identifiable individual.[7] *Personal Information* is also close to the notion of *personally identifiable information* (or "PII"), which can be found in various U.S. sectoral laws.[8] This definition (or very similar definitions) can be found in transnational policy instruments such as in the 1980 *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*[9] ("Convention 108"), the 1981 *Guidelines for the Protection of Privacy and Transborder Flows of Personal Data*[10] ("OECD Guidelines"), and more recently, in the Asia-Pacific Economic Cooperation (or "APEC") Privacy Framework[11] ("APEC Privacy Framework"). Going back in time,

[7] Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, art. 2(1) (Can.) (defining personal information as "information about an identifiable individual . . . ."); *see also* Personal Information Protection Act, S.A. 2003, c. P-6.5, art. 1(1)(k) (Can.) (defining personal information as "information about an identifiable individual."); An Act Respecting the Protection of Personal Information in the Private Sector, R.S.Q. 1993, c. P-39.1, s. 2 (Can.) (defining personal information as "any information which relates to a natural person and allows that person to be identified."); Directive 95/46, 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, 1995 O.J. (L 281) 31, 38 [hereinafter Directive 95/46] (defining personal data as "any information relating to an identified or identifiable natural person . . . .").

[8] *See generally* Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814, 1829–36 (2011) (referencing the Children's Online Privacy Protection Act, 15 U.S.C. § 6501–6506 (Supp. 2006); the Gramm-Leach Bliley Act, 15 U.S.C. §§6801–6809; the Video Privacy Protection Act, 18 U.S.C. § 2710; the HITECH Act, Pub. L. No. 111-5, 123 Stat. 226 (2009) (codified as amended in scattered sections of 42 U.S.C.)).

[9] Council of Europe, Convention for the Protection of Individual with Regard to Automatic Processing of Personal Data, 1981, Doc. No. 28.1, at art. 2(a) (1981) [hereinafter Convention 108] ("'[P]ersonal data' means any information relating to an identified or identifiable individual '(data subject)'").

[10] Organisation for Economic Co-Operation and Development (OECD), *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* (1980), *available at* http://www.oecd.org/internet/ieconomy/oecdguidelinesonthe-protectionofprivacyandtransborderflowsofpersonaldata.htm [hereinafter *OECD Guidelines*] (defining personal data as "any information relating to an identified or identifiable individual"). The OECD was created in 1961, which brings together the governments of countries committed to democracy and the market economy from around the world. *See History*, OECD, http://www.oecd.org/about/history (last visited Feb. 2, 2014).

[11] APEC, *Privacy Framework*, 1, 5 (2005), *available at* http://www.apec.org/Groups/Committee-on-Trade-and-Investment/~/media/Files/Groups/ECSG/05_ecsg_privacyframewk.ashx ("Personal information means any information

we can note that identical or at least similar definitions of *personal information* were in fact already used in various European jurisdictions in the seventies.[12] This illustrates that the definition of *personal information* as "information relating to an identifiable individual" was already elaborated about forty years ago, and has not been modified since.[13]

The circumstances have changed fundamentally in the last forty years.[14] Individuals constantly give off personal information through their use of the Internet, which reaches billions of people around the world and serves as a virtual marketplace for products, information, and ideas.[15] The second generation of the Internet makes possible greater interaction and connectedness among online users, and individuals are becoming increasingly involved in managing their own data through online social networks ("OSNs").[16] There are also recent technological developments triggering the emergence of new identification

about an identified or identifiable individual."). APEC was established in 1989 to further enhance economic growth and prosperity, is the premier forum for facilitating economic growth, cooperation, trade and investment in the Asia-Pacific region. *Mission Statement*, APEC, http://www.apec.org/About-Us/About-APEC/Mission-Statement.aspx (last visited Feb. 2, 2014).

[12] *See* SIR NORMAN LINDOP, CHAIRMAN, COMM'N ON DATA PROT., REPORT OF THE COMMITTEE ON DATA PROTECTION 154 (1978) ("Accordingly, we have come to the conclusion that the only feasible definition of 'personal information' for this purpose is any information which relates to any data subject who is, or can be, identified—including the information whereby he can be identified . . . . Here again, we are reinforced in our conclusion by the fact that the foreign statutes all adopt similar definitions. The US privacy Act, for example, uses 'any information about an individual that contains his name . . . or identifying particulars', the Swedish Acts speaks of 'information concerning an individual' and the Norwegian Bill defines it as 'information and assessments which are directly or indirectly traceable to identifiable individuals, associations or foundations'. France, Austria, Denmark and West Germany all use similar terms in their proposed or enacted legislation."); *see also* Committee of Ministers (EC) Resolution (74) 29 on the Protection of the Privacy of Individuals vis-à-vis Electronic Data Banks in the Public Sector of 20 September 1974 [hereinafter EC Resolution (74) 29] ("'[P]ersonal information' means information relating to individuals (physical persons) . . ."); Committee of Ministers (EC) Resolution (73) 22 on the Protection of the Privacy of Individuals vis-à-vis Electronic Data Banks in the Private Sector of 26 September 1973 [hereinafter EC Resolution (73) 22] ("'personal information' means information relating to individuals (physical persons) . . .").

[13] *See generally* LINDOP, *supra* note 12.

[14] Schwartz & Solove, *supra* note 8, at 1820.

[15] *See generally* ÉLOÏSE GRATTON, UNDERSTANDING PERSONAL INFORMATION: MANAGING PRIVACY RISKS 21 (2013).

[16] *Id.* at 24–25.

tools, which allow for easier identification of individuals.[17] Data-mining techniques and capabilities are reaching new levels of sophistication, and the convergence of different technologies now makes it possible for organizations to collect information that are of far more personal nature than before.[18] Because it is now possible to interpret almost any data as *personal information* (any data can in one way or another be related to some identifiable individual)[19] the question arises as to how much data should be considered as *personal information*.

In the first part, I elaborate on how a literal interpretation of the definition of *personal information* and the notion of an *identifiable* individual is no longer workable and may trigger many negative outcomes: triggering an over-inclusive outcome in certain situations (DPLs protecting all personal information, regardless of whether this information may be harmful to individuals or is worthy of protection);[20] triggering an under-inclusive outcome in certain situations (certain information, "on their own," may not qualifying as *personal information* and certain profiles may fall outside of the scope of the definition, although they may still be used to single out an individual and treat him or her differently);[21] and triggering various uncertainties, especially in light of new types of data and collection tools which have recently emerged (which may relate to a device or object).[22] In light of this, I will be proposing a new

[17] *Id.* at 26, 34; *see also* Bradley Malin, *Betrayed By My Shadow: Learning Data Identity via Trail Matching*, J. PRIVACY TECH. 1, 1 (2005) (discussing how in a data-driven society anonymity is impossible as there are simple algorithms that can capture identifying information. For example, highway video cameras record automobiles, IP address are captured when websites are visited, and patient DNA is sequenced and recorded in hospital records); TJ McIntyre, *Alternative Routes to Identifying "Anonymous" Online Users*, IT LAW IN IRELAND BLOG (Feb. 18, 2010), http://www.tjmcintyre.com/2010/02/alternative-routes-to-identifying.html ("The key insight is that sites typically embed multiple external services (such as advertising, stats counters and video hosting) which may either individually or in combination enable the identity of particular users to be pinned down[.]");

[18] *See generally* GRATTON, *supra* note 15, at 35, 38.

[19] *See* Schwartz & Solove, *supra* note 8, at 1816 (stating that due to the advances in computer science, technologists can now take information that on its face may seem to be non-identifiable and make it identifiable information).

[20] *See* discussion *infra* Part II.A.1.

[21] *See infra* Part II.A.2 (There are a growing number of cases where information about an individual may not be directly personally identifiable, but where the individual has some interest based on the use of the information).

[22] *See* discussion *infra* Part II.A.3.

method of interpreting the notion of *personal information*, one which takes into account the ultimate purpose behind the adoption of DPLs which was to protect individuals against a risk of harm which may take place upon organizations collecting, using, or disclosing their personal information.[23] I will elaborate on how the proposed approach will ensure that only information which were meant to be protected by DPLs will in fact be protected, and how this approach may be used to ensure that information qualifying as *personal* is managed in light of its overall sensitivity (for instance by requiring a less stringent consent before being collected used or disclosed, or less stringent security measures for information which present a lower risk of harm).[24]

In the second section, I will demonstrate how this potential harm (or what I refer to as a *risk of harm*) is different depending on the data handling activity at stake.[25] I will elaborate on how the *collection* and *disclosure* of personal information may trigger a subjective kind of harm, while the use of personal information may usually trigger an objective kind of harm.[26] I will then offer a way forward, proposing a decision-tree test useful when deciding whether certain information should qualify as *personal information*.[27]

II. DECONSTRUCTING AND CONSTRUCTING THE DEFINITION

I will first elaborate on the types of problems triggered by a literal interpretation of the definition of personal information.[28] Then, I will present a new approach to interpreting the definition of personal information, one which takes into account the fact that DPLs were ultimately meant to protect individuals against a risk of harm triggered by organizations collecting, using and disclosing their information.[29]

[23] *See* discussion *infra* Part II.B.1.
[24] *See* discussion *infra* Part II.B.2.
[25] *See* discussion *infra* Part III.A.
[26] *See* discussion *infra* Parts III.A, III.B.
[27] *See* discussion *infra* Parts III, III.A.1.a.
[28] *See* discussion *infra* Part II.A.
[29] *See* discussion *infra* Part II.B.1.

A. Challenges with Literal Interpretation Of "Identifiable"

It is my contention that a literal interpretation[30] of the definition of *personal information* and of the term "identifiable" has in many instances either an over-inclusive outcome, an under-inclusive one, or may trigger uncertainty as to which kind of information is in fact "identifiable."

1. Over-Reaching Definition

Reviewing the context of the elaboration of the definition of *personal information*, it is important to bear in mind that this definition was initially meant to be broad.[31] In *Wyndowe v. Rousseau*,[32] the Canadian Court mentions that in light of the fact that "[p]ersonal information is defined . . . as meaning '*information about an identifiable individual*' [] [t]he Act is therefore very far reaching."[33] As Paul Ohm states: "[n]o matter how effectively regulators follow the latest re-identification research, folding newly identified data fields into new laws and regulations, researchers will always find more data field types they have not yet covered. The list of potential PII will never stop growing until it includes everything."[34]

The definition of *personal information* is so broad that almost any information can qualify as *personal*.[35] As a treatise on Canadian privacy law summarizes, "In essence, almost any information in any form that can be attributed to an identified

---

[30] *See* PIERRE-ANDRÉ CÔTÉ ET AL., INTERPRÉTATION DES LOIS 4 (4th ed. 2009) (Can.) (defining literal interpretation as a process which is based on the exact wording of the law).

[31] *See e.g.*, Article 29 Data Protection Working Party, *Opinion 4/2007 on the Concept of Personal Data*, 4 (June 20, 2007), *available at* http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf [hereinafter Article 29 Working Party Opinion 4/2007] (noting that the European Commission and Council both felt that "personal data" should be defined as general and broad as possible so that it could be applied to all identifying information).

[32] Wyndowe v. Rousseau, [2008] F.C. 39, (Can. Ont. Fed. Ct. App.).

[33] *Id*. at para. 40.

[34] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1742 (2010) (footnotes omitted).

[35] *See* OFFICE OF THE PRIVACY COMMISSIONER OF CANADA, A PRIVACY HANDBOOK FOR LAWYERS: PIPEDA AND YOUR PRACTICE 2 (2011) [hereinafter OPCC HANDBOOK] (stating that personal information is a broadly defined term and thus making it difficult to determine what information is and is not personal information).

individual is caught by this expansive definition."[36] In Canada, the federal Privacy Commissioner plays a key role in deciding whether information is "identifiable."[37] The general tendency has been expansionist.[38] As the OPCC stated in its annual report to Parliament in 2001–2002: "[t]he definition is deliberately broad, and in my findings I have tended to interpret it as broadly as possible. . . . I am inclined to regard information as personal even if there is the smallest potential for it to be about an identifiable individual."[39]

According to Canadian case law, information will be "about" an individual when it is not just the subject of that individual, but also relates to or concerns the individual.[40] An individual will also be "identifiable" "where there is a serious possibility that [they] could be identified through the use of that information, alone or in combination with other available information."[41]

Even the mere fact that an individual is wearing a red shirt can constitute an item of personal information. Bercic and George are illustrating this excessive broadness with the following examples:

> The fact that John Smith drives a car of XYZ brand is undisputedly personal data. A related question is whether the fact that this car has an engine capacity of 2000 cm^3 can also be considered personal data. At first sight, the engine's capacity is not personal data (and it is not if taken by itself). Surprisingly, it becomes personal data

---

[36] BARBARA MCISAAC ET AL., THE LAW OF PRIVACY IN CANADA 4–7 (2011); *see also* JEFFREY A. KAUFMAN, ED., PRIVACY LAW IN THE PRIVATE SECTOR: AN ANNOTATION OF THE LEGISLATION IN CANADA 15 (2007) ("It is, therefore, important to note at the outset that the definition of 'personal information' [in PIPEDA] is extremely broad."); STEPHANIE PERRIN ET AL., THE PERSONAL INFORMATION PROTECTION AND ELECTRONIC DOCUMENTS ACT: AN ANNOTATED GUIDE 54 (2001) ("The definition in the Act is limitless in terms of what can be information about an identifiable individual").

[37] *About the Office of the Privacy Commissioner,* OFFICE OF THE PRIVACY COMMISSIONER OF CANADA, http://www.priv.gc.ca/au-ans/index_e.asp (last modified Jul. 19, 2010).

[38] *See* Ohm, *supra* note 34.

[39] GEORGE RADWANSKI, OFFICE OF THE PRIVACY COMMISSIONER OF CANADA, ANNUAL REPORT TO PARLIAMENT 2001–2002 56 (2003); *see also* Gordon v. Can. (Health), [2008] F.C. 258, para. 25 (Can. Ont. Fed. Ct.) (interpreting personal information under PIPEDA); Wyndowe v. Rousseau, [2008] F.C. 39, para. 40 (Can. Ont. Fed. Ct. App.).

[40] Dagg v. Can. (Minister of Finance), [1997] 2 S.C.R. 403, para. 94 (Can. Ott.); *see also* Info. Comm'r of Can. v. Exec. Dir. of the Can. Transp. Accident Investigation & Safety Bd., [2007] 1 F.C.R. 203, paras. 35–57 (Can. Ott. Fed. Ct. App.).

[41] *Gordon*, [2008] F.C. at para. 34.

> as soon as we know that this car is driven or owned by an individual. . . . Similarly, the fact that a piece of land X that is owned by James Moore is worth €100.000 is also personal data. . . . [T]he fact that the water on the piece of land is potable (or not) can become personal data if we know whose piece of land it is or who lives on it. Many other absurd examples like this can be constructed (e.g., the fact that Paris is the capital of France can become personal data if we relate it to John Smith who lives in Paris, the capital of France).[42]

In Europe, the Article 29 Working Party suggests that the concept of personal data includes data providing any sort of information including more general kinds of information,[43] and that the term "any information" contained in Directive 95/46/EC "clearly signals the willingness of the legislator to design a broad concept of *personal data* [and that t]his wording calls for a wide interpretation.[44]

Referring to Murphy's definition of *personal information*[45] (which is consistent with the definition of *personal information* discussed herein) Solove claims that "[it] is too broad because there is a significant amount of information identifiable to us that we do not deem as private."[46] In his own words: "For example, the fact that a person is a well-known politician is identifiable to her, but is not private. Murphy's definition thus provides no reasonable limitation in scope."[47] According to Inness, not all personal information is private as "it is the *intimacy* of this information that identifies a loss of privacy."[48] Nissenbaum argues that "[t]he widely held conception of a right to privacy as a right to control information about oneself [can

---

[42] Boštjan Bercic & Carlisle George, *Identifying Personal Data Using Relational Database Design Principles,* 17 INT'L J. L. & INFO. TECH. 233, 248 (2009).

[43] Article 29 Working Party Opinion 4/2007, *supra* note 31, at 6 ("From the point of view of the content of the information, the concept of personal data includes data providing any sort of information. This covers of course personal information considered to be 'sensitive data' in Article 8 of the directive . . . but also more general kinds of information.").

[44] *Id.*

[45] *See* Richard S. Murphy, *Property Rights in Personal Information: An Economic Defense of Privacy,* 84 GEO. L.J. 2381, 2383 (1996) (defining personal information as "any data about an individual that is identifiable to that individual . . . .").

[46] Daniel J. Solove, *Conceptualizing Privacy*, 90 CALIF. L. REV. 1087 at 1111–12 (2002).

[47] *Id.* at 1112.

[48] JULIE C. INNESS, PRIVACY, INTIMACY AND ISOLATION 58 (1992).

also apply to] protections even in categories of so-called public information, public spaces, and against non-governmental agents."[49]

In certain cases, judges have come to the conclusion that the definition of personal information was to be interpreted more narrowly.[50] For example, in the U.K. case of *Durant v. Financial Services Authority,*[51] the Court of Appeal issued a landmark ruling narrowing the interpretation of what makes data "personal" as information which: "is biographical in a significant sense;" has to have the individual as its focus; and has to affect an individual's privacy "whether in his personal family life, business or professional [activity]."[52]

"In the United States, however, the government is constitutionally prohibited under the First Amendment from interfering with the flow of information, except in the most compelling circumstances."[53] Many authors outline the benefits of a society dominated by open information flows.[54] Protecting all information would ignore the importance of the flow of data for society as a whole, as such flow would be important in economic efficiency.[55] Personal information would also be increasingly used

[49] Helen Nissenbaum, *Privacy as Contextual Integrity,* 79 WASH. L. REV. 119, 154 (2004).

[50] *See* Durant v. Fin. Servs. Auth., [2003] EWCA 1746, [28] (Eng.).

[51] *Id.*

[52] *Id.* at para. 28. Please note that the case has been taken before the European Court of Human Rights as a breach of Article Eight of the European Convention of Human Rights. Article Eight states that everyone has the right to respect to his private and family life, his home and his correspondence. Council of Europe, Convention for the Protection of Human Rights and Fundamental Freedoms, Nov. 4, 1950, 213 U.N.T.S. 222.

[53] Fred H. Cate, *The Changing Face of Privacy Protection in the European Union and the United States,* 33 IND. L. REV. 173, 179–80 (1999).

[54] *Id.* at 220–21; *see also id.* at 174 n.a1 ("My vision is dominated instead by the benefits we all share of a society dominated by open information flows, the wide range of valuable services that such flows make available, the broad array of steps that the very technologies and markets that Professor Krotoszynski laments make available to me to protect my privacy, and fear of burdensome and costly government regulation to protect privacy, such as Europe now enjoys."); Pierre Trudel & Karim Benyekhlef, Approches et Stratégies pour Améliorer la Protection de la Vie Privée dans le Contexte des Inforoutes 4 (1997) (unpublished essay presented to La Commission de la Culture de L'Assemblee Nationale at the University of Montreal) (on file with author) (suggesting that personal information would become social data in the sense that their aggregation would provide valuable information to society).

[55] It would enable cost cutting in the private and/or public sector (identifying individuals with bad credit, thereby protecting lenders as well as the financial system, i.e., the collective). *See* NEIL ROBINSON ET AL., RAND CORP., REVIEW OF

in healthcare, particularly in research and large-scale epidemiological studies.[56]

Moreover, using a literal interpretation of the notion of "identifiable" individual may trigger a system in which organizations and industry players will incur additional costs for complying with DPLs, which have nothing to do with the protection of individuals.[57] Additionally, it may bring about a situation whereby even new types of data will be governed by DPLs, implying certain obligations for organizations managing this data, which may be problematic in certain cases. For instance, Poullet and his colleagues from the *Comité consultatif de la convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel* question how it would be possible to provide a privacy disclosure pertaining to the collection and use of these new types of data and obtain consent from individuals without actually first identifying them.[58] They also raise that it may be difficult for an organization collecting new types of data to grant access if this data has not even been processed. Another "issue with granting access to website recording navigational or *clickstream* data as an online user moves from page to page on its website is that the data collected through these devices does not necessarily belong to one single individual. This entails that providing access [to this data] to an online user [requesting it] may breach the privacy of the other users of the same computer"[59] (since the

THE EUROPEAN DATA PROTECTION DIRECTIVE 12–13 (2009), *available at* http://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR710.pdf; Ron A. Dolin, *Search Query Privacy: The Problem of Anonymization*, 2 HASTINGS SCI. & TECH. L.J. 137, 144 (2010).

[56] *See, e.g.*, ROBINSON ET AL., *supra* note 55, at 14 & n.38.

[57] According to Microsoft: "As data flows increase in volume and complexity, the application of blanket rules will not make sense in many circumstances—they will increase costs without meaningfully enhancing the protections provided to data subjects." MICROSOFT CORP., MICROSOFT RESPONSE TO THE COMMISSION CONSULTATION ON THE LEGAL FRAMEWORK FOR THE FUNDAMENTAL RIGHT TO PROTECTION OF PERSONAL DATA 6 (2009), *available at* http://ec.europa.eu/justice/news/consulting_public/0003/contributions/organisations/microsoft_corporation_en.pdf.

[58] Yves Poullet et al., Comité Consultatif de la Convention pour la Protection des Personnes à L'égard du Traitement Automatisé des Données à Caractère Personnel, Rapport sur L'Aapplication des Principes de Protection des Données aux Réseaux Mondiaux de Télécommunications 34 (2004), *available at* http://www.coe.int/t/dghl/standardsetting/dataprotection/Reports/Poullet_report_2004_fr.pdf.

[59] Éloïse Gratton, *Personalization, Analytics, and Sponsored Services: The*

profile data, *clickstream* data, and other data that could be collected might reveal information of intimate nature).[60]

2. Under-Inclusive Definition

"A strict literal interpretation of [the notion of "identifiable" individual] may result in excluding new types of data since they relate to a device or an object (instead of an individual) or because the identity (name or contact information) of the individual to which they relate is unknown."[61] Such literal interpretation could also encourage a piecemeal approach instead of looking at the big picture.[62] While every piece of information taken "on its own" may not qualify as *personal information*, when considering all the data together as a whole, the profile data may end up identifying an individual.[63] The same reasoning can apply if we consider the volume of information readily available[64] and data correlation and data-mining techniques now available,[65] which trigger the situation whereby a single, insignificant piece

Challenges of Applying PIPEDA to Online Tracking and Profiling Activities, 8 CANADA J.L. & TECH. 316, 316 (2010) (Can.) (footnote omitted).

[60] *See id.* (using the example of an employee who suffers from an embarrassing disease, and shares their workspace with a colleague. In the event that the information concerning the employee's disease is given to online users that request it, the employee would be humiliated).

[61] GRATTON, *supra* note 15, at 108.

[62] Jean-Francois Blanchette & Deborah G. Johnson, *Data Retention and the Panoptic Society: The Social Benefits of Forgetfulness,* 18 INFO. SOC'Y. 33, 33 (2002).

[63] *See* Renée M. Pomerance, *Redefining Privacy in the Face of New Technologies: Data Mining and the Threat to the "Inviolate Personality,"* 9 CANADIAN CRIM. L. REV. 273, 287 (2005) (Can.) ("[I]t is by no means clear that data-mining would be found to offend section 8 of the *Charter*, given that: 1) any single piece of information, standing alone, might not be sufficiently intimate, personal or private to trigger section 8 protection; and 2) because much of the information that is accessed or 'mined' is within the public domain."); *see also* Ian Kerr & Jenna McGill, *Emanations, Snoop Dogs and Reasonable Expectation of Privacy*, 52 CRIM, L. Q., May 2007, at 392, 431 (2007) (Can.) (characterizing the typical approach taken by courts when deciding whether certain information is entitled to a reasonable expectation of privacy as a "jigsaw," since the courts analyze individual pieces of information and, instead, advocating for an analysis that bundles all of the searched for information and then decides whether the bundled information "attract[s] a reasonable expectation of privacy.").

[64] *See* GRATTON, *supra* note 15, at 21; Peter Fleischer, *The Data Deluge*, PETER FLEISCHER: PRIVACY . . . ? (Apr. 21, 2010, 12:55PM), http://peterfleischer.blogspot.com/2010/04/data-deluge.html?spref=tw.

[65] *See e.g.*, GRATTON, *supra* note 15, at 27 (discussing how location data can now be used to identify individuals).

of information may end up identifying an individual.[66]

As it happens, the content of search queries have indeed been found to identify individual people in certain situations.[67] In the AOL privacy scandal in which AOL Research published a compressed text file on one of its websites containing twenty million search keywords punched into AOL's search engine for over 650,000 AOL anonymous users over a 3-month period for research purposes, it was found possible to identify single users on the basis of the content of their combined search queries.[68] AOL ultimately apologized for the disclosure and recognized that it had violated the privacy of its users despite its attempts to anonymize the data.[69] Thus, although isolated pieces of information may not qualify as "personal," the fact that there is a great volume of data easily available may further heighten the ability to trace personal information to an individual.[70]

---

[66] *See e.g.*, *Chapter One: Getting Started The Internet and Privacy*, CTR. FOR DEMOCRACY & TECH. (Oct. 22, 2009), http://www.cdt.org/privacy/guide/start (explaining how use of the network generates detailed information about the individual—revealing where they 'go' (via URLs), who they associate with (via list-servs, chat rooms and news groups), and how they engage in political activities and social behavior); Jerry Berman & Deirdre Mulligan, *Privacy in a Digital Age: Work in Progress*, 23 NOVA L. REV. 551, 554 (1998) ("The data trail, known as transactional data, left behind as individuals use the Internet is a rich source of information about their habits of association, speech, and commerce. Transactional data, click stream data, or 'mouse droppings,' as it is alternatively called, can include the Internet protocol address ('IP address') of the individual's computer, the browser in use, the computer type, and what the individual did on previous visits to the Web site, or perhaps even other Web sites.").

[67] *See, e.g.*, International Data Protection and Privacy Commissioners' Conference, London, United Kingdom, Nov. 2–3, 2006, *Resolution on Privacy Protection and Search Engines*, *available at* http://privacyconference2011.org/htmls/adoptedResolutions/2006_London/2006_L4.pdf.

[68] *Id.*; *see also* Nate Anderson, *AOL Releases Search Data on 500,000 Users*, ARS TECHNICA (Aug. 7, 2006, 11:39 AM), http://arstechnica.com/uncategorized/2006/08/7433 (discussing how easy it was identify users based on their search histories); Michael Barbaro & Tom Zeller, Jr., *A Face is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES (Aug. 9, 2006), http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all&_r=0 (reporting on the identify of one the AOL users whose search history was leaked and how that information was used to identify said user).

[69] Anick Jesdanun, *AOL: Breach of Privacy Was a Mistake*, WASH. POST (Aug. 7, 2009, 9:11 PM), http://www.washingtonpost.com/wpdyn/content/article/2006/08/07/AR2006080700790.html.

[70] *See* GRATTON, *supra* note 15; *see also* Gratton, *supra* note 59, at 299 (discussing how advances in technology make it possible to gather more varied data, which can be used to create a profile for specific users, and in turn identify them).

There are a growing number of cases where information about an individual may not be directly personally identifiable, but where the individual has some interest based on the use of the information.[71] Profiles of individuals, although they may be anonymous and not covered under the definition of *personal information* in all cases (for instance, the profile may contain no name or contact information), may still be used, for instance, to make decisions about a profile, therefore impacting on the individual behind the profile.[72] For example, behavioral advertising may often involve the collection of IP addresses and the processing of unique identifiers (through the use of cookies).[73] "The use of such devices with a unique identifier allows the tracking of users of a specific computer even when dynamic IP addresses are used. In other words, such devices enable data subjects to be targeted or 'singled out', even if their real names [or contact information] are not [necessarily] known." [74] Similar concerns can take place in the offline world, using location data or RFID technology to profile individuals.[75]

---

[71] Certain scholars have raised that RFID tracking without additional identifiers should not be governed by DPLs, while the Article 29 Working Party disagrees on the basis that individuals may have some interest based on the use of the information. *See* Article 29 Data Protection Working Party, *Results of the Public Consultation on Article 29 Working Document 105 on Data Protection Issues Related to RFID Technology* (Sep. 28, 2005) [hereinafter Results of Article 29 Working Party].

[72] *See* Poullet et al., *supra* note 58, at 28. *See generally* Roger Clarke, *Profiling: A Hidden Challenge to the Regulation of Data Surveillance*, 4 J.L. & INFO. SCI. 403, 406–07 (1993) (Austl.) (providing step-by-step analysis of the profiling system used to identify users based off their online activity).

[73] Lisa J. Sotto & Melinda L. McLellan, *Online Behavioral Advertising: A User's Guide,* IP LITIG., Nov.–Dec. 2012, at 1, 1–2.

[74] Article 29 Data Protection Working Party, *Opinion 2/2010 on Online Behavioural Advertising* 9 (June 22, 2010), http://ec.europa.eu/justice/data-protection/index_en.htm.

[75] *See* Article 29 Data Working Party, *Working Document on Data Protection Issues Related to RFID Technology*, 7 (Jan. 19, 2005) [hereinafter Article 29 Working Party Working Doc. on RFID Tech.] ("A further example could be where the use of RFID tags can lead to the processing of personal data, even when RFID technology does not involve the use of other explicit identifiers. Take the hypothesis where person Z walks into Shop C with a bag of RFID tagged products from Shops A & B. Shop C scans his bag and the products in it (more likely a jumble of numbers) are revealed. Shop C keeps a record of the numbers. When person Z returns to the shop the next day, he is rescanned. Product Y, that was scanned yesterday, is revealed today—the number is for the watch he always wears. Shop C sets up a file using the number of product Y as a 'key'. This allows them to track when Person Z enters their shop, using the RFID number of his watch as a reference number for him. This allows shop C to

It has been raised that with new technologies, certain profiles, considered as anonymous, may be used to make decisions about an individual (or a profile).[76] For instance, Amazon was accused of practicing *adaptative pricing* using cookies that would raise the price of certain items in accordance with the profile of the potential purchaser.[77] In this case, although the identity of the individual impacted by this pricing decision is unknown, this individual may still be subject to some type of discrimination or other type of harm, which DPLs may have been meant to address.[78]

It is interesting to note that in Sweden, the Personal Data Act 1998 defines personal data as "all kinds of information that directly or indirectly may be referable to a natural person who is alive."[79] This definition does not refer to the fact that the data needs to "identify" an individual. For example, a website that would propose life insurance policies online, could conclude, rightfully or not, that a particular online visitor is homosexual and is afflicted with AIDS, based on the profile information collected by cookies.[80] The Swedish DPL would therefore apply if

set up a profile of Person Z (whose name they don't know) and to track what he has in his shopping bag on subsequent visits to Shop C. By doing this, Store C is processing personal data and data protection law will apply.").

[76] *See* Clarke, *supra* note 72, at 403; Edward C. Baig, *Internet Users Say, Don't Track Me*, USA TODAY (Dec. 14, 2010, 9:12 PM) (discussing how "[a]nonymous monitoring has become common practice and big business."); Annie Lowery, *How Online Retailers Stay a Step Ahead of Comparison Shoppers*, WASH. POST (Dec. 12, 2010), http://www.washingtonpost.com/wp-dyn/content/article/2010/12/11/AR2010121100143.html (explaining how retailers use a user's cookies to determine their buying history and make predictions for future purchases). Consumers and users of the Internet have expressed their disdain at being monitored and tracked online without their knowledge. *See* FTC, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS 18 (2012) [hereinafter FTC 2012 Recommendations] (explaining that consumers have objected to technologies that can track them, despite learning their name, and creating profiles); *see also* Poullet et al., *supra* note 58, at 25, 28; Scott Cleland, *Americans Want Online Privacy: Per New Zogby Poll*, THE PRECURSOR BLOG (June 8, 2010, 5:28 PM), http://www.precursorblog.com/content/americans-want-online-privacy-new-zogby-poll (stating that 88% of Americans disagree with online tracking).

[77] Poullet et al., *supra* note 58, at 29.

[78] *Id*. The Comité consultatif takes the position that for these reasons, online profiles, even if potentially anonymous, should be covered by the definition of *personal information*.

[79] 3 § Personuppgiftslag (Svensk författningssamling [SFS] 1998:204) (Swed.).

[80] Poullet et al., *supra* note 58, at 33–34.

the notion of "gay person who probably has AIDS" relates, at the time of connection, to a physical person alive, even if such a person is not identifiable by name.[81] Probably in order to address this type of concern (information not identifying the individual although the information may be used against the profile, thereby impacting the individual behind the profile), the Article 29 Working Party very recently issued an Opinion providing further input on the recent 2012 data protection reform discussions in the EU, in which it suggests clarifying that personal data, on top of protecting information "identifying" an individual, also covers "any information allowing a natural person to be singled out and treated differently."[82]

3. "Identifiable" Triggering Uncertainty

The definition of *personal information* is also rather vague since it is not always clear at what point a piece of data can be said to be *identifying* an individual.[83] This legal uncertainty surrounding the notion of "identifiable" individual is problematic for organizations that manage personal information, since they do not know if the data that they are handling is personal information, in which case they would have an obligation to comply with the relevant DPLs.[84]

The OPCC has recently admitted that, "It is not always straightforward to determine whether or not information is 'personal information' for the purposes of PIPEDA."[85] Authors Patrick Lundevall-Unger and Tommy Tranvik ("Lundevall-Unger

[81] *See* Personuppgiftslag (SFS 1998:204) (Swed.).

[82] *See* Article 29 Data Protection Working Party, *Opinion 08/2012 Providing Further Input on the Data Protection Reform Discussions* 5 (Oct. 5, 2012), *available at* http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp199_en.pdf [hereinafter Article 29 Working Party Opinion 08/2012]. The Working Party also recommends changing Recital 24 to explicitly consider IP addresses and cookies as personal data. *Id.* at 5–6.

[83] Bercic & George, *supra* note 42, at 235.

[84] *See id.* at 235–36, 248 (discussing how the definition of personal data is a crucial in step in determining whether or not DPLs, like the EU directive on data protection, applies). If it is determined that the data is personal information than those organizations handling the information would have to comply with certain obligations, which would include informing the individual of which information is collected and for what purpose and obtaining his or her consent; protecting this information using adequate security measures; providing access to the information upon request, etc.

[85] OPCC HANDBOOK, *supra* note 35, at 2.

and Tranvik") rightfully articulate the view that: "[t]he challenge for Fleischer, European data protection agencies and everybody else trying to determine which data are personal (and which are not), is to make sense of the Data Protection Directive's definition of what personal data is."[86]

The Article 29 Working Party in Europe has conducted an analysis of the concept of "personal data" since they noticed that current practices in EU Member States suggested that there was some uncertainty on this issue and more specifically with the notion of "identifiable."[87] Schwartz and Solove have recently suggested that the concept of "identifiability" is complex in part because of the changing landscape of technology.[88]

An illustration of the uncertainty raised by certain pieces of data can be made with IP addresses.[89] As a matter of fact, European privacy advocates don't even agree on whether IP addresses constitute personal data. While some argue that IP addresses should qualify as "personal data" under Directive 95/46/EC,[90] European officials are not consistent on the question. Courts and regulators in Sweden[91] and Spain[92] hold that IP addresses fall within Directive 95/46/EC. In Germany[93] and the U.K.,[94] the opposite position is favored. Even within the same

[86] Patrick Lundevall-Unger & Tommy Tranvik, *IP Addresses: Just a Number?*, 19 INT'L J. L. & INFO. TECH. 53, 54 (2011).

[87] Article 29 Working Party Opinion 4/2007, *supra* note 31, at 3.

[88] Schwartz & Solove, *supra* note 8, at 1836.

[89] Lundevall-Unger & Tranvik, *supra* note 86, at 53.

[90] Article 29 Working Party Opinion 4/2007, *supra* note 31, at 16–17.

[91] John Oates, *Sweden: IP Addresses are Personal . . . Unless You're a Pirate*, THE REGISTER (June 18, 2009), http://www.theregister.co.uk/2009/06/18/sweden_ip_law.

[92] Agencia Española de Protección de Datos, *Statement on Internet Search Engines*, 1 n.1, 2 (Dec. 1, 2007), *available at* https://www.agpd.es/portalwebAGPD/canaldocumentacion/recomendaciones/common/pdfs/declaracion_aepd_buscadores_en.pdf.

[93] Jeremy Mittma, *German Court Rules That IP Addresses Are Not Personal Data*, PROSKAUER (Oct. 17, 2008), http://privacylaw.proskauer.com/2008/10/articles/european-union/german-court-rules-that-ip-addresses-are-not-personal-data. *Cf.* Aoife White, IP Addresses Are Personal Data, E.U. Regulator Says, WASH. POST (Jan. 22, 2008), http://www.washingtonpost.com/wp-dyn/content/article/2008/01/21/AR2008012101340.html (article on Germany's data-protection commissioner, Peter Scharr's, view on IP addresses as personal data).

[94] *See* INFO. COMM'R'S OFFICE, PERSONAL INFORMATION ONLINE, CODE OF PRACTICE 9 (2010) (U.K.) (explaining that in multi-person households where computer logins are shared amongst all the users, IP address cannot be traced to one specific individual, an as a result IP address are not considered personal data).

jurisdiction such as in France, courts are not unanimous on the issue of whether IP addresses are *personal information*. The Cour d'appel de Paris in April and May 2007 took the position that IP addresses were not personal information.[95] In August 2007, the French CNIL issued a press release voicing concern over these two decisions and stating that IP addresses should be considered as *personal information*.[96] In May 2008, the Cour d'appel from Rennes decided that IP addresses were personal information.[97] In January 2009, the Cour de cassation reversed this decision, stating that IP addresses did not constitute personal information.[98] In June 2009, the Tribunal de Grande Instance from Paris took the position that IP addresses are indeed personal information.[99] In February 2010, the Paris Appeal Court, re-aligning with the position of the Cour de Cassation, took the position that IP addresses were not personal information.[100] An analysis of this case law shows that although these French cases all shared similar facts, it was the literal interpretation of the definition (either strict or broad) of *personal information* which was inconsistent throughout French courts and therefore triggered contrary decisions on the same issue, within the same jurisdiction.[101]

A first issue to discuss is whether illegal means that may be used to identify an individual should be considered. The Article 29 Working Party seems to take the position that illegal means

[95] Florence Chafiol-Chaumont, *Do IP Addresses Still Qualify as Personal Data?,* LEXOLOGY (Nov. 1, 2007), http://www.lexology.com/library/detail.aspx?g=0574cac3-7c73-48bb-a6c6-8baa1600cb0b.

[96] Press Release, CNiL, L'Adresse IP est une donnée à Caractère Personnel pour L'ensemble des CNIL Européenes (Aug. 2, 2007) (on file with author). This issue was also recognized by the European Commission. European Commission, Comparative Study on Different Approaches to New Privacy Challenges, in Particular in the Light of Technological Developments: France, 2, 2 n.5 (2010) (by Douwe Korff) [hereinafter Korff Comparative Study].

[97] time.lex, *Study of Case Law on the Circumstances in which IP addresses are Considered Personal Data* 102–03 (May 2, 2011), *available at* http://www.timelex.eu/frontend/files/userfiles/files/publications/2011/IP_addresses_report_-_Final.pdf.

[98] Cour de Cassation [Cass.] [Supreme Court for Judicial Matters] Crim., Jan. 13, 2009, Bull. Crim., No. 13 (Fr.).

[99] Tribunial de Grande Instance [TGI] [Ordinary Court of Original Jurisdiction] Paris, 3e ch., June 24, 2009 (Fr.)

[100] Cour d'Appel [CA] [Regional Court of Appeals] Paris, 12 e. ch., Feb. 1, 2009 (Fr.).

[101] *See* time.lex, *supra* note 97, at 100–08 (providing a summary of relevant French case law).

should be taken into account when evaluating whether data is "identifiable" when it reasons that IP addresses should (almost) always and everywhere be regarded as personal data, in order "to be on the safe side."[102] Various case law in Europe, which were rendered evaluating article 2 of the Directive 95/46/EC (definition of *personal data*) with recital 26 (which suggest that "whereas, to determine whether a person is identifiable, account should be taken of all the means *likely reasonably to be used* either by the controller or by any other person to identify the said person"), also adhere to this view.[103] But the case law rendered on this issue in Europe is contradicting.[104] As a matter of fact, the Paris Appeal Court, in two cases already discussed above, which concern the alleged infringement of copyright by members of a file-sharing network published in April and May 2007, rejected the complainants' arguments, and ruled that IP addresses are not *personal data* arguing that illegal means of unmasking the users of IP addresses should play no part in the "identifiability" assessment.[105] In a 2008 court case, the District Court of Munich also concluded that dynamic IP addresses are not *personal data*, on the basis that "IP addresses are characterized by what the court called 'intrinsic determinability,'" and because "dynamic IP addresses are not personal data because Internet portal or website operators cannot link 'names and faces' to IP addresses

[102] *See* Article 29 Working Party Opinion 4/2007, *supra* note 31, at 17 ("So, unless the Internet Service Provider is in a position to distinguish with absolute certainty that the data correspond to users that cannot be identified, it will have to treat all IP information as personal data, to be on the safe side."). *See generally* Article 29 Data Protection Working Party, *Working Document: Privacy on the Internet—An Integrated EU Approach to Online Data Protection*, 21–30 (Nov. 21, 2000) (discussing the general legalities of personal data on the internet).

[103] In a case argued before the District Court and the Regional Court of Berlin and pertaining to the legal status of IP addresses, the court emphasized that illegal means of linking "names and faces" to IP addresses should not be excluded from the decision-making process. Amtsgericht Berlin-Mitte [AG Berlin-Mitte] [District Court of Berlin-Mitte] Mar. 27, 2007, Az. 5 C 314/06 (Ger.). In a trial case in 2005, the Stockholm Lænsrætt came to the same conclusion as the District Court and the Regional Court of Berlin did, two years later. Lundevall-Unger & Tranvik, *supra* note 86, at n.34.

[104] time.lex, *supra* note 97, at 6.

[105] The IP address does not allow the identification of the individuals using the computer since only the legitimate authority of investigation (the law-enforcement authority) may obtain the user identity from the ISP. Cour' d'appeal [CA] [regional court of appeal] Paris, 13e. ch., Apr. 27, 2007, No. 06/02334 (Fr.).

by employing 'normally available tools.'"[106] The court mentioned that "'normally available tools does not encompass illegal methods of identification, or, more precisely, the possibility that a third party—[such as] an Internet Access Provider—gives portal or website operators access to information about the identity of customers that have been assigned IP addresses by that particular access provider."[107]

Authors Lundevall-Unger and Tranvik argue that illegal means of linking "names and faces" to "name and faceless" IP addresses should never be taken into account when assessing whether or not IP addresses are personal data.[108] They believe that "only legal methods of identification should form the basis of these decisions,"[109] but as discussed earlier, the views are not unanimous on this issue.

A second issue is what kind of costs and resources should be used by an organization to determine if certain data can "identify" an individual and is therefore covered under the definition? The FTC, in its recent 2012 Report, states that, "One industry organization asserted, for instance, that if given enough time and resources, any data may be linkable to an individual."[110] With new, sometimes sophisticated, technologies and the Internet, web 2.0, OSNs and the new trend towards increased cross-site profile linkage, certain data that could not identify an individual may now be able to.[111] The degree of difficulty in identifying an unknown Internet user that should be taken into

[106] Lundevall-Unger & Tranvik, *supra* note 86, at 62. In this case, "[t]he portal operator registered and stored IP addresses in log files, not just for the duration of an individual browsing session, but also after the session had been terminated. The plaintiff argued that this violated the German Data Protection Act because the log files contained [personal data]" (the processing of which is regulated by German law) since the dynamic IP addresses and other information (like date and time of use and websites visited) could unmask the identity of individual users. *Id.* at 62. The full decision can be read in German. *See* MEDIEN INTERNET UND RECHT, http://medien-internet-und-recht.de/volltext.php?mir_dok_id=1769 (last visited Feb. 8, 2014).

[107] "[T]he District Court excludes the use of illegal methods of identification, particularly the possibility of getting unlawful access to unique and identifying information from a third party." Lundevall-Unger & Tranvik, *supra* note 86, at 63.

[108] *Id.* at 58.

[109] *Id.*

[110] FTC 2012 Recommendations, *supra* note 76, at 19–20 (footnote omitted).

[111] *See* GRATTON, *supra* note 15, at 33 ("Today, technological developments are triggering the emergence of new identification tools that allow for easier identification of individuals.").

account when decisions about the identifiability of individuals are made is not always clear.[112]

In 1980, Convention 108 clarified in its Explanatory Report that "identifiable persons" refer to individuals who can be "easily" identified and that it did not cover identification of persons "by means of very sophisticated methods."[113] But this Report does not elaborate on what "sophisticated methods" may entail.[114] Some argue that these methods must consist of an assessment of factors like time, money, expertise, and manpower.[115] Similarly, the Council of Ministers of the Council of Europe adopted (in 1997) a Recommendation on the protection of medical data, which states that natural persons are not identifiable "if identification requires an unreasonable amount of time and manpower."[116] Many European DPLs provide that "identification" must be subject to a reasonableness standard.[117] For example, a definition such as that given in the *German Federal Data Protection Act* could be used as a basis for this interpretation: "[r]endering anonymous mean[s] the [modification] of personal data so that the information concerning personal or material circumstances cannot be attributed to an identified or identifiable person or that such attribution would require a *disproportionate amount of time*, expense and [labour]."[118]

The U.K. DPL has also adopted a similar "reasonableness" test since data are deemed personal if the individual to whom they

---

112 *See* Lundevall-Unger & Tranvik, *supra* note 86, at 57 (describing diverse arguments regarding the ability to identify an internet user, and the use of IP addresses to identify users).

113 Council of Europe, Convention for the Protection of Individuals with Regard tto Automatic Processing of Personal Data, Explanatory Report art. 2 para. 28, Jan. 28, 1981, E.T.S No. 108, *available at* http://conventions.coe.int/Treaty/EN/Reports/Html/108.htm.

114 *See id.* (failing to define "sophisticated methods" in the introductory definitions of the report).

115 *See*, *e.g.*, CHRISTOPHER KUNER, EUROPEAN DATA PRIVACY LAW AND ONLINE BUSINESS 50 (2003) (noting how the "clauses have been drafted from a purely data protection point of view, without taking into account the commercial issues which businesses need to consider when engaging in international commercial transactions.").

116 *Committee of Ministers Recommendation No. R (75) 5 on the Protection of Medical Data* (Feb. 13, 1997), *available at* http://www1.umn.edu/humanrts/instree/coerecr97-5.html.

117 *See infra* note 120 and accompanying text.

118 BUNDESDATENSCHUTZGESETZ [BDSG] [Federal Data Protection Act], Jan. 14, 2003, EBANZ. at 6, part 1, § 3 (Ger.).

relate is identifiable "from those data, [and] other information [] in the possession of, or likely to come into the possession of, the data controller."[119] In Slovenia, the DPL also specifies a reasonableness standard: "where the method of identification does not incur large costs or disproportionate effort or require a large amount of time."[120]

Directive 95/46/EC states at recital 26 that to determine whether a person is "identifiable," account should be taken of "all the means likely reasonably to be used either by the controller or by any other person to identify the said person[.]"[121] The Article 29 Working Party suggests that the criterion of *"all the means likely reasonably to be used*" should in particular take into account all the factors at stake, namely:

> The cost of conducting identification is one factor, but not the only one. The intended purpose, the way the processing is structured, the advantage expected by the controller, the interests at stake for the individuals, as well as the risk of organizational dysfunctions (e.g. breaches of confidentiality duties) and technical failures should all be taken into account.[122]

Still, the challenge is then to determine which means should be deemed "likely and reasonable." Lundevall-Unger and Tranvik ask whether this should include sophisticated, cutting-edge, and expensive means or only off-the-shelf and inexpensive tools and methods.[123]

In Canada, jurisprudence on the concept of *personal information* takes the view that an individual will be "identifiable" where there is a *serious possibility* that they "could be identified through the use of that information, *alone or in combination with other available information*."[124] Unfortunately this jurisprudence offers no further guidance on what kind of efforts should be undertaken by an organization to determine this notion of "serious possibility" or evaluate whether the

---

[119] Data Protection Act, (1998) part 1, § 1, 1 (Eng.).

[120] Personal Data Protection Act of the Republic of Slovenia, (2004) art. 6, § 2.

[121] Directive 95/46, *supra* note 7, at 33.

[122] Committee on Bioethics of the Council of Europe Symposium, Strasbourg, Fr., June 19–20, 2012, *Biobanks and Biomedical Collections: An Ethical Framework for Future Research* 44, *available at* http://www.coe.int/t/dg3/healthbioethic/Activities/10_Biobanks/Proceedings%20biobanks%20e%20final.pdf.

[123] Lundevall-Unger & Tranvik, *supra* note 86, at 56.

[124] Gordon v. Can. (Health), [2008] F.C. 258, para. 34 (Can. Ont. Fed. Ct.) (emphasis added).

information at stake qualifies as *personal information* in light of "other data available."[125]

A few European courts have also looked into this issue.[126] In a case argued before the District Court and the Regional Court of Berlin, the court concluded that dynamic IP addresses must be considered *personal data* under the German *Data Protection Act* since all means of identification, regardless of whether these means are controlled by a third party (an Internet Access Provider) or by the portal operator himself, must be taken into account when making decisions about the identifiability of address-holders.[127] Thus, the only relevant criteria for evaluating the status of IP addresses according to them was the effort (or costs) involved in the identification process.[128] In a trial case in 2005, the Stockholm Lænsrætt landed on the same conclusion: dynamic IP addresses in the hands of Internet portal or website operators are personal data, since "portal or website operators could, without investing too much effort, contact a third party (an Internet Access Provider) and get illegal access to identifying billing information controlled by that third party."[129] In both these cases, the issue of identifiability was reduced to a "likely reasonable" test and it is the amount of effort or costs needed to link "names and faces" to IP addresses that was used as the sole criteria for separating personal from anonymous data.[130]

This issue of identification is closely linked to the issue of anonymization of personal information. Data rendered anonymous is usually no longer subject to substantive rights and obligations elaborated by DPLs.[131] But more and more, there is a blurring of the distinction between personal and anonymized

---

[125] *See id.*

[126] *See supra* notes 104, 106–08 and accompanying text.

[127] *See supra* notes 107–08 and accompanying text.

[128] According to Patrick Lundevall-Unger and Tommy Tranvik, "This implies, for instance, that since the illegal transfer of identifying data from an Internet Access Provider to a portal operator would not strain the resources of either party (a telephone call or an exchange of e-mails is enough), then the 'name and faceless' IP addresses stored by that portal operator are personal data (even if the portal operator has no interest in finding out the identity of individual address-holders)." Lundevall-Unger & Tranvik, *supra* note 86, at 64.

[129] *Id.* at 65.

[130] There are also a few other cases where the IP-addresses-as-personal-data issue has been addressed. *See* Lundevall-Unger & Tranvik, *supra* note 86, at 70–72 (discussing which data are considered personal, how to separate personal from anonymous data, and the authors' approach to the issue).

[131] Directive 95/46, *supra* note 7, at 33.

information. In Canada, in *PIPEDA Case Summary #2009-018*, the OPCC took the position that the psychologist's anonymized peer review notes were the *personal information* of the patient, since "de-identified data will not constitute 'truly anonymous information' when it is possible to subsequently link the de-identified data back to an identifiable individual."[132] In Europe, according to the Article 29 Working Party, "[a]nonymous data in the sense of the Directive [95/46/EC] . . . would [also] be anonymous data that previously referred to an identifiable person, but where that identification is no longer possible."[133] In the U.S., the FTC recently commented that "the traditional distinction between PII and non-PII continues to lose significance due to changes in technology and the ability to re-identify consumers from supposedly anonymous data."[134]

Organizations may employ various techniques to "anonymize" (or de-identify) the personal information they collect before using the data or selling it to third parties.[135] A challenge is that anonymization methods can vary.[136] For example, Google and the Article 29 Working Party recently disagreed on what anonymization of data actually means. After Google revealed its anonymization process,[137] the Article 29 Working Party clarified

---

[132] *PIPEDA Case Summary #2009-018, Psychologist's Anonymized Peer Review Notes are the Personal Information of the Patient*, OFF. OF THE PRIVACY COMM'R OF CANADA (Feb. 23, 2009), http://www.priv.gc.ca/cf-dc/2009/2009_018_0223_e.asp.

[133] Article 29 Working Party Opinion 4/2007, *supra* note 31, at 21.

[134] FTC, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS 43 (2010) [hereinafter FTC 2010 Recommendations]; *see also* FTC 2012 Recommendations, *supra* note 76, at 19 ("[T]he traditional distinction between PII and non-PII has blurred and that it is appropriate to more comprehensively examine data to determine the data's privacy implications.").

[135] *See* JEFF SEDAYAO, INTEL IT, ENHANCING CLOUD SECURITY USING DATA ANONYMIZATION 4–5 (2012), *available at* http://www.intel.com/content/dam/www/public/us/en/documents/best-practices/enhancing-cloud-security-using-data-anonymization.pdf (outlining seven different techniques to anonymize data).

[136] *See id.* at 5 (chart of various methods of anonymization).

[137] Letter from Peter Fleischer, Global Privacy Counsel, Google Inc., to Peter Schaar, Chairman, Article 29 Data Prot. Working Party (June 10, 2007) at 5 (responding to Article 29 Data Protection Working Party's letter dated May 16, 2007: "We are putting significant resources into creating processes for reliably anonymizing data. Although we are still developing our precise technical methods and approach, we can confirm that we will delete some of the bits in logged IP addresses (i.e., the final octet) to make it less likely that an IP address can be associated with a specific computer or user. And while it is difficult to guarantee complete anonymization, the network prefixes of IP

that "such [a process] must be completely irreversible for [Directive 95/46/EC] to no longer apply."[138]

Various studies have challenged the reliability of anonymization, demonstrating that by using publicly available data, anonymized information about a user's online history can be "de-anonymized" to identify users.[139] A recent case in point involves the identification of Netflix customers using anonymized data.[140] This anonymized information, generally related to movie preferences, combined with digital trails left on blogs, chat rooms, and Twitter were used to positively identify Netflix customers.[141] Thus, even a small amount of de-identified data on an individual, once combined with another dataset available either publicly or privately through sale, may still serve to re-identify the individual. With the volume of data available online, it is easier than ever to identify individuals.[142]

Ohm has recently published an article entitled *Broken Promises of Privacy* in which he articulates the view that we

addresses do not identify individual users. Logs anonymization will not be reversible. We will intentionally erase, rather than simply encrypt, logs data so that no one (not even Google) can read it once it has been anonymized. Finally, logs anonymization will apply retroactively and will encompass all of Google's search logs worldwide.").

[138] Article 29 Data Protection Working Party, *Opinion 1/2008 on Data Protection Issues Related to Search Engines*, [2008] 00737/EN WP 148 at 20 [hereinafter Article 29 Working Party Opinion 1/2008].

[139] *See, e.g.*, Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets* 1, 1 (2008), *available at* http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf.

[140] *Id.* at 8–13.

[141] Netflix carried out a project providing a monetary incentive for researchers to improve their movie-recommendation system. The company provided data on 500,000 of its subscribers' ratings of various movies and removed the subscribers' names and other PII. Two researchers at the University of Texas collated this data with reviews found in the database of the International Movie Database (or IMDb) and established the identity of two Netflix subscribers (IMDb's terms of use prevented them from executing a more comprehensive search of their records). According to the study, even attempting to complicate the re-identification task by inserting errors into the dataset would not overwhelm the researchers' algorithm used, which could theoretically identify up to 99% of the Netflix subscribers. *Id.*; *see also* Natasha Singer, *When 2+2 Equals a Privacy Question,* N.Y.TIMES (Oct. 17, 2009), http://www.nytimes.com/2009/10/18/business/18stream.html?_r=0 (discussing the results of the Univ. of Texas researches in their quest to identify Netflix users). The contest ended up causing privacy concerns as a result Netflix cancelled it. Steve Lohr, *Netflix Cancels Contest After Concerns Are Raised About Privacy,* N.Y. TIMES (Mar. 12, 2010), www.nytimes.com/2010/03/13/technology/13netflix.html.

[142] *See generally* GRATTON, *supra* note 15.

should abandon the very concept of PII since it is a fatally-flawed concept given that so much non-PII can be re-identified.[143] Amongst other things he refers to is a landmark study by Latanya Sweeney entitled *Uniqueness of Simple Demographics in the U.S. Population*, which suggests that for 87 percent of the American population, no individual shares their specific combination of ZIP code, birth date (including year), and gender with any other individual.[144] Therefore, these three pieces of often easily accessible information would uniquely identify an individual.

With technologies that are becoming more and more sophisticated and may enable to link an individual to certain data, the notions of "identification" and "anonymization" of data are being challenged.[145] Experts claim that there is always a risk of re-identification with new technologies,[146] and that as the semantic web continues to evolve and tools become more sophisticated, re-identification arguably could become easier.[147]

A third issue is whether information should be evaluated alone or in correlation with other information available when attempting to determine if this information is "identifiable." In most cases, certain information "on its own" may not affect an individual's privacy or be potentially harmful, since it can almost always be associated with more than one individual.[148] For

---

143 Ohm, *supra* note 34, at 1742.

144 Latanya Sweeney, *Uniqueness of Simple Demographics in the U.S. Population* 2 (Carnegie Mellon Univ., Lab. for Int'l Data Privacy, Working Paper No. 3, 2000), *available at* http://dataprivacylab.org/projects/identifiability/paper1.pdf.

145 *See* Ohm, *supra* note 34, at 1742–43.

146 JOEL R. REIDENBERG & PAUL M. SCHWARTZ, DATA PROTECTION LAW AND ONLINE SERVICES: REGULATORY RESPONSES 34 (1998), *available at* http://www.paulschwartz.net/pdf/onlinesvcs_schwartz-reidenberg.pdf ("Yet, anonymity in a network environment is not necessarily absolute. The mapping functions that render data anonymous are not always irreversible.").

147 Ohm, *supra* note 34, at 1716 ("About fifteen years ago, researchers started to chip away at the robust anonymization assumption, the foundation upon which this state of affairs has been built. Recently, however, they have done more than chip away; they have essentially blown it up, casting serious doubt on the power of anonymization, proving its theoretical limits and establishing what I call the easy reidentification result.").

148 *See Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, U.S. DEP'T OF HEALTH AND HUM. SERVICES, http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html (last visited Aug. 19, 2013) (mentioning that identifying information alone (names, address, phone number) would not be

example, many people have the same name, share the same birth date, or share the same address.[149] It is usually the correlation between two information elements that creates a privacy issue, as this reduces the number of individuals from the group sharing the same information elements or ends up referring to a unique individual.[150] An individual may feel that his or her privacy is adequately protected if they can be identified in a group of 10 people, while others may feel that their privacy is adequately protected if they can be identified within a group of 100 people.[151]

The fact that there is a huge volume of data available that can be used to make a link between a piece of data and an individual, triggers a debate as to how certain pieces of data should be treated and what kind of correlation is needed between data and an individual in order for this data to qualify as *personal information*.[152] In the U.S., the FTC has raised the issue as follows:

> Another question is whether applying the framework to data that can be 'reasonably linked to a specific consumer, computer, or other device' is feasible, particularly with respect to data that, while not currently considered 'linkable,' may become so in the future. If not feasible, what are some alternatives? Are there reliable methods for determining whether a particular data set is linkable or may become linkable?[153]

As more pieces of *personal information* become available, it may become easier to link this data to other data since there will likely be more common data elements.[154] The definition as it

considered protected health information).

[149] *See Top Baby Names for 2012*, SOC. SECURITY, http://www.socialsecurity.gov/mobile/content/en/baby-names.html (last visited Aug.19, 2013) (listing the most common baby names in the U.S. for 2012); *see also* Steven Strogatz, *It's My Birthday Too, Yeah*, N.Y.TIMES OPINION BLOG (OCT. 1, 2012, 9:00 AM), http://opinionator.blogs.nytimes.com/2012/10/01/its-my-birthday-too-yeah/?_r=0 (discussing how common it is for individuals, even in the same family, to share the same birthday).

[150] *See* NATIONAL RESEARCH COUNCIL OF THE NATIONAL ACADEMIES, ENGAGING PRIVACY AND INFORMATION TECHNOLOGY IN A DIGITAL AGE 39–40 (James Waldo et al., eds., 2007) (explaining that privacy is a relative concept characterized by the ability to be identified in a certain range, and that it is the combination of personal information, not the information itself, that is a "unique identifier").

[151] *Id.* at 39.

[152] *See id.* at 40.

[153] FTC, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: A PROPOSED FRAMEWORK FOR BUSINESSES AND POLICYMAKERS 43 (2010) [hereinafter FTC Proposed Framework].

[154] As Paul Ohm notes: "The accretion problem is this: Once an adversary has linked two anonymized databases together, he can add the newly linked data to

stands now does not provide clear guidance as to whether correlation is needed between two pieces of information when evaluating them and, if so, whether the correlation should be made taking into account third party data.[155] If one was to take the position that correlation needs to be taken into account, then we must determine what kind of correlation is necessary in order for the information to qualify as *personal*. Some believe that only the data actually available to the data controller should be taken into account.[156] Others, such as U.K. privacy expert Chris Pounder, suggest that the data "likely to become available" to the data controller should be taken into account.[157] Finally, some, such as the Article 29 Working Party, take the position that data in the hands of third parties should be taken into account as well,[158] but Peter Fleischer, CPO of Google Inc., disagrees with this last interpretation.[159]

---

his collection of outside information and use it to help unlock other anonymized databases." Ohm, *supra* note 34, at 1746.

[155] *See* Lundevall-Unger & Tranvik, *supra* note 86, at 56 ("[W]hich third parties, possessing potential means of identification, should be included when determining the question of identifiability?").

[156] The CPO of Google seems to suggest that only the data available to the data controller should be taken into account. *See* Peter Fleischer, *Are IP Addresses "Personal Data"?*, PETER FLEISCHER: PRIVACY . . . ? (Feb. 5, 2007), http://www.peterfleischer.blogspot.com/2007/02/are-ip-addresses-personal-data.html.

[157] In the U.K., personal data means data which relate to a living individual who can be identified from those data, or from those data and other information which is in the possession of, "or is likely to come into the possession of, the data controller." Data Protection Act, (1998) part 1, § 1, 1 (Eng.). According to Chris Pounder, this would mean for example, in the context of the IP address, that the question is whether the individual is identifiable not from the IP address but rather "from other information in the possession of Google" (e.g., a history of transactions). *See* Blog Response from Chris Pounder to Alma Whitten, *Are IP addresses personal?* GOOGLE PUBLIC POLICY BLOG (Feb. 25, 2008, 6:13 AM), http://googlepublicpolicy.blogspot.com.

[158] Article 29 Working Party Opinion 4/2007, *supra* note 31, at 17 (stating that until ISPs can adequately match up users to data, all IP information should qualify as personal data); *see also* Article 29 Data Protection Working Party, *Opinion 2/2002 on the Use of Unique Identifiers in Telecommunications Terminal Equipments: the Example of IPv6*, 3 n.4 (2002) (offering that recital 26 of Directive 95/46 requires for data to become "personal data" the moment there is a link between the data and identity of the user (user of the IP address). These considerations will apply equally to search engine operators. Article 29 Working Party Opinion 1/2008, *supra* note 138, at 8.

[159] Fleischer, *supra* note 156, at 2–3 ("The Working Party have assumed that if an IP address is identifiable by one company (e.g., an ISP) it is personal data as far as all other companies are concerned, even if they have no access to the information that permits an association to the individual. But this assumption

A last issue to consider is whether information identifying a device or an object should qualify as information identifying an individual. In the context of the Internet or new technologies, new types of data or collection tools may relate to an inanimate object.[160] Part of the uncertainty, therefore, results from the fact that these new types of data or tools are more and more divorced from a unique and identifiable individual.[161] They may relate to a machine (clickstream data or data collected from cookies), to an Internet connection (IP address) or a web account.[162] Other types of data relating to ambient technologies may relate to a wireless device (ex: location data) and an object (ex: RFID).[163] These devices or objects may be used by one or more individuals.[164]

A current tendency from the industry is to consider that unique identifiers, and basic biographical information pertaining to these unique identifiers, do not refer to identifiable individuals.[165] In the debate on the status of IP addresses, we can note the stance taken by certain French courts. In determining that IP addresses did not qualify as *personal information*, these courts considered that this data merely identified a machine, and it did not qualify under the definition.[166] In June 2009, a Seattle court in the context of a class action lawsuit brought by consumers against Microsoft also took the position that IP addresses are not personal information based on the same premise, since they can only identify a computer.[167]

---

is very questionable.").

[160] *See* Julie E. Cohen, *Privacy, Ideology, and Technology: A Response to Jeffrey Rosen,* 89 GEO. L.J. 2029, 2039 (2001).

[161] *See* Nissenbaum, *supra* note 49, at 132–34 (using advances in technology (wire-tapping, thermal imaging devices, surveillance) as examples of how drawing the line between what is and what is not private space varies between societies, cultures, and time).

[162] *Id.* at 121–22.

[163] *See generally* Article 29 Working Party Working Doc. on RFID Tech., *supra* note 75, at 4–5 (explaining the uses of RFID technology in varies industries).

[164] *Id.* at 2.

[165] Poullet et al., *supra* note 58, at 31. Many industry players make a distinction between personal information or PII and information collected by electronic means (such as information collected by cookies), which they qualify as Non-PII.

[166] *See* Korff Comparative Study, *supra* note 96, at 2 (providing a definition of personal data and explaining that "the Court of Appeal in Paris twice ruled that IP addresses do not constitute personal data.").

[167] *See* Wendy Davis, *Court: IP Addresses Are Not 'Personally Identifiable' Information*, MEDIAPOST NEWS (July 6, 2009, 6:47 PM), http://www.mediapost.com/publications/article/109242/#axzz2dJ0ZklEa

Data generated through the use of a device may be the result of interventions by a number of individuals; perhaps the members of an extended family each making use of a home PC, a whole student body utilizing a library computer terminal, or potentially hundreds of people purchasing from a networked vending machine.[168] In certain cases, the cookies or IP addresses will be linked with additional information such as a web user account.[169] The data collected would therefore identify an individual since, logically, there can be an assumption that the data relates to a specific individual, the owner of the web account.[170] As for other cases, the interpretation to be given to the definition is not clear.

Also, with certain types of information such as *clickstream* data associated with a group of individuals, will the information be considered anonymous if the aggregation is small? Certain courts and authors have taken the position that if an object or device is linked to a small number of individuals (such as a license plate number), it should be treated as *personal information*.[171] Still, it is debatable what constitutes a "small"

(quoting U.S. District Court Judge Richard Jones, "In order for 'personally identifiable information' to be personally identifiable, it must identify a person. But an IP address identifies a computer [.]").

[168] *See* REIDENBURG & SCHWARTZ, *supra* note 146, at 31–32, (discussing how cookies and IP addresses are not linked to individual people, but rather specific computers, and analogizing IP addresses to license plate numbers to further demonstrate that the owner is not necessarily the user/operator).

[169] *Id.* at 31.

[170] Joel Reidenberg and Paul Schwartz suggest that, to be able to identify a particular user, the information in the cookie file must be linked with other data such as a registration entry at the web site, which is increasingly a typical practice for websites. *Id.* In Canada, "[t]he Commissioner was satisfied [in PIPEDA Case Summary No 2003-162] that the information stored by the temporary and permanent cookies qualified as *personal information* for the purposes of [PIPEDA]" as it pertained to a website customer's profile, which was created when the customer signs in. *PIPEDA Case Summary #2003-162: Archived-Customer Complains About Airline's Use of "Cookies" on its Web Site*, OFFICE OF THE PRIVACY COMM'R OF CAN. (2003), *available at* http://www.priv.gc.ca/cf-dc/2003/cf-dc_030416_7_e.asp (emphasis added). In Europe, the Article 29 Working party has taken a similar approach as they believe that "[w]hen a cookie contains a unique user ID, this ID is clearly *personal data*." *See* Article 29 Working Party Opinion 1/2008, *supra* note 138, at 9 (emphasis added).

[171] *See* Lundevall-Unger & Tranvik, *supra* note 86, at 55 ("It is sufficient, according to Article 2 (a) of the Directive, that he or she *can be* identified (i.e., is "identifiable")—that it is possible to reveal the identify of the person. Knowing someone's car registration number will usually, and without too much effort, allow us to identify the owner of the car if that is our intention. Therefore, the registration number should be treated as personal data . . . ."). *But cf.* Bercic &

enough group to make certain data (such as a computer's IP address) qualify as "identifiable."

Some have taken the position that in the event that a computer is registered against a number of individuals through an IP address, then it is not personal data within the meaning of the definition, because a single individual cannot be identified from such use.[172] At the same time, some have raised that "[w]hile there may be difficulties in determining whether clickstream data correlates with a specific individual, the technologies have become so sophisticated that it is possible to extract personal information from clickstream data and identify specific individuals through this process,"[173] therefore further illustrating

George, *supra* note 42, at 241–43 (arguing that partial identifiers, for example a first and last name, do not rise to the level of personal data, and that "[t]he fact that an individual could be uniquely identified in a few further steps arguably should not mislead one into believing that data related to this as yet unidentified individual should qualify as personal data . . . ."). In France, the CNIL rejected a proposed intelligent transport system in part because of the reliance on collecting and tracking data matched by license plate numbers. The CNIL felt that while a "license plate number merely identifies the owner of the car, and not the actual person driving the car" at any given time, the information is nonetheless "linked to a small group of people (possible drivers of a particular car)" and therefore it had to be considered as being "identifiable." REIDENBERG & SCHWARTZ, *supra* note 146, at 32. In Quebec, the courts have taken a similar view, that license plate numbers should be considered personal information. Syndicat de Autobus Terremont Ltée c. Autobus Terremont Ltée et Paul Imbeau, [2010] QCCA 1050 (Can. Que. C.A.). At the Canadian federal level, the OPCC states: "Specifically, organizations should not collect unique identifying numbers appearing on government-issued documents (driver's licences, health cards, *licence plates*, etc.), for purposes other than those intended by the issuers of these documents," therefore implying that license plate numbers should qualify as personal information. *PIPEDA Case Summary #2010-006: Rapid-oil-change Shop Unnecessarily Scanned Customers' Vehicle Registration Information*, OFFICE OF THE PRIVACY COMM'R OF CAN. (Feb. 9, 2010), http://www.priv.gc.ca/cf-dc/2010/2010_006_0209_e.asp. This view is not synonymous throughout Canada. In March of 2011, the Alberta Court of Appeal held in the appeal of a 2009 lower court decision (dismissing an application for judicial review of a decision by an adjudicator appointed by the Alberta Privacy Commissioner), license plate information is *not* "personal information." Leon's Furniture Ltd. v. Info. & Privacy Comm'r of Alberta, [2011] ABCA 94 (Can. Alta. C.A.). This decision has generated considerable debate over the definition of personal information and the types of information that organizations are allowed to collect, use, or disclose without consent of the individual.

[172] LEE BYGRAVE, DATA PROTECTION LAW: APPROACHING ITS RATIONALE, LOGIC AND LIMITS 318 (Brent Hugenholtz et al., eds., 2002) ("[T]he chance of an IP address (and other clickstream data registered against that address) constituting personal data will be diminished if a multiplicity of persons are *registered* against that address.").

[173] Daniel B. Garrie & Rebecca Wong, *Demystifying Clickstream Data: A*

how the notion of "identifying" data is changing with the Internet and new technologies.

Aside from the fact that certain data may be linked to more than one individual (this would be the case, for example, with clickstream data collected from cookies, if more than one person is using the same device),[174] another and last issue is the degree of accuracy needed to be able to consider the data collected as being "identifying." Some claim that IP addresses are not precise enough in many instances to qualify as *personal information* for at least two reasons which are: multiple users and multiple locations.[175] Many individuals may use the same computer, and thus share the same IP address and without an actual username/password login, no actual identification is facilitated.[176] There could be more than one machine using the same Internet facing IP address *at the same time*.[177] These individuals would all technically appear to be under the same IP address.[178] Although subscriber data pertaining to an IP address "can be helpful to law

European and U.S. Perspective, 20 EMORY INT'L L. REV. 563, 580 (2006); *see also* Daniel B. Garrie, *The Legal Status of Software,* 23 J. MARSHALL J. COMPUTER & INFO. L. 711, 727 (2005) (describing the ease with which someone could use cookies to "hijack" passwords and other personal information).

174 Garrie & Wong, *supra* note 173, at 580.

175 "The first reason, the case of multiple users of the same IP address, is exemplified by a public computer, say at a library. There, many people use the same computer, and thus share the same IP address. A new cookie may be generated each time the web browser is re-opened after a prior user closes it, allowing the search engine to detect a possible change in user. However, without an actual username/password login, no actual identification is facilitated. The second reason that an IP address alone may be insufficient to track a user's queries, multiple locations for the same user, is exemplified by someone using the same laptop from different locations. A user may scatter his queries across multiple IP addresses, some of which he may own, some not. Again, without cookie information, and, in particular, an actual login, the user would not have access to his complete search history via IP address data alone. IP addresses are still informative, however, as they can often be mapped to a small geographical region such as a county or zip code without requiring any non-public information." Dolin, *supra* note 55, at 149.

176 *Id.*

177 *See* Alma Whitten, *Are IP Addresses Personal?* GOOGLE PUB. POL'Y BLOG (Feb. 22, 2008), http://googlepublicpolicy.blogspot.com/2008/02/are-ip-addresses-personal.html (explaining that people can use their laptops at the corner café).

178 The Canadian Federal Court recognised the fact that given the unreliability of the evidence matching IP addresses and pseudonyms to account holders, it would be irresponsible to order disclosure of the identity of an account holder and expose that individual to a lawsuit. *See* BMG Canada v. John Doe, [2004] 3 F.C.R. 241 (Can. Fed. Ct.).

enforcement agents, [] sometimes mistakes are made.[179] In the fall of 2006, "an [ISP] mismatched a customer and an IP address, resulting in a guns-drawn raid by a child-porn squad on a farmer in rural Virginia."[180] This illustrates how the quality of the identifying method will play an important role when linking certain data to an individual.

In light of the above, the application of the definition of *personal information* and the term "identifiable," when using a literal interpretation, can lead to unpredictable or counterintuitive results. In 2007, the Article 29 Working Group issued an opinion on the concept of "personal data" in which they proposed a more relative interpretation of the definition.[181] "While the relative interpretation [proposed by the Article 29 Working Group] is more flexible than the literal one, the three criteria are still very broad,"[182] and this type of interpretation may definitely create over-inclusiveness with the problems previously raised in Part II.A.1.[183]

In order for a law to be efficient, it usually has to provide a result, which adequately translates its goal or purpose.[184] Many DPLs and data protection transnational policy instruments adopted for the last thirty or forty years (the OECD Guidelines, Convention 108, Directive 95/46/EC and the APEC privacy framework to name a few) claim to have been adopted for one of the main purposes of protecting the privacy of individuals.[185] I maintain, in Part II.A.1, that the ultimate purpose was in fact broader than protecting privacy and that it was the protection of individuals against the *risk of harm*, which may take place upon organizations collecting, using, and disclosing their personal information.[186] In this context, it is reasonable to wonder if the definition of personal information, *information that relates to an identifiable individual*, properly translates this main goal of

---

[179] Ellen Nakashima, *The Legal Tangles Of Data Collection*, WASH. POST (Jan. 16, 2007), http://www.washingtonpost.com/wp-dyn/content/article/2007/01/15/AR2007011501301_pf.html.

[180] *Id.*

[181] Article 29 Working Party Opinion 4/2007, *supra* note 31, at 4.

[182] ROBINSON ET AL., *supra* note 55, at 27.

[183] *See supra* Part II.A.1.

[184] Lyria Bennett Moses, *Recurring Dilemmas: The Law's Race to Keep Up With Technological Change*, 2007 U. ILL. J.L. TECH. & POL'Y, 239, 271–72 (2007).

[185] *See, e.g., OECD Guidelines*, *supra* note 10, at Preface; Convention 108, *supra* note 9, at Preamble; Directive 95/46, supra note 7, at paras. 1–2, 10; APEC, *supra* note 11.

[186] *See infra* Part II.A.1; *infra* note 223.

DPLs.

B. Proposing an Interpretation of "Identifiable" Taking Into Account Underlying Risk of Harm

Van den Hoven, as are many others, is of the view that the current legal definition of *personal information* (or *personal data* in Europe) provides no guidance on which data should be governed by DPLs.[187] He suggests that it is essential to the ethics, law, and technology of data protection to identify the parcels of information that actually warrant protection.[188]

1. Using a Purposive Approach to Interpreting Personal Information

Rather than proposing a redrafting of the current definition or worse, as suggested by Ohm, to completely abandon it,[189] a possible alternative would be to discard the literal method of interpreting *personal information* and more specifically the term "identifiable," which has led to so many unwanted outcomes and contradictory results.[190] An interpretation, which takes into account the purposes behind DPLs leaves much more room for interpretation.

The main purpose of adopting a very broad definition of personal information was initially meant to ensure that the law would keep up with technological developments.[191] The interpretation of the FIPs was crucial and would largely

187 van den Hoven, *supra* note 1, at 309.

188 *Id.*

189 *See* Ohm, *supra* note 34, at 1742 (arguing that the concept of personal identifiable information (PII)—the U.S. counterpart of *personal information*—is unworkable and unfixable).

190 *See* Arvind Narayanan & Vitaly Shmatikov, *Privacy and Security: Myths and Fallacies of "Personally Identifiable Information*,*"* 53 COMM. ACM, 24, 26 (2010) (discussing the difficulty in discerning a technical meaning for the phrase "personally identifiable").

191 LINDOP, *supra* note 12, at 13 ("Because the lifetime of the legislation on which we are asked to advise will be substantial, we have informed ourselves both about the current state of the art and about foreseeable developments in it, to ensure that the legislation will not need to be amended by reason of technical changes alone."); *id.* at 18 ("We took these considerations into account when deciding upon our recommendations for data protection legislation. An approach which would have been appropriate in 1970 and 1975 would not be suitable for the technology of 1980 and 1985. Technological developments are happening with increasing speed and economy; this requires flexibility in the mechanics of control to allow new potential threats to be contained.").

determine the effect of the objectives of the FIPs as implemented.[192] In the 1970s, as the FIPs were being established and DPLs began to emerge in certain jurisdictions, it was already very clear that a certain flexibility was required and necessary in the application of FIPs.[193] For instance, the Lindop Report explained that the FIPs were drafted in broad terms, specifically in order to provide some type of flexibility.[194]

As early as the 1970s, businesses and various organizations were already raising warning flags over potential restrictions to their data processing activities, if the objectives of the FIPs were to be given a strict, literal interpretation.[195] In answer to these concerns, certain documents from the 1970s, generated from European jurisdictions, illustrate that the FIPs and their underlying obligations on users were to be imposed only "as far as is reasonably practicable."[196] Clearly, the original intention of DPLs was not to ensure that every conceivable data handling activity be covered.[197]

Proposing an interpretation has various benefits. It is always less disturbing to provide a solution which will be incorporated within the current legal framework, such as a proposed interpretation, than proposing something completely new. Many industry players are vouching for global privacy standards.[198]

---

[192] *Id*. 45–46.

[193] *See* Fred H. Cate, *The Failure of Fair Information Practice Principles*, *in* CONSUMER PROTECTION IN THE AGE OF 'INFORMATION ECONOMY' 344, 344 (Jane K. Winn ed., 2006) (discussing the history of broad FIPP regulations dating back to the 1970s).

[194] LINDOP, *supra* note 12, at 199; *see also id*. at 147–48 (explaining the broad scope of FIP regulations).

[195] *Id.* at 200 ("The great majority of users who commented on the objectives set out in paragraph 34 of the White Paper were concerned to explain to us the problems which would be imposed on their data processing activities if those objectives were to be strictly applied, across the board, to everyone and with literal adherence to their wording.").

[196] *Id*. at 199.

[197] *See id*. at 147 (stating that regulation should be "as light and flexible as possible," but only strictly enforced when there is an extraordinary circumstance to encroach upon individual privacy).

[198] Jose Vilches, *Google Proposes Global Privacy Standard*, TECHSPOT (Sept. 14 2007, 10:22 AM), http://www.techspot.com/news/27032-google-proposes-global-privacy-standard.html ("As the Commission considers if, and how, the Data Protection Directive should be reformed, we encourage examination of regulatory developments in other jurisdictions and movement towards a more harmonised global regime. As the patchwork of worldwide data protection laws has become increasingly difficult to navigate, Microsoft has repeatedly called for a comprehensive, workable global privacy framework that is consistent, flexible,

Many jurisdictions are also aligning their practices with one another in an attempt to promote some type of global consistency in the data protection arena.[199]

For instance, in Canada, the OPCC recently concluded that *work product* should not be exempt from the definition of personal information in PIPEDA, one of the reasons being that the current definition of personal information is based on known Canadian and "International precedent and consensus."[200] More specifically, the OPCC was concerned that the introduction of a *work product* exemption would mean that Canada would be taking a position different from that taken in other jurisdictions, particularly in Europe.[201] While this concern (promoting consistency across jurisdictions) makes even more sense in today's world, with the web and related technologies, we can note that this concern has been around for a while.[202]

transparent and principles-based."); MICROSOFT CORP., *supra* note 57 at 9; *see also* INT'L BUS. MACHINES CORP. (IBM), IBM COMMENTS TO FTC STAFF PRELIMINARY REPORT ON PROTECTING CONSUMER PRIVACY-FILE NO. P095416 2 (2011) (advocating for an internationally harmonious approach to privacy practices); NUALA O'CONNOR KELLY, COMMENTS OF THE GENERAL ELECTRIC COMPANY (2010) (encouraging international harmonization).

[199] *EU Ministers Agreed to Seek Better Data Protection,* EU2013.LT (Jul. 19, 2013, 1:58 PM), http://www.eu2013.lt/en/news/pressreleases/eu-justice-ministers-agreed-to-seek-better-data-protection.

[200] *The Privacy Commissioner of Canada's Position at the Conclusion of the Hearings on the Statutory Review of PIPEDA*, OPCC (Feb. 22, 2007), http://www.priv.gc.ca/parl/2007/sub_070222_03_e.cfm [hereinafter OPCC, Privacy Comm'r of Can. Position] (discussing how the Canadian definition of "personal information" was based on its stability in Canadian law and similarity to European law).

[201] While the OPCC admitted that it had not investigated whether a change in PIPEDA's definition of personal information would affect the perception that PIPEDA was sufficiently harmonized with European law, it noted that during a recent review of the Directive 95/46/EC, the EC was asked to add a "work product" exemption to the Directive 95/46/EC's definition of personal information and that in general, the EC advised against modifying it. *Id.*

[202] Even back in the 1970s, certain legislators or committees in charge of analyzing data protection issues, such as the Lindop Committee in the U.K., were very cautious about ensuring the cross-jurisdictional consistency of how *personal information* was defined: "Accordingly, we have come to the conclusion that the only feasible definition of 'personal information' for this purpose is any information which relates to any data subject who is, or can be, identified—including the information whereby he can be identified, as for example his name, address, date of birth, or telephone number . . . . Here again, we are reinforced in our conclusion by the fact that the foreign statutes all adopt similar definitions. The US Privacy Act, for example, uses 'any information about an individual that contains his name . . . or identifying particulars', the Swedish Acts speaks of 'information concerning an individual' and the

In Canada, the OPCC has also chosen not to implement a literal interpretation of the notion of *personal information* when evaluating information produced in the context of an individual's employment and instead, favour an approach they refer to as the "total context approach."[203] Ostensibly, the OPCC has implemented what it sees as a more practical and logical approach.[204] In Europe, the Article 29 Working Group suggests that it is important to keep the ultimate purposes of Directives 95/46/EC and Directive 2002/58/EC[205] in mind when interpreting and applying the rules of both instruments.[206]

In his book entitled *Purposive Interpretation in Law*, leading judge and legal theorist, Aharon Barak, argues that while legal philosophers and jurists apply different theories of interpretation to statutes and rules, a purposive interpretation would probably be more beneficial.[207] He suggests that this method would allow jurists and scholars to approach all legal texts in a similar

Norwegian Bill defines it as 'information and assessments, which are directly or indirectly traceable to identifiable individuals, associations or foundations'. France, Austria, Denmark and West Germany all use similar terms in their proposed or enacted legislation." LINDOP, *supra* note 12, at 154.

[203] OPCC, Privacy Comm'r of Can. Position, *supra* note 200. For examples of the OPCC's "total context" approach, *see PIPEDA Case Summary #2003-220, Telemarketer Objects to Employer Sharing Her Sales Results with Other Employees,* OFFICE OF THE PRIVACY COMM'R OF CAN., (Sept. 15, 2003) http://www.priv.gc.ca/cf-dc/2003/cf-dc_030915_e.asp; *PIPEDA Case Summary #2005-303*, *Real Estate Broker Publishes Names of Top Five Sales Representatives in a City,* OFFICE OF THE PRIVACY COMM'R OF CAN. (May 31, 2005), http://www.priv.gc.ca/cf-dc/2005/303_20050531_e.asp.

[204] OFFICE OF THE PRIVACY COMM'R OF CAN., A GUIDE FOR SUBMITTING PRIVACY IMPACT ASSESSMENTS TO THE OFFICE OF THE PRIVACY COMM'R OF CAN.: EXPECTATIONS 4 (2003).

[205] Directive 2002/58, of the European Parliament and of the Council of 12 July 2002 on the Processing of Personal Data and the Protection of Privacy in the Electronic Communications Sector (Directive on Privacy and Electronic Communications, 2002 O.J. (L 207) 37, 38 43 [hereinafter Directive 2002/58].

[206] "Articles 1 of Directive 95/46/EC and of Directive 2002/58/EC clearly state the ultimate purpose of the rules contained therein: to protect the fundamental rights and freedoms of natural persons and in particular their right to privacy, with regard to the processing of personal data. This is a very important element to take into account in the interpretation and application of the rules of both instruments. It may play a substantive role in determining how to apply the provisions of the Directive to a number of situations where the rights of individuals are not at risk, and it may caution against any interpretation of the same rules that would leave individuals deprived of protection of their rights." Article 29 Working Party Opinion 4/2007, *supra* note 31, at 4.

[207] AHARON BARAK, PURPOSIVE INTERPRETATION IN LAW 31 (Sari Bashi trans., 2005).

manner, while remaining sensitive to important differences.[208] Gautrais and Trudel articulate the view that legal interpretation should place an emphasis on the global context of any given law, its purpose as well as the original intent of the lawmaker.[209] They refer to Gregory Mandel, who suggests that we should verify what the rationale behind a legal construct actually is, when interpreting such construct.[210]

I elaborate in Part II.B.2 on the fact that the ultimate purpose behind DPLs was to protect individuals against a *risk of harm* that may be triggered by organizations collecting, using and disclosing their personal information.[211] But this *risk of harm* triggered by the handling of personal information is a function of several variables, such as: situation-specific circumstances, the intentions of the parties involved, the kind of information being sought after and the way it is processed.[212] For a full contextual approach, all of these elements would need to be taken into consideration in an integrated manner.

A contextual approach may pave the way for a more flexible framework necessary to adequately address the *kind* of harm resulting from the processing of personal information.[213] However, before elaborating on this full contextual approach, we first need to understand what kind of so-called "harms" DPLs were looking to address in the first place. My analysis should

---

[208] *Id.* at 88.

[209] VINCENT GAUTRAIS & PIERRE TRUDEL, CIRCULATION DES RENSEIGNEMENTS PERSONNELS ET WEB 2.0 at 48 (Montréal: Éditions Thémis, 2010).

[210] Gregory N. Mandel, *History Lessons for a General Theory of Law, Science & Technology*, 8 MINN. J. L. SCI. & TECH. 551, 556 (2007) ("[A] decision-maker must consider the rationale for the existing legal categories in the first instance, and then determine whether that rationale applies to the new technology. Legal categories (such as common carrier[s]) are only that—legal constructs. Such constructs may need to be revised in the face of technological change.").

[211] *See infra* Part II.B.2 (indicating that the ultimate purpose of DPLs is the protection of individuals' personal information from harm).

[212] Other relevant variables may include the historical context, the particular type of technology at stake, the political environment, the nature of the information within a given context, the vulnerability of the individual, the long term as well as the short-term impact on the individual affected, on what terms the information is shared, the terms of further dissemination, the purpose of disclosure, the expectations of the individual, the identity of the recipient, whether the recipient has an interest in knowing the information disclosed, etc.

[213] *See* OPCC, Privacy Comm'r of Can. Position, *supra* note 200 (discussing the fact that in Canada, the OPCC has been using a "total context approach" to interpret the notion of "personal information" in certain situations, claiming that doing so, it has been able to take into account the broader and more important context of the collection, use and disclosure of information).

therefore be viewed as a first step towards a contextual model. While a contextual approach may bring more subjectivity as to which information is in fact "identifiable" (different organizations may take a different approach and ultimately only courts or privacy commissioners would be in a position to qualify information) my approach can be used as a guide for organizations that handle personal information. More specifically, I propose a framework under which certain additional criteria (which specifically pertain to the information) will provide guidance as to which data may create a risk of harm for individuals. I believe that such interpretation, which takes into account the purposes behind DPLs, will leave more room for interpretation.

2. Determining Risk of Harm as Purpose Behind the Protection of Personal Information

In proposing a new interpretation of *personal information*, which will address the challenges brought on by the application of DPLs in the context of new technologies and the Information Age, the idea is to aim for a "level of generality [which] corresponds with the highest level goal that the lawmakers [initially] wish to achieve.[214]

Privacy debates have quite naturally focused on information and on constraining its use and dissemination.[215] As a matter of fact, when referring to DPLs, "[c]ontemporary privacy scholarship links data collection with privacy invasion so frequently that this assumption has become second nature to many scholars."[216] We usually refer to "privacy laws," when we are in fact referring to DPLs. It has been raised that "[t]he connection between the collection of personal data and personal privacy is straightforward: the more personal data that websites collect, store, and use, the less privacy that data subjects have."[217] Van den Hoven and Vermaas state that, "No data, no

[214] For example, "[h]igh level goals, such as, *preserv[ing*] *human life or improve[ing] economic efficiency* are relatively immune to waves of technological change." Moses, *supra* note 184, at 273–76 (emphasis added).

[215] *See* Stan Karas, *Privacy, Identity, Databases*, 52 AM. U. L. REV. 393, 403–05 (discussing the tendency of arguments regarding the issue of privacy to be focused on precluding data collection).

[216] *Id.* at 403.

[217] Steven Hetcher, *Changing the Social Meaning of Privacy in Cyberspace*, 15 HARV. J.L. & TECH. 149, 158 n.29 (2001).

need for data protection; no personal information, no need for informational privacy."[218]

Privacy and data protection are two fields that definitely overlap.[219] DPLs address personal privacy, as it relates to personal-data record keeping.[220] But privacy and data protection are not one and the same. In Europe, the Article 29 Working Party has reiterated the fact that the right to the protection of personal data is "separate and different from the right to private life."[221] The Article 29 Working Party also makes a distinction between the "violation of human dignity" and "data protection rights" which may take place with some applications of RFID technology.[222] This implies that they may therefore potentially be two separate types of rights.

While "privacy" includes various aspects of "data protection," I argue that data protection is in fact broader than privacy.[223] As early as the 1970s, at the time that DPLs were being discussed and adopted, certain committees which were in charge of analysing these issues, were clear on the fact that the notion of data protection is broader than privacy, as illustrated by this exert from the U.K. Lindop Report: "we believe that data protection goes further than the protection of privacy in its narrowest sense: it serves to protect many interests of the data subject, of which his privacy is only one."[224] The Article 29 Working Party also raises the fact that the data protection concept is much broader that the right to privacy:

---

[218] Jeroen van den Hoven & Pieter E. Vermaas, *Nano-Technology and Privacy: On Continuous Surveillance Outside the Panopticon,* 32 J. MED. & PHIL. 283, 292.

[219] LINDOP, *supra* note 12, at 9 ("The Younger Committee has to deal with the whole field of privacy. Our task has been to deal with that of data protection. In fact, the two fields overlap, and the area of overlap can be called 'information privacy' or, better, 'data privacy'. It is an important area . . . . But it is not by itself the whole field of data protection, and we have had to consider some matters which do not directly raise questions of privacy.").

[220] *See* WILLIS WARE, CHAIRMAN, SEC'Y'S ADVISORY COMM. ON AUTOMATED PERS. DATA SYS., DEP'T OF HEALTH, EDU., AND WELFARE at § III (1973), *available at* http://www.epic.org/privacy/hew1973report.

[221] Article 29 Working Party Opinion 4/2007, *supra* note 31, at 7.

[222] *See* Article 29 Working Party Working Doc. on RFID Tech., *supra* note 75, at 2 ("On the data protection front, Working Party 29 [] is concerned about the possibility for some applications of RFID technology to violate human dignity as well as data protection rights.").

[223] *Contra* GRATTON, *supra* note 15, at 204.

[224] LINDOP, *supra* note 12, at 204.

> On the one hand, it has to be considered that the concept of private and family life is a wide one, as the European Court on Human Rights has made clear. On the other hand, the rules on protection of personal data go beyond the protection of the broad concept of the right to respect for private and family life. . . . This is consistent with the terms of Article 1.1, aimed at protecting "the fundamental rights and freedoms of natural persons, and *in particular* [but not exclusively] their right to privacy."[225]

Various lawmakers and experts mandated to analyze data protection issues in the 1970s were in fact initially struggling with the distinction (privacy vs. data protection) since it was difficult to deal with the concept of privacy, as applied to records and databases.[226]

Although DPLs or transnational policy instruments pertaining to the protection of personal information that have been adopted since the early 1980s (OECD Guidelines, Convention 108, Directive 95/46/EC and the APEC privacy framework) all claim to protect the privacy of individuals,[227] they underline a much broader purpose: the one of protecting individuals from the risk of harm resulting from the collection, use and disclosure of their personal information.[228]

---

[225] Article 29 Working Party Opinion 4/2007, *supra* note 31, at 7 (footnotes omitted).

[226] *See* WARE, *supra* note 220 ("Dictionary definitions of privacy uniformly speak in terms of seclusion, secrecy, and withdrawal from public view. They all denote a quality that is not inherent in most record-keeping systems. Many records made about people are public, available to anyone to see and use. Other records, though not public in the sense that anyone may see or use them, are made for purposes that would be defeated if the data they contain were treated as absolutely secluded, secret, or private. Records about people are made to fulfill purposes that are shared by the institution maintaining them and the people to whom they pertain. Notable exceptions are intelligence records maintained for criminal investigation, national security, or other purposes. Use of a record about someone requires that its contents be accessible to at least one other person-and usually many other persons. Once we recognize these characteristics of records, we must formulate a concept of privacy that is consistent with records.").

[227] *See OECD Guidelines*, *supra* note 10, at Preface; Convention 108, *supra* note 9, at Preamble; Directive 95/46, supra note 7*,* at Whereas (1), (2), (10); APEC, *supra* note 11.

[228] *See* APEC, *supra* note 11, at 6 ("The APEC Privacy Framework applies to persons or organizations in the public and private sectors who control the collection, holding, processing, use, transfer or disclosure of personal information. Individual economies' definitions of personal information controller may vary. However, APEC economies agree that for the purposes of this Framework, where a person or organization instructs another person or organization to collect, hold, use, process, transfer or disclose personal

While documents from the 1970s leading to the elaboration of the FIPs and the adoption of DPLs mention the fact that protection of the privacy of individuals was a central element, they also make reference and discuss the broader notion of *risk of harm* in great length.[229] For instance, in the U.S., the Secretary's Advisory Committee on Automated Personal Data Systems[230] "was asked to analyze and make recommendations about: [the] [h]armful consequences that may result from using automated personal data systems; [s]afeguards that might protect against these potentially harmful consequences; [and] [m]easures that might afford redress for any harmful consequences."[231]

In Europe, working documents from the 1970s, leading to the elaboration of Convention 108, also include this notion of *risk of harm*.[232] The Resolutions (73) 22 and (74) 29 refer to electronic data processing which can be "harmful" to individuals,[233] to information that "may cause serious damage,"[234] that "may lead to unfair discrimination,"[235] to "unreasonably long retention of data that could be harmful,"[236] to "retention of information [that],

information on its behalf, the instructing person or organization is the personal information controller and is responsible for ensuring compliance with the Principles.").

[229] *Id.* at 11.

[230] The Secretary's Advisory Committee was established by former Secretary of Health, Education, and Welfare Elliot L. Richardson in response to growing concern about the *harmful consequences* that may result from uncontrolled application of computer and telecommunications technology to the collection, storage, and use of data about individual citizens. WARE, *supra* note 220, at Transmittal Letter to Secretary.

[231] *Id.* at Preface.

[232] OECD, Report on the Cross-Border Enforcement of Privacy Laws 6 (2006).

[233] EC Resolution (73) 22, *supra* note 12, at para. 12; EC Resolution (74) 29, *supra* note 12, at para. 24.

[234] EC Resolution (73) 22, *supra* note 12, at para. 18; *see also* EC Resolution (74) 29, *supra* note 12, at para. 18 ("It has been emphasized that the processing of sensitive information should be governed by special rules in view of the damage which individuals might suffer in case of misuse.").

[235] EC Resolution (73) 22, *supra* note 12, at para. 19; *see also* EC Resolution (74) 29, *supra* note 12, at para. 24 ("A further reason for imposing certain restrictions on the exercise of the right to know the information may be that this right may in turn degenerate into a source of unjust discrimination and be harmful to individuals . . . .").

[236] EC Resolution (73) 22, *supra* note 12, at para. 24; *see also* EC Resolution (74) 29, *supra* note 12, at para. 21 ("The first paragraph of this principle deals with the time-limits for keeping and using the information. In the public sector, just as in the private sector, individuals have a legitimate interest in seeing certain kinds of information concerning them, particularly that which is harmful to them, wiped off or rendered inoperative after a certain time has

even if not intended for use, presents a certain risk (for example, in case of accidental leaks)."[237] Resolutions (73) 22 and (74) 29 also mention that in the context of making "exceptions in the interests of science and of historiography," these exceptions had to "be reconciled with the interests which citizens have against the preservation of data harmful to them."[238] They suggested that "processing of sensitive information should be governed by special rules in view of the damage which individuals might suffer in case of misuse."[239] These concerns which FIPs were meant to address refer to a *risk of harm* to an individual that may take place if certain information is inappropriately used or disclosed.[240]

More recently, under DPLs, organizations are prohibited to disclose personal information to third parties.[241] Certain Canadian DPLs mention the potential harm that may take place in such an event, for instance if the disclosure "may *seriously harm* that third person."[242] Under the B.C. DPL, an organization may disclose, without the consent of the individual, personal information for research purposes, including statistical research, only if linkage of the personal information to other data "is not *harmful* to the individuals identified by the personal information."[243]

---

passed.").

[237] EC Resolution (73) 22, *supra* note 12, at para. 25; *see also* EC Resolution (74) 29, *supra* note 12, at para. 21(suggesting that not all information should be subject to time limits because some information continues to be valid, such as names, birth dates, etc.).

[238] EC Resolution (74) 29, *supra* note 12, at para. 22; *see also* EC Resolution (73) 22, *supra* note 12, at para. 36 (outlining parties who could obtain access to personal information on a limited basis).

[239] EC Resolution (74) 29, *supra* note 12, at para. 18; *see also* EC Resolution (73) 22, *supra* note 12, at para. 19 ("Examples of information concerning a person's intimate private life are: information about his behaviour at home, his sexual life, his opinions etc. An example of information which may lead to unfair discrimination is that about his state of health, or his past criminal record. The text of this principle makes a distinction between the keeping and the release of this kind of information.").

[240] EC Resolution (74) 29, *supra* note 12, at para. 24; EC Resolution (73) 22, *supra* note 12, at para. 12.

[241] *See generally* DATA PROTECTION COMM'R, DATA PROTECTION ACTS 1988 AND 2003: INFORMAL CONSOLIDATION 8–10 (2009) (outlining Ireland's policies in the Data Protection Act prohibiting third party disclosure and the exceptions that may apply).

[242] An Act Respecting the Protection of Personal Information in the Private Sector, R.S.Q. 1993, c. P-39.1, s. 2 para. 40 (Can.) (emphasis added).

[243] Personal Information Protection Act, S.B.C. 2003, c. 63, 21(1)(c) (emphasis

In recent years, the few Canadian security breach notification laws that have been introduced or adopted stipulate that organizations notify affected individuals when security breaches occur, where personal information has been compromised and "where a reasonable person would consider that there exists *a real risk of significant harm to an individual*."[244] It is interesting to note that when legislators are attempting to limit the scope of DPLs, they are inclined to focus on the notion of "risk of harm," probably because this is in fact the main goal at the heart of DPLs and its ultimate purpose.[245] In Europe, certain security rules notably state that in deciding what level of security is appropriate, organizations handling data must assess the nature of the personal data in question, and, interestingly, the *harm* that might result from the unauthorized use, disclosure or loss of the data.[246]

Some "regulators are also looking at the '*preventing harm*' principle as a valid way forward."[247] In 2007, the U.K. Information Commissioner published its data protection strategy which emphasized the fact that the U.K. regulator's actions will give priority to tackling situations where there is a real

added).

[244] Personal Information Protection Act, S.A. 2003, c. P-6.5, art. 34.1(1) (Can.) (emphasis added). This section requires an organization to notify individuals in circumstances where the "*real risk of significant harm*" to an individual as a result of the loss or unauthorized access or disclosure is obvious and immediate. *See* Safeguarding Canadians' Personal Information Act, H.C. C-12, 41st Parl. §§ 10.01–02 (2011) (Can.) (emphasis added); *see also* An Act to Amend the Personal Information Protection and Electronic Documents Act, H.C. C-475, 41st Parl. §§ 10.01 (2013) (similar proposed legislation).

[245] *See, e.g.*, INFO. COMM'R'S OFFICE, THE GUIDE TO DATA PROTECTION 82–92 (2009) (outlining the aims of the U.K.'s Data Protection Act to prevent risks of harm to individuals).

[246] According to such rules, "[I]t is reasonable for organizations to weigh up the costs of security measures against the other factors so that if the *risks* of security breaches are low, and the *likely harm* that would arise is trivial or minor, then a data controller might justifiably decide not to invest a great deal of money in state-of-the-art security measures. Conversely, if the risks of security breaches (or attempted breaches) are high, and/or the *likely harm* to an individual would be high, then the organization should invest in robust security measures . . . ." *Security Measures for Personal Data: A Guide to the New Data Protection Rules,* DATA PROTECTION COMM'R (2001), http://www.dataprotection.ie/ViewDoc.asp?DocID=39&UserLang=GA [hereinafter *Security Measures*] (emphasis added).

[247] Peter Fleischer, *Global Privacy Standards Should Focus on Preventing Harm to Consumers*, GOOGLE PUB. POL'Y BLOG (Nov. 14, 2007), http://googlepublicpolicy.blogspot.com/2007/11/global-privacy-standards-should-focus.html (emphasis added).

likelihood of serious harm.[248] In 2009, RAND Corporation was mandated by the U.K. Information Commissioner to evaluate Directive 95/46/EC in light of recent technological advancements.[249] One of its main recommendations was to enforce the regulatory framework only in cases "where significant risk of harm or actual harm exists."[250] Sweden, following a review, has recently adopted a set of regulations using a risk-based approach toward the misuse of personal data.[251]

Recently, this notion of *harm* has been included in several national and transnational policy instruments.[252] The Treasury Board of Canada Secretariat recently released a guidance document to assist organizations in determining whether a contract involving personal information would result in harm, in which it proposes an invasion-of-privacy test which should take into account the sensitivity of the information, expectations of the individuals, and probability and gravity of injury.[253] The *risk of harm* resulting from this test is then categorized as being no risk, low risk, medium risk, or high risk.[254] In 2005, APEC confirmed having developed a Framework on information privacy protection in recognition of the importance of developing appropriate privacy protections for personal information; particularly from the "harmful" consequences of unwanted intrusions and the misuse of personal information, which includes a "Preventing Harm" principle.[255]

Prompted by concern over offline data privacy threats and the increasing convergence of online and offline data systems, the

---

[248] By acknowledging that DPLs are over-reaching, the U.K. has been proposing to enforce them based on their original goal and purpose: "Being a strategic regulator means that, in so far as we have a choice, we have to be selective with our interventions. We will therefore apply our limited resources in ways that deliver the maximum return in terms of a sustained reduction in data protection risk. That is the risk of harm through improper use of personal information. There are priorities we have to set. We need to focus most attention on situations where there is a real likelihood of serious harm." INFO. COMM'R'S OFFICE, DATA PROTECTION STRATEGY: CONSULTATION DRAFT 5 (2007).

[249] ROBINSON ET AL., *supra* note 55, at ii.

[250] *Id.* at 41.

[251] *Id.*

[252] *See infra* text accompanying notes 253–58 (discussing national and transnational policy instruments' inclusion of the notion of harm).

[253] *Guidance Document: Taking Privacy into Account Before Making Contracting Decisions*, TREASURY BD. CAN. SECRETARIAT (Jul. 8, 2010), https://www.tbs-sct.gc.ca/atip-aiprp/tpa-pcp/tpa-pcp00-eng.asp.

[254] *Id.*

[255] APEC, *supra* note 11, at 11.

FTC has, since 2000, decided on a privacy approach evolved to include a focus on specific consumer harms as the primary means of addressing consumer privacy issues.[256] The FTC admits that this harm-based approach does have its limitations.[257] However, the FTC's harm-based model was meant to target practices that caused or were likely to cause physical or economic harm, or unwarranted intrusions into consumers' daily lives, while omitting the more costly notice-and-choice requirements characteristic of DPL's.[258]

I discuss the notion of "risk" of harm, instead of simply referring to the notion of "harm." This is because in evaluating whether certain pieces of information qualify as *personal information* or can "identify" an individual, I maintain that we first need to assess if a certain data handling activity may be harmful to an individual. Since this harm is only potential, I find it proper to refer to a "risk" of harm.[259] In the next section, I will propose criteria in order to assist in determining whether certain information should be considered as *personal information*, depending on whether there is the presence of a "risk of harm" and if so, the extent of such risk or harm.[260]

III. IMPLEMENTING THE RISK OF HARM APPROACH TO THE NOTION OF PERSONAL INFORMATION

The notion of "identifiable" individual, has been the subject of much debate and controversy and is interpreted differently

---

256 *See* FTC 2012 Recommendations, *supra* note 76, at 2 (explaining that the FTC has developed various models of consumer protection, each with its own flaws).

257 *Id.*

258 FTC Proposed Framework, *supra* note 153, at 9; *see also* Chairman Timothy J. Muris, FTC, Remarks At the Privacy 2001 Conference (Oct. 4, 2001), *available at* http://www.ftc.gov/speeches/muris/privisp1002.shtm. Chairman Muris then identified various harms caused by the misuse of consumer data—for example, risks to physical security from stalking; economic injury resulting from identity theft; and commercial intrusions into daily life by unwanted solicitations. *Id.*

259 This is also because in certain situations, the harm will be subjective in nature, in which case whether the harm does in fact takes place (for instance whether an individual feels embarrassed following a disclosure of his or her information) will depend on each individual and their personal sensitivities. Whether information available will be used against an individual in order to inflict an objective harm (such as discriminating this individual for certain employment for instance) is only "a risk" as it may or may not take place in the future.

260 *See infra* discussion Part III.

between jurisdictions (and sometimes even within the same jurisdiction).[261] The notion of "identifiability" is therefore a complex issue and also a subjective one. Moreover, while the European system (contrary to Canada) does in fact have a test to provide guidance on what should be taken into account when determining what counts as "identifiable" information with recital 26, the handful of cases that address the interpretation of the Directive 95/46/EC's Article 2 (a) (definition of *personal data*) in conjunction with recital 26 *"all the means likely reasonably to be used*" reveal that European courts have approached this issue in a number of ways, leading to contradictory and confusing conclusions.[262] To add to all the uncertainty surrounding this notion of the "identifiable individual," the Article 29 Working Party also maintains that the European test is a dynamic one, and that we should consider the state of the art in technology at the time of the processing, and the possibilities for development during the period for which the data will be processed.[263]

Various authors are proposing potential guidance on some of the issues raised above. For example, the work performed by Bercic and George is examining how knowledge of relational database design principles can help to understand what is and what is not *personal data*.[264] Lundevall-Unger and Tranvik

[261] *See infra* Part II.A.3 which elaborates on this issue.

[262] *See* Lundevall-Unger & Tranvik, *supra* note 86, at 61–65 (featuring a detailed analysis of these court decisions).

[263] Article 29 Data Protection Working Party Opinion 4/2007, *supra* note 31, at 15 ("Identification may not be possible today with all the means likely reasonably to be used today. If the data are intended to be stored for one month, identification may not be anticipated to be possible during the 'lifetime' of the information, and they should not be considered as personal data. However, it they are intended to be kept for 10 years, the controller should consider the possibility of identification that may occur also in the ninth year of their lifetime, and which may make them personal data at that moment. The system should be able to adapt to these developments as they happen, and to incorporate then the appropriate technical and organizational measures in due course.").

[264] Bercic & George, *supra* note 42, at 235–236, 238 (These authors suggest that, in relational database theory, there would be a record structurally consisting of two parts: (i) the record identifier (primary key) and (ii) data related to it. The identifier is usually unique or full, which means that an individual is identified uniquely (e.g., name and surname, often together with added information such as residence) or a unique number such as one provided by the government. They suggest to make a distinction between "explicit" identifier (name, surname, and residence if needed) and "implicit" identifier (such as a social security number or national ID number). They suggest to also make a distinction between "full" and "partial" identifiers when qualifying data

propose a different and practical method for deciding the legal status of IP addresses (with regard to the concept of *personal data*); a test that can apply to other types of data as well.[265] Briefly, their proposed method consists of two steps: (i) first, a legality test under which illegal means of linking "names and faces" to IP addresses are not taken into account when assessing whether or not IP addresses are personal data (only legal methods of identification should form the basis of these decisions); and (ii) second, a "likely reasonable" test. More specifically, under this second step, the authors suggest that the question of *personal data* should be resolved by assessing the costs (in terms of time, money, expertise, etc.) associated with employing legal methods of identification.[266] In a more recent article, Professors Schwartz and Solove argue that the current approaches to PII are flawed and propose a new approach called "PII 2.0," which accounts for PII's malleability.[267] Based upon a standard rather than a rule, PII 2.0 would be based upon a continuum of "risk of identification" and would regulate information that relates to either an "identified" or "identifiable" individual (making a distinction between the two categories), and they establish different requirements for each category.[268]

My contribution in providing guidance on this notion of "identifiability" in the context of the Internet and related technologies has to do with interpreting the notion of "identifiable individual" depending on the purpose behind the data handling activity regulated by the DPLs. Regulating the "disclosure" and the "use" of personal information serve very different ends; protecting against subjective harm in the case of the former and objective harm for the latter.[269]

Calo articulates the view that "the vast majority of privacy harms fall into just two categories—one subjective, and the other

and determining whether information is "personal.").

[265] Lundevall-Unger & Tranvik, *supra* note 86, at 53.

[266] *Id.* at 58 ("If the costs of employing these methods are exceedingly high, then the likelihood of identifying who is using which IP address is low. Hence, IP addresses are not personal data. But if the costs are more modest, then the chance of identifying individual Internet users increases, and we should conclude that IP addresses are indeed personal data.").

[267] Schwartz & Solove, *supra* note 8, at 1865.

[268] *Id.* at 1877.

[269] M. Ryan Calo, *The Boundaries of Privacy Harm*, 86 IND. L.J.1131, 1131 (2011).

objective."[270] Although he includes the "objective" type of harm resulting from the use of personal information under the notion of "privacy," we are both undoubtedly in agreement that there are clearly two distinct types of harms that result from the collection, use and disclosure of personal information.[271] Esther Dyson also points out that it is possible to distinguish between objective harms resulting from the use of personal information (ex. the denial of a service, fraud) from subjective privacy harms (ex. the knowledge of certain intimate details pertaining to an individual by a second or third person which is experienced as an injury).[272]

Evidence that two types of harms (subjective vs. objective) were targeted by DPLs can be found in documents prepared in the context of the elaboration of the FIPs in the 1970s.[273] The following excerpt illustrates that the subjective/objective distinction was at the very heart of the original goals of the FIPs. Resolution (74) 29 mentions that: "[e]specially when electronic data banks process information relating to the intimate private life of individuals or when the processing of information might lead to unfair discrimination, their existence must have been provided for by law . . . ."[274]

In this resolution, when referring to the "intimate private life of individuals," the authors are in fact alluding to a more subjective kind of harm.[275] Conversely, the reference to an "unfair discrimination" relates to a more objective harm.[276] Around the same period, the Lindop Report (U.K., 1978) addressed the issue of privacy in relation to data subjects, mentioning that: "'[p]rivacy' means, in relation to any data subject, his interest to determine for himself what data relating to him should be known to what other persons, and upon what terms as to the use which those persons may make of those data."[277]

This assertion also refers to a subjective kind of harm when it states: "what data relating to him should be known to what other persons," and it refers to a more objective kind of harm when it

[270] *Id*. at 1133.
[271] *Id*.
[272] ROBINSON ET AL., *supra* note 55, at 3.
[273] EC Resolution (74) 29, *supra* note 12, at principle 3 of Annex.
[274] *Id*.
[275] *Id*.
[276] *Id*.
[277] LINDOP, *supra* note 12, at 204.

states: "and upon what terms as to the use which those persons may make of those data."[278] More recently, the U.K. Information Commissioner published a report on its data protection strategy in 2007, in which it is emphasizing on the need to judge the seriousness of the risks of individual harm which can present itself in different ways, also making a distinction between objective vs. subjective harms: "Sometimes it will be tangible and quantifiable, for example the loss of a job" (which implies an objective kind of harm), while: "At other times it will be less defined, for example damage to personal relationships and social standing arising from disclosure of financial circumstances" which implicitly refers to a more subjective kind of harm.[279]

In the face of uncertainty presented by new types of data,[280] the Article 29 Working Party recently commented on information generated by RFID tags, and when this data should be considered as relating to an individual: "data relates to an individual if it 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."[281]

When the Article 29 Working Party suggests that the data "refers to the identity, characteristics or behaviour of an individual," it implicitly refers to a risk of harm that is of a more subjective nature.[282] On the other hand, when referring to the information "used to determine or influence the way in which that person is treated or evaluated," the Article 29 Working Party refers to a more objective kind of harm.[283]

Solove has put together a "privacy taxonomy" in order to assist the legal system in grappling with the concept of privacy. He believes that "[s]ince the goal of the law is to have privacy protections that best prevent and redress particular problems, we need to first understand the problems in order to evaluate the effectiveness of the protections."[284] In devising a taxonomy,

---

[278] *Id*.

[279] INFO. COMM'R'S OFFICE, *supra* note 245, at 7.

[280] *See supra* Part II.A.3 (elaborating on the issue of the face of uncertainty presented by new types of data).

[281] Article 29 Working Party Opinion 4/2007, *supra* note 31, at 10 (footnote omitted).

[282] *See supra* Part III.A (elaborating on this kind of harm).

[283] *See* discussion *infra* Part III.B (elaborating on this kind of harm).

[284] Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 485 (2006).

although there are many different ways to go about carving up the landscape, he has decided to focus on the activities that invade privacy and create problems.[285] This taxonomy is comprised of four basic groups of harmful activities: information collection, information processing, information dissemination, and invasion.[286] It is interesting to note that three out of the four groups relate to data handling activities, which illustrates that this is a good starting point when attempting to determine the type of harm that may take place. What may also be implied is that each data handling activity has its own set of problems that DPLs were looking to address.

If we are looking to identify the ultimate purpose of DPLs (in order to guide us in determining which kind of data should qualify as *personal information*), we need to be sensitive towards the particular types of data handling activities and whether the underlying *risk of harm* is subjective or objective.

I will now demonstrate how this harm is different depending on the data handling activity at stake, and propose a decision-tree which may be useful when having to decide whether certain information should qualify as *personal information* or whether or not a certain data handling activity may be harmful.

A. Subjective Harm Associated with Personal Information

The first type of harm that may be triggered by DPLs is of a subjective nature and usually, is linked with two types of data handling activities: the collection of personal information and the disclosure of this information, as detailed below. It is subjective in nature, as it often relates to an emotional or psychological type of harm.[287]

Warren and Brandeis in their famous article about privacy and

---

[285] *Id.*

[286] *Id.* at 488.

[287] In 1972, the Scottish Justice Committee stated that: "the notion of privacy has a substantial emotive content in that many of the things which we feel the need to preserve from the curiosity of our fellows are feelings, beliefs or matters of conduct which are themselves irrational." Justice Committee on privacy, *"Privacy and the Law*" at 5, para. 18, discussed in Home Office, Lord Chancellor's Office, Scottish Office (Chairman The Rt. Hon, Kenneth Younger), REPORT OF THE COMM. ON PRIVACY, presented to Parliament by the Secretary of State for the Home Department, the Lord High Chancellor and the Secretary of State for Scotland by Command of Her Majesty, July 1972, at 17, para. 47 [hereinafter REPORT OF THE COMM. ON PRIVACY].

the right to be let alone, referred to the disclosure of private facts in new press, contending that privacy involved "injury to the feelings."[288] William L. Prosser ("Prosser") discusses how the common law recognizes a tort of privacy invasion in cases where there has been a "[p]ublic disclosure of embarrassing private facts about the plaintiff."[289] According to Calo, the subjective category of privacy harm (which is included in the activity of collecting and disclosing personal information) is "the [unwanted] perception of . . . observation, broadly defined."[290] Observation may include the activity of collecting personal information but this also includes the disclosure of personal information.[291] Calo suggests that many of the harms we associate with a person seeing us, such as "embarrassment, chilling effects or a loss of solitude, flow from the mere belief that one is being observed."[292] Gavison refers to an observation with an "inhibitive effect on most individuals that makes them more formal and uneasy."[293] Recently, in *Jones v. Tsige*,[294] the Court of Appeal for Ontario hinted that there was a subjective component to an invasion of privacy, assimilated to "distress, humiliation or anguish" (which is therefore subjective in nature).[295]

The first data handling activity typically regulated by DPLs involves the collection of personal information. Although the term "collection" is not specifically defined in the Canadian and European DPLs discussed herein, it usually relates to the activity or the means by which personal information is gathered or obtained.[296] The *risk of harm* resulting from the collection of personal information is usually of a subjective nature: it can be

---

288 *See* Samuel Warren & Louis Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 197–98 (1890) ("[O]ur system . . . does not afford a remedy even for mental suffering which results from mere contumely and insult . . . .").

289 William L. Prosser, *Privacy*, 48 CALIF. L. REV. 383, 389 (1960).

290 Calo, *supra* note 269, at 14.

291 *Id.* (adding that unwanted observation includes knowing a person's "preferences, associations and whereabouts.").

292 *Id.* at 16 (footnote omitted).

293 Ruth Gavison, *Privacy and the Limits of Law*, 89 YALE L.J. 421, 447 (1980).

294 Jones v. Tsige, 2012 ONCA 32 (Can. Ont. C.A.).

295 *Id.* at para. 71. The court also mentioned that "proof of harm to a recognized economic interest is not an element of the cause of action," therefore implying that a subjective kind of harm may take place upon an invasion of privacy, even in the absence of an objective (financial) harm. *Id.*

296 *See* GAUTRAIS & TRUDEL, *supra* note 209, at 112–25 (examining the notion of "collection" in the Quebec public sector DPL and under PIPEDA).

assimilated with a psychological type of harm, similar to a feeling of being observed (or under surveillance).[297] I discuss elsewhere why regulating the activity of collecting personal information is challenged in light of the Information Age, and why DPLs were not specifically aiming at addressing this kind of harm.[298] Given the volume of personal information readily available today,[299] we should be focusing on the type of harm which can take place through other data handling activities, namely the types of harm triggered by the use or the disclosure of *personal information*. As a matter of fact, if an organization collects personal information without ever actually "using" it (for instance to take a decision which will impact the individual) and adequately protects the information against any potential disclosure (or a disclosure of the information would not be harmful to the individual), then the risk of harm at the "collection" level is either minimal, or it should be regulated by tools other than DPLs.[300] Therefore, I maintain that upon information being collected, the analysis which should take place in order to determine whether the information collected is *personal*, is whether the information collected may create a *risk of harm* upon being disclosed (for instance in the context of a security breach) or upon being used, in which case it will qualify as *personal information*. This translates in data *collected* only having to be disclosed to individuals (and their consent having to be obtained) if the data creates a risk of harm at the "disclosure" or "use" levels.[301]

A second activity that is regulated by DPLs is the disclosure (or

---

[297] *See* GRATTON, *supra* note 15, at 229 ("We can separate the types of harm resulting from the activity of collecting personal information into two categories. The first category relates to a feeling of being under surveillance or observation. The second category relates to an individual who is not aware that certain information is being collected about him or her, in which case the harm is more of a dignitary harm.").

[298] *See id.* at 247.

[299] *See id.* at 21.

[300] *See id.* at 247. Although the collection may increase the risk of harm resulting from the disclosure or use of the personal information, the type of harm that the collection in itself usually triggers is more likely to be associated with some type of psychological harm (such as the "feeling of being under surveillance") or some type of dignitary harm. Since DPLs were not meant to address the first kind of harm (feeling of being under surveillance), and that they have proven to be inadequate in addressing dignitary harm (through the inefficient notice and choice model) I argue that we should focus on the risks of harm which may take place at the "disclosure" and "use" levels.

[301] *See* discussion *infra* Parts III.A.1, III.B.1 (discussing the tests to use at the disclosure and use levels).

dissemination) of personal information. The notion of disclosure usually refers to the giving of information, the making available of information, the exchange of information or the sharing of knowledge.[302] Solove refers to this activity as "information dissemination," where the data holders transfer the information to others or release the information, resulting in the data moving further away from the control of the individual.[303] In many situations, the information disclosed by a party may have in fact been already available to a certain extent, an activity that I refer to as making information "increasingly available," this activity being included as a *disclosure* for the purpose of the present analysis.[304]

In his *Taxonomy of Privacy,* Solove suggests that the kind of harm resulting from the dissemination of information is more often than not of a psychological nature.[305] He mentions that an "exposure" involves divulging certain physical and emotional attributes about a person that people view as deeply primordial; this often creates "embarrassment and humiliation [] [such as g]rief, suffering, trauma . . . ."[306] According to Solove, we have developed social practices to conceal aspects of life that we find animal-like or disgusting (for example: nudity or going to the bathroom).[307] Individuals being "exposed" could therefore experience a severe and sometimes "debilitating humiliation and loss of self-esteem."[308]

1. "Identifying" Taking Into Account the Overall Sensitivity of Information

In order to be harmful to an individual, a disclosure of personal information would therefore have to create some type of humiliation or embarrassment, as discussed in the previous section. Given that the *risk of harm* upon a disclosure is highly

[302] *See* GAUTRAIS & TRUDEL, *supra* note 209, at 96–97 (discussing the meaning of the verb "communicating"); *see also* BLACK'S LAW DICTIONARY 316 (8th ed. 2004).

[303] Solove, *supra* note 284, at 488.

[304] *See* discussion *infra* pp. 165–68 (discussing the increasing availability of data and how the results may not be as harmful as assumed).

[305] Solove, *supra* note 284, at 525.

[306] *Id.* at 533.

[307] *See* Anita L. Allen, *Lying to Protect Privacy*, 44 VILL. L. REV. 161, 177 (1999) ("Sex is an area in which we encounter our desires, prejudices and shame, and cloak these emotions in privacy.").

[308] Solove, *supra* note 284, at 535.

contextual and can be difficult to isolate, I propose to interpret the notion of "identifiable" in light of the overall sensitivity of information in question. More specifically, I propose additional criteria relating to the information which may be used when interpreting the notion of "identifiable" and which may be essential to the identification of this kind of harm: These additional criteria are the "intimate" nature of the information, and the extent of its "availability" to third parties or the public upon being disclosed.[309]

To trigger the feeling of humiliation or embarrassment upon being disclosed, the data usually needs to focus on something of an "intimate nature."[310] As mentioned earlier, Warren and Brandeis were specifically concerned with protecting information about "the private life, habits, acts, and relations of an individual"[311] and Prosser discussed the presence of a tort of privacy invasion in cases where there had been a "[p]ublic disclosure of embarrassing private facts."[312] In the late 1960s, the conclusions of the Nordic Conference on the Right of Privacy (1967) referred to the kind of harm resulting from an attack on the honour and reputation of an individual and the "disclosure of irrelevant embarrassing facts relating to his private life."[313] In Europe, Resolution 428 (1970) *containing a declaration on mass communication media and human rights* suggested that the right to privacy was the protection of one's "private, family and home life" which consisted, among other things, of the "non-revelation of irrelevant and embarrassing facts, unauthorized publication of private photographs, . . . [and the] protection from disclosure of information given or received by the individual confidentially."[314]

In Europe, Directive 95/46/EC has included at article 8

---

309 Nissenbaum, *supra* note 49, at 128 (noting that the proposed criteria are very close to what Nissenbaum prescribes when she discusses how the principle of restricting access to personal information usually focuses on data that is "intimate," "sensitive," or "confidential").

310 Trudel & Benyekhlef, *supra* note 54, at 5.

311 Warren & Brandeis, *supra* note 288, at 216.

312 Prosser, *supra* note 289, at 389.

313 REPORT OF THE COMM. ON PRIVACY, *supra* note 287, at 327 (including as a definition of privacy the conclusions from the 1967 Nordic Conference on the Right of Privacy, "The right of the individual to lead his own life protected against . . . the disclosure of irrelevant embarrassing facts relating to his private life . . . .").

314 Eur. Parl. Ass. Resolution 428 (1970) on Containing a Declaration on Mass Communication on Media and Human Rights of 23 January 1970 [hereinafter Eur. Parl. Ass. Resolution 428 (1970)].

categories of "sensitive" data, and acknowledge that certain types of personal data are more privacy sensitive and more likely to harm the data subject in cases of unauthorized processing.[315] These categories include data "revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life."[316] The categories of so-called inherently "sensitive" information are usually of an "intimate" nature.[317] Interestingly, these categories are similar to the categories or elements determined by Canadian courts as relating to the intimate or the private life of individuals.[318]

In Canada, in *Stevens v. SNF Maritime Metal Inc.*,[319] the Federal Court of Canada took the position that the individual had not put into evidence the fact that his personal information disclosed in breach of PIPEDA triggered a subjective harm, since the information at stake was not "deeply personal" or "intimate." In the recent case of *Jones v. Tsige*,[320] the Court of Appeal for Ontario illustrates that in the case of an invasion of privacy, the fact that the information disclosed is of "intimate" nature is crucial:

> A claim for intrusion upon seclusion will arise only for deliberate and significant invasions of personal privacy. . . . it is only intrusions into matters such as one's financial or health records, sexual practices and orientations, employment, diary or private correspondence that, viewed objectively on the reasonable person

315 Directive 95/46, supra note 7, at art. 8, para. 1.

316 Id.

317 *See* GRATTON, *supra* note 15, at 291 (elaborating on the fact that information of *intimate* nature would include medical and health information, information pertaining to one's family and personal life, information pertaining to love, sex and sexual orientation, religion, political and philosophical opinions, race and ethnicity, personal affiliations, financial information, private communications and location information).

318 *See* Pierre Trudel, *Privacy Protection on the Internet: Risk Management and Networked Normativity*, *in* REINVENTING DATA PROTECTION? 317, 322 (Serge Gutwirth et al., eds. 2009); *see also* Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, at Principle 4.3.4 (Can.) ("Although some information (for example, medical records and income records) is almost always considered to be sensitive, any information can be sensitive, depending on the context. For example, the names and addresses of subscribers to a newsmagazine would generally not be considered sensitive information. However, the names and addresses of subscribers to some special-interest magazines might be considered sensitive.").

319 Stevens v. SNF Maritime Metal Inc., 2010 FC 1137 (Can. Ont. Fed. Ct.).

320 Jones v. Tsige, 2012 ONCA 32 (Can. Ont. C.A.).

standard, can be described as highly offensive.[321]

In the U.S., there is no general DPL overseeing all commercial activities of organizations (such as there are in Canada and France) although so-called "sensitive" information is accorded special recognition through a series of sectoral privacy statutes.[322] More specifically, the particular categories of information most likely to require protection against disclosure to third parties are often information which would be considered as being of "intimate" nature: government records,[323] cable company records,[324] video rental records,[325] and personal health information.[326] Various U.S. states would also restrict the disclosure of particular forms of information, such as medical data and alcohol and drug abuse.[327]

If a given set of information is already in circulation or already available to the party receiving the information, then the risk of subjective harm that may be triggered by the disclosure of information is less substantial (in the sense that individuals will rarely be embarrassed nor humiliated following the disclosure of information already available).[328]

---

[321] *Id*. at para. 72.

[322] *See* Natasha Singer, *An American Quilt of Privacy Law, Incomplete*, N.Y. TIMES (Mar. 30, 2013), http://www.nytimes.com/2013/03/31/technology/in-privacy-laws-an-incomplete-american-quilt.html?pagewanted=all&_r=0 ("The American system involves a patchwork of federal and state privacy laws . . . . The European Union, on the other hand, has one blanket data protection directive . . . .").

[323] *See* The Privacy Act of 1974, 5 U.S.C. § 552a(e)(10) (2000) (prohibiting governmental agencies from disclosing information about an individual without his or her prior written consent).

[324] *See* The Cable Communications Policy Act of 1984, 47 U.S.C. §§ 551(b)–(c) (2000) (limiting the extent to which a cable service may collect or disclose PII about subscribers).

[325] The Video Privacy Protection Act of 1988, 18 U.S.C. § 2710(b)(1) (2013) creates civil liability for video stores that disclose PII about any customer and protects against unconstrained dissemination of video rental records.

[326] The Health Insurance Portability and Accountability Act of 1996, 42 U.S.C. §§ 1320d–1320d-8 (2000) protects the privacy of personal health information in transactions.

[327] The California Health and Safety Code prohibits the disclosure of HIV test results. CAL. HEALTH & SAFETY CODE § 120980 (2006). The New York Public Health Law permits the release of medical records of minors relating to sexually transmitted diseases and abortion upon written request, but prohibiting the disclosure to parents without consent. N.Y. PUB. HEALTH LAW § 17 (McKinney 1999). The Pennsylvania Drug and Alcohol Abuse Control Act prohibits the disclosure of all records prepared during alcohol or drug abuse treatment. PA. STAT. ANN. tit. 71, § 1690.108 (West 1990).

[328] In the U.S., in *Sipple v. Chronicle Publishing Co.*, newspapers disclosed

Some are claiming that changes with regards to how individuals view their privacy have recently taken place and contend that the social changes inherent to web 2.0 (with individuals voluntarily sharing their personal information) may perhaps reflect a changing mentality with regards to privacy.[329] As early as 1970, Resolution 428 *containing a declaration on mass communication media and human rights* suggested that individuals who "by their own actions, have encouraged indiscreet revelations about which they complain later on, cannot avail themselves of the right to privacy."[330] Certain European jurisdictions (such as France) provide for certain exclusions (no consent is required) for personal data rendered public by the individual concerned.[331] This provision implicitly acknowledges the fact that the disclosure of personal information, that was already made available or rendered public by the individual, may not be as harmful as the disclosure of information that has remained confidential.[332]

In the Information Age, with new technologies and the web, most information that is disclosed may have been previously available to a certain extent. Instead of data being "disclosed," we can therefore speak of data being "increasingly available." Solove suggests that in such situation "[o]ne must focus on the extent to which the information is made more accessible."[333]

In a 2009 finding, the Canadian OPCC allowed enrichment of phone book information with demographic information from Statistics Canada and refused to impute a consent

the fact that Oliver Sipple, who heroically saved President Ford from an assassination attempt, was homosexual. The court concluded that his sexuality was not private because it was already known in the gay community. 201 Cal. Rptr. 665 (Cal. Ct. App. 1984).

329 "[F]or example individuals willing to give up personal information for small gains such as by telling personal stories to become part of a trusted community of shared interests, and sharing content increasingly via user-friendly and accessible platforms such as YouTube and SNS." ROBINSON ET AL., *supra* note 55, at 15; L. Gordon Crovitz, *Privacy Isn't Everything on the Web*, WALL ST. J. (May 24, 2010), http://online.wsj.com/article/SB10001424052748704546304575260470054326304.html.

330 Eur. Parl. Ass. Resolution 428 (1970), *supra* note 314, at C(2).

331 Loi 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés [Law 78-16 of January 6, 1978 Relating to Data, Files and Freedom] at c. II, s. 2, art. 8 (II) (4), *available at* http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000886460.

332 *See id*. (implying that if the data was released by the subject it is no longer as harmful nor confidential).

333 Solove, *supra* note 284, at 540.

requirement.[334] The PIAC has shared its concern with this decision involving "publicly available" information enriched by data aggregators and data miners.[335] In France, a different outcome took place in a similar situation, when publicly available directory data was to be merged with other available information. France's Data Protection Authority, the CNIL, announced on September 23, 2011 that it had found the French provider of universal telephone directory services, *Pages Jaunes,* guilty of violating several provisions of the French DPL.[336] Pages Jaunes' web crawler function captured information contained on Facebook, Twitter and LinkedIn profiles of individuals having the same name as the individual being looked up in the directory service and "more complete profiles" were made available online without the requisite consent.[337] The CNIL's decision illustrates the concerns which can take place with the "availability" criteria.[338] More specifically, an organization, prior to disclosing information, must assess if the data to be disclosed has been mined, analyzed and whether the disclosure of the information will release additional information or increase the "knowledge"

[334] *PIPEDA Case Summary #2009-004: No Consent Required for Using Publicly Available Information Matched with Geographically Specific Demographic Statistics*, OFFICE OF THE PRIVACY COMM. OF CAN., https://www.priv.gc.ca/cf-dc/2009/2009_004_0109_e.asp (last visited Oct. 5, 2013).

[335] *2010 Consumer Privacy Consultations: Understanding Online Tracking, Profiling and Targeting*, PUB. INTEREST ADVOCACY CTR. (Mar. 15, 2010), http://www.piac.ca/privacy/piac_comments_to_privacy_commissioner_of_canada_on_behavioural_targeting (stating that in "the OPCC, in bestowing the title of 'publicly available' upon this type of personal information (directory information) and then refusing to require consent for the new use the information after its 'enrichment' with yet more personal information simply guts PIPEDA Principle 4.5. It ignores the general safeguards that the CRTC sought to uphold over the years in many decisions on directories. It allows an entire industry to be constructed with the express purpose of doing indirectly what PIPEDA forbids directly.").

[336] The CNIL did not fine Pages Jaunes, but published a detailed warning, listing each privacy violation that the CNIL had identified during its investigation of Pages Jaunes's activities. NATIONAL CNIL, *Red Card for the Yellow Pages*, (Sept. 23, 2011), http://www.cnil.fr/linstitution/actualite/article/article/carton-rouge-pour-les-pages-jaunes.

[337] For example, if someone were to look up the telephone number of Éloïse Gratton, Pages Jaunes would show Gratton's phone number, and would also show information on social media sites relating to individuals named Éloïse Gratton. The information displayed included photos, the name of employer, schools attended, geographic location, profession, etc. *Id.*

[338] *See id*. (describing the various sources of social networks where information was gathered).

with regards to the individual concerned.

a. Identifying Using Illegal Methods?

An important issue is whether the data should be evaluated taking into account the possibility of an illegal act or a security breach rendering certain pieces of data "identifiable."[339] I already discuss, in Part II.A.3, how this issue is not yet resolved.

I argue that when assessing if certain information qualifies as *personal,* one should focus on the extent of the risk of subjective harm that may arise following the disclosure. This risk should then be taken into account when determining whether to consider any illegal means involved in making certain data "identifiable." For example, if the data to be disclosed is not of an "intimate" nature and is widely "available," illegal means should not be taken into account in assessing if this information qualifies as *personal information*. On the other hand, if the information is of a very "intimate" nature and is not "available," then one should be more reluctant to dismiss considering the illegal means which may be used to determine if this data is "identifiable" or not.

This question of illegal means was raised in a recent case where key-coded clinical trial data, which had been anonymized, was to be transferred from Europe to the United States.[340] While some European agencies interpreted this as a transfer of *personal information* because the clinical trial data had not been "reversibly anonymized," or because the trial participants could be identified by the U.S. pharmaceutical company who had been in illegal contact with someone from the European clinical trial investigator, others disagreed.[341] While the data was *identifying* for the European company, it was potentially anonymous for the U.S. based partner (unless we consider that illegal means should be taken into account when determining whether this

[339] For instance, whether the mere possibility (such as a third party giving illegal access to identifying information) is enough to qualify strings of non-identifying numbers as *personal information*.

[340] *See* Commission Decision 2000/520/EC, O.J. (L 215) 7, *available at* http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000D0520:EN:HTML (discussing how this transfer of data would not constitute a transfer of personal data) [hereinafter Commission Decision].

[341] *Id.* at 24; *see also* RICHARD MORGAN & RUTH BOARDMAN, DATA PROTECTION STRATEGY: IMPLEMENTING DATA PROTECTION COMPLIANCE 40 (1st ed. 2003) ("[I]f the data is about a living individual who cannot be identified, the data is not personal data so far as the Act is concerned . . . .").

information is personal).[342] Using the proposed approach, since the information was of an "intimate" nature (i.e., health data) and not already "available" to the U.S. company, I maintain that the notion of "identifiability" should be interpreted less rigidly because the *risk of harm* upon this data being disclosed, once identified, is on the high side. Perhaps therefore, since it may be relatively easy to make a link between the clinical data and an individual, even using illegal methods, the key-coded clinical data should have been considered as being *personal information* even for the U.S. company.

The proposed approach can also be illustrated using the case of "IP addresses." These addresses by themselves may not qualify as *personal information* (for instance, if we don't take into account the illegal means of identifying the individual behind IP addresses).[343] If these addresses are linked with a profile that contains information of an "intimate" nature, then perhaps "illegal means," which may be used to put a name and a face to the profile behind an IP address, should be taken into account. The threshold to "identify" the individual should be lower (the information being considered as "personal" more easily) if the disclosure of this kind of "intimate" data is potentially much more "harmful." On the other hand, if by using the IP address as a point of collection, other more trivial information is collected (information that does not qualify as "intimate") the IP address and the information linked with this address may not be considered as *personal information.* Moreover, the illegal means of linking this profile with an actual person should not be taken into account in the overall assessment, given that the risk of harm is rather minimal. In the hands of the relevant ISP, that also has access to subscriber information, the information in question would be considered as *personal* (since it would definitely be identifiable).[344] However, the same information in the hands of another website that collects trivial information in

[342] *See* Commission Decision, *supra* note 340, at 10 (discussing how the Commissions decisions ensures a level of security that information transferred to other nations have some level of security).

[343] ISPs are usually prohibited by law to disclose the identity of the subscribers to which IP addresses have been assigned to.

[344] *See* Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, art. 2 (Can.) (defining "personal information" as information that pertains to an identifiable individual.); *see also* Personal Information Protection Act, S.A. 2003, c. P-6.5, art. 1(1)(k) (Can.) (stating that "'personal information' means information about an identifiable individual").

connection with dynamic IP addresses (which it then uses for operational purposes that has no negative impact for users, such as remembering the language of users or visitors) would not be considered as *personal information* (since it is not highly identifiable information and it contains no information of an intimate nature).

b. Efforts to Identify

Part II.A.3 details the fact that it is not always clear what kind of costs and efforts (or even resources) should be taken into account when determining if certain data is "identifiable."[345]

Using the proposed method of interpretation, as the risk of subjective harm increases, the "effort and costs" necessary to consider this data as "identifiable" tend to decrease. Interestingly, certain industry players are already suggesting or implying that the extent of the *risk of harm* (upon the information being disclosed) should be taken into account when determining whether certain information qualifies as *personal*.[346] Certain authors have also articulated similar views: Lundevall-Unger and Tranvik propose that the "likely reasonable" test in recital 26 of the Directive 95/46/EC "refers to the proportionality principle, which is well established in European Community Law."[347] The word "necessary" in Article 5 of the Treaty, according to Unger and Tranvik, is synonymous with "likely reasonable" in recital 26 of the Directive 95/46/EC, in the sense that both terms point towards the same type of assessment: the weighing of pros and cons so that a balanced and fair result can be achieved.[348] The "likely reasonable" test, therefore, would have to assess the effort and costs associated with linking "names and faces" to various pieces of information (like IP addresses) as well

345 I also discuss the fact that European courts, various academics and industry players do not agree on what "identifiability" actually means. *See* discussion *supra* Part II.A.3 (discussing how various parties define "identifiability").

346 Fleischer, *supra* note 156 ("As long as there is little or no chance of disclosure by the controller to a third party of information that could lead, in combination with data held by that person, to re-identification of individuals, then this approach seems more than reasonable.").

347 Lundevall-Unger & Tranvik, *supra* note 86, at 70.

348 *Id.* ("Particularly, Article 5 of the European Community Treaty provides that action taken by the Community shall not go beyond what is 'necessary' to achieve the objectives of the Treaty.").

as the "privacy risks" that this linking may entail.[349] The "privacy risks" which are referred to by these authors share some similarities with the "risk of subjective harm" test which I propose to take into account when interpreting the notion of "identifying."

Lundevall-Unger and Tranvik argue that in the context of IP addresses, the "likely reasonable" test should primarily consist of the effort and costs associated with putting "names and faces" to certain pieces of data with cost factors including "time, money, expertise, manpower, etc."[350] They argue that "the higher the costs, the less likely it is that the information is *personal* data (and the other way around)."[351] But more interestingly, these authors are also of the opinion that the nature of the information in question and the retention period should play a role in the evaluation.[352] They maintain that it is, for instance, reasonable that the threshold-value, the point where anonymous information becomes personal information (and vice versa), should be lower when we are talking about "sensitive information" (which can be assimilated to information of "intimate" nature) compared to when we are dealing with more trivial information.[353] They note, "Similarly, extending the retention period may facilitate the collection of additional information that will make the original data personally identifiable."[354] These authors' (Lundevall-Unger and Tranvik) views are consistent with the proposed approach, which maintains that if the information is of an "intimate" nature and

---

[349] *Id.*

[350] *Id.*

[351] *Id.* at 72 (emphasis added)

[352] *Id.*

[353] *Id.*

[354] *Id. See* Article 29 Working Party Opinion 4/2007, *supra* note 31, at 13. ("The name is a piece of information that reveals that the individual uses that combination of letters and sounds to distinguish himself and be distinguished by other persons with whom he establishes relations. The name may also be the starting point leading to information about where the person lives or can be found, may also give information about the persons in his family (through the family name) and a number of different legal and social relations associated with that name (education records, medical records, bank accounts). . . . All these new pieces of information linked to the name may allow someone to zoom in on the flesh and bone individual, and therefore through the identifiers the original information is associated with a natural person who can be distinguished from other individuals.").

not already "available," then the point where information becomes "personal information" is lower (information becoming more easily considered personal) then if we are dealing with information which is not of an "intimate" nature nor already "available."

In other words, if "connecting the dots" between information and the identity of an individual is relatively easy, then the information will most likely be considered as *personal*.[355] Using the proposed approach, if an anonymous IP address includes or is linked to profile information which is of an "intimate" nature and, or, includes information which is not generally "available," then the data (including the IP address) may be considered as *personal information*, even if the efforts or costs to link this information to a unique individual is relatively important or costly, because the disclosure of this information, if linked to an individual, may trigger a higher risk of subjective harm.

The challenge, then, is to identify the factors (effort and costs) that should be weighed against the potential "privacy risks" or what I refer to as the "risk of subjective harm" test. I leave it to better minds than mine to determine these factors, but I suggest that the simple rules which should be adhered to are the following: as the effort and costs increase, the less likely it is that information will qualify as *personal*, and as the "intimate" nature of the information and its non "availability" factors increase, the more likely it is that the information will qualify as *personal*.

c. Taking Into Account Potential Correlation

Part II.A.3 discusses how the definition of *personal information*, as it now stands, does not provide clear guidance as to whether correlation is needed for certain information to qualify as *personal*.[356] Trivial bits and pieces of very common

[355] For example, information may be disclosed (published) about a former criminal case without mentioning any name (or other identifier) linked to the individuals involved. If, for example, this case won much public attention in the past, then it would not seem unreasonably difficult to gain additional information (e.g., by looking up newspapers from the relevant time period) allowing one to find out the identity of the individuals involved. In such case, as suggested by Article 29 Working Party, it would seem justified to consider the information as being information about identifiable persons and as such, *personal information*. *Id.* at 14.

[356] Furthermore, what pool of data should be taken into account when assessing this correlation: data actually available to the data controller, data "likely to become available" to the data controller, or data in the hands of third

information may rarely qualify as "identifiable" *personal information*. Consider the name "John Smith" for instance. There may be well over hundreds of people in Canada that share this name. Therefore, "John Smith" will in fact very rarely relate to an identifiable individual. According to the Article 29 Working Group, the question of identifiability depends on the circumstances of the case:

> . . . the extent to which certain identifiers are sufficient to achieve identification is something dependent on the context of the particular situation. A very common family name will not be sufficient to identify someone—i.e. to single someone out—from the whole of a country's population, while it is likely to achieve identification of a pupil in a classroom. Even ancillary information, such as "the man wearing a black suit" may identify someone out of the passers-by standing at a traffic light. So, the question of whether the individual to whom the information relates is identified or not depends on the circumstances of the case.[357]

Also, the information "by itself" will rarely create a risk of subjective harm upon its disclosure. If someone was to post the name "John Smith" on a website, it would not create any type of harm unless the website included additional details. It is the correlation between "John Smith" and another piece of information, such as being afflicted with a particular disease or being a member of a special interest group, that may in fact create a risk of subjective harm upon being disclosed.

When assessing the notion of "identifying," data availability and correlation should be prime factors. Certain legislators have in fact taken the position that the disclosure of the name of an individual by "itself" creates no harm.[358] Therefore, in interpreting the notion of "identifiability" which is necessary in assessing whether the disclosure of certain pieces of data will create a risk of harm, correlation is a key factor.[359] This point is

parties as well? Moreover, it has now become considerably easier to link certain data to individuals, simply from the sheer availability of enormous amounts of data on the Internet, from the correlation of data across various online services and from the use of new identification tools.

[357] *Id*. at 12–13.

[358] The Quebec public sector DPL for instance actually takes the position that the name of an individual is not *personal information*, except where it appears "in conjunction" with other information concerning this individual (or where the mere mention of this individual's name would disclose something personal concerning him or her). See *An Act Respecting Access to Documents Held by Public Bodies and the Protection of Personal Information*, R.S.Q., c. A-2.1, s. 56.

[359] The privacy threat is in the aggregation of consumer records, as outlined

further illustrated by van den Hoven with the following example:

> Let's consider the following two claims C1: "X is in a restaurant A at time ti" and C2: "Y is in Restaurant A at time tr". Is C2 about X? C2 presents itself obviously as information about Y. When looked at in isolation, "Y is in Restaurant at time ti" does not tell you anything about X, but when combined with C1, it does provide information about X which was not contained in C1. Good detective stories often present information to the reader which is seemingly irrelevant to the crime or to the biography of the protagonist, but later turns out to be, in an unexpected sense, about the murderer or his victim. As the story unfolds and the plot unravels, the insignificant piece of information is situated in a context where it suddenly picks out an individual. We suddenly see how the insignificant and seemingly irrelevant piece of information suddenly applies to the protagonist.[360]

The true impact of data-mining can only be meaningfully assessed when taking into consideration other data available.[361] Data volume will play a role in increasing the potential for identification because it increases the potential for data correlation.[362] As the volume of data increases, so too do the chances for identifiability. A good illustration of this can be made using the 2006 AOL breach discussed in Part II.A.2.[363] While a

---

by Stan Karas who suggests that "[a] single retailer's consumer file may be extensive, but its scope is unlikely to be comprehensive enough for a true representation of our consumer identities" and that the true danger is in the compilation of our transactional data. *See* Karas, *supra* note 215, at 445.

[360] van den Hoven, *supra* note 1, at 307.

[361] Ian Kerr and Jenna McGill share similar views. They argue that: "In fact, as new and emerging information technologies continue to come before the courts, we predict that the current reductionist inclination which asks whether the intercepted data is, *on its own*, meaningless will and ought to give way to the very opposite approach, namely: whether the bundle of information that is made available by means of the search, *once assembled,* ought to attract a reasonable expectation of privacy." Kerr & McGill, *supra* note 63, at 430–31.

[362] *See* GRATTON, *supra* note 15, at 21; *see also* Ohm, *supra* note 34, at 1766–67 ("Most privacy laws regulate data quality but not quantity. Laws dictate what data administrators can do with data according to the nature, sensitivity, and linkability of the information, but they tend to say nothing about how much data a data administrator may collect, nor how long the administrator can retain it. Yet, in every reidentification study cited, the researchers were aided by the size of the database. . . . Thus, lawmakers should consider enacting new quantitative limits on data collection and retention. They might consider laws . . . limiting the total quantity of data that may be possessed at any one time."); *see also* Article 29 Working Party Opinion 1/2008, *supra* note 138, at 19 (arguing that search engines should store queries for a maximum of six months).

[363] *See supra* Part II.A.2.

single piece of data by itself may be meaningless (in this case, a single "web search") it may nevertheless be possible, because of the volume of data available (i.e., "all searches made" by a given profile over a three month period) to actually make the link between these searches and an identifiable individual, even if the name of the individual is not revealed.

Clearly, the more work required to make a link between a piece of information and an individual, the less likely that information may be considered as being "identifying." To have one piece of a complex puzzle is one thing—but the ease with which additional pieces can be obtained must always be given consideration;[364] in light of the overall risk of harm that may take place upon the information being disclosed. For example, upon the merger of organizations, this correlation should be taken into account especially since the link can be made effortlessly.[365] In such cases, if a business transaction triggers the merging of databases which will result in highly identifiable profiles, perhaps this information (each database and definitely, the "resulting profiles") should be considered as personal; depending, of course, on the intimate nature and availability of the information in question.[366] Therefore, under the proposed approach, we need to evaluate the ease with which correlation can occur, along with the "intimate" nature of the information and whether it is already "available."[367]

---

[364] Kerr & McGill, *supra* note 63, at 431 (articulating the position that the jigsaw nature of the data/information/knowledge/wisdom chain and the importance of each piece of the puzzle in telling a story despite the fact that no single piece could do so on its own should be recognized.).

[365] Article 29 Working Party Doc. on RFID Tech., *supra* note 75, at 9 ("Other scenarios that can lead to identifiability are mergers, data losses and the increasing availability on the Internet of personal data in combination with IP addresses.").

[366] For example, when DoubleClick and Abacus announced a merger one of the executives proudly states, "The goal is to have the most complete picture of the consumer you can." *Quoted in* Beth Givens, *Privacy Expectations in a High Tech World*, 16 SANTA CLARA COMPUTER & HIGH TECH. L.J. 347, 352 (2000). Consolidation of information databases may also happen through sales of consumer lists of defunct dot coms, a phenomenon more common in the current economic climate. *See* Richard A. Beckmann, *Privacy Policies and Empty Promises: Closing the "Toysmart Loophole,"* 62 U. PITT. L. REV. 765, 772–773 (2001).

[367] *See supra* Part III.A.2. (in the sense that according to the proposed approach, the more important is the *risk of subjective harm*, the less work is in fact required for the data to qualify as "personal.").

d. Dealing with New Types of Data

The first step in determining "identifiability" begins with the proper qualification of information.[368] In Part II.A.3, I already discuss how certain data (which may qualify as "identifiers" or "points of collection") may not always be considered as personal information by courts or industry players.[369] The Comité consultatif produced a report in which they propose that the notion of "identity" can mean three different things: (i) first, characteristic traits, such as age, information pertaining to family, hobbies, employer, professional qualifications, movements, purchases, etc.; (ii) second, a point of collection or an identifier that may allow a linkage to different data and biographical characteristics from the same person (this could mean a permanent cookie, a client number, a number identifying a terminal); (iii) third, a point of contact that would enable a third party to take the initiative to contact an individual (by email, mail, fax, phone, etc.).[370] The Comité consultatif suggests that with time and throughout its life cycle, the status of a certain piece of data may change: For example, a dynamic IP address may be a point of collection for a short period.[371] They suggest that an address may be both a point of collection as well as a point of contact, and the vulnerability of postal addresses would result from the fact that this kind of data would accumulate the three properties above.[372]

Nowadays, "points of collection" or "identifiers" can also be supermarket loyalty cards, RFID tags or mobile coupons that track customers.[373] The definition of personal information found in most DPLs does not specify whether these identifiers (or points of collection) constitute personal information. When discussing the notion of "identifiability," it is important not to ignore the fact that devices acting as points of collection or identifiers (such as IP addresses, RFID tags, cookies, wireless

368 *See infra* Part III.A.

369 *See* discussion *supra* Part II.A.3.

370 Poullet et al., *supra* note 58, at 30–31.

371 *Id.* at 31.

372 *Id.*

373 *See* Article 29 Working Party Working Doc. on RFID Tech., *supra* note 75, at 5–6; Ariana Eunjung Cha, *Mobile Coupons Track Customers*, WASH. POST (June 27, 2010), http://www.washingtonpost.com/wp-dyn/content/article/2010/06/26/AR2010062600223.html.

devices, etc.) may reveal information of "intimate" nature.[374]

The first step in ascertaining a risk of harm with the disclosure of information under the "identifiability" criteria, is to begin with qualifying the kind of information in question (biographic information, point of collection, point of contact, or some or all of the above criteria).[375] We can't be too quick to disregard an IP address as personal information simply because it belongs to a device instead of an individual (which may or may not be "identifiable"). For example, an IP address may be a simple point of collection with no additional information attached to it and therefore, less harmful if disclosed; therefore, it should not qualify as personal information (if not linked to any biographic or contact information). An IP address leading to a device may also not be considered personal information if it doesn't reveal "intimate" details (for instance, it is only used to remember the preferred language of its website users and is not otherwise made available to third parties). An IP address coupled with biographic information becomes more "sensitive," especially when linked with biographic information that is of an "intimate" nature and becomes more "sensitive" if this information was not already "available." In this case, the IP address (together with the information linked to it) would clearly qualify as personal information. This IP address would become even more potentially harmful when associated with a point of contact (such as an email address, a user account or a physical address).

Recent technological advancements have opened the way for data to be collected by a certain device that may be associated with a group of individuals; for instance, an IP address linked to a computer used by a few co-workers, family members or library users (vs. a unique individual). It is not always clear in such cases whether the IP address is "identifiable," as detailed in Part II.A.3.[376] In order to determine when data belonging (or potentially belonging) to a group of individuals should be covered by the definition of personal information, one should take into

[374] *See* Article 29 Working Party Working Doc. on RFID Tech., *supra* note 75, at 7 ("Belongings of a person are very personal and hold information whose knowledge by third parties would invade the privacy of the person who owns the object. The following examples illustrate this hypothesis. Consider the case where anyone in possession of a reader can detect banknotes, books, medicines or valuable objects of passers by. The knowledge of this information by third parties will invade the privacy of the person who owns the object.").

[375] *See supra* Part II.A.3.

[376] *See supra* Part II.A.3.

account its intimacy and availability.[377] For example, while information of a very "intimate" nature which is not otherwise "available" should be considered personal even if it belongs to a small group of individuals (ex: a handful of individuals using the same computer), this should not be the case if the information is more trivial and more easily "available." In this last case, the information should only be considered as personal information if it can be linked to a unique individual, since the risk of harm is much lower.

Another issue further discussed in Part II.A.3 is that it is not always clear how accurate the link must be between certain information and an individual in order for the data to qualify as "identifying."[378] The name of an individual is indeed the most common identifier. In practice, the notion of an "identified individual" usually implies a reference to the individual's name. As discussed, "in order to ascertain this identity, the name of the person sometimes has to be combined with other pieces of information (date of birth, parents, an address or a photograph) to prevent confusion between that individual and possible namesakes."[379] With regards to the notion of "indirectly" identified or identifiable persons, as detailed in article 2 of the Directive 95/46/EC, the Article 29 Working Party articulates the view that "this category typically relates to the phenomenon of unique combinations, whether small or large in size."[380] The

[377] *See supra* Part III.A.1.

[378] *See supra* Part II.A.3.

[379] Article 29 Working Party Opinion 4/2007, *supra* note 31, at 13 (illustrating this idea with an example, "The name is a piece of information that reveals that the individual uses that combination of letters and sounds to distinguish himself and be distinguished by other persons with whom he establishes relations. The name may also be the starting point leading to information about where the person lives or can be found, may also give information about the persons in his family (through the family name) and a number of different legal and social relations associated with that name (education records, medical records, bank accounts). . . . All these new pieces of information linked to the name may allow someone to zoom in on the flesh and bone individual, and therefore through the identifiers the original information is associated with a natural person who can be distinguished from other individuals.").

[380] *See id.* (The Article 29 Working Party articulates the view that "[i]n cases where *prima facie* the extent of the identifiers available does not allow anyone to single out a particular person, that this individual might still be "identifiable" because that information combined with other pieces of information (whether the latter is retained by the data controller or not) will allow the individual to be distinguished from others.").

Directive 95/46/EC comes in with "one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."[381] The Article 29 Working Party maintains that while some characteristics are so unique that someone can be identified with relative ease ("present Prime Minister of Spain"), "a combination of details on a categorical level (such as age category, regional origin, etc.) may also be pretty conclusive in some circumstances, particularly if one has access to additional information of some sort."[382]

Using the proposed approach, if the information evaluated (or the bundle of information) reveals information of an "intimate" nature (i.e., John Smith from Montreal who suffers from AIDS) and this information is not already "available," then the fact that the data relates to a small group of individuals (for example ten individuals named John Smith who live in Montreal) may be sufficient to argue that this data is personal. As a matter of fact, since the information linked to the name John Smith is of a very "intimate" nature (i.e., being afflicted with AIDS) and not already in circulation or "available," then the interpretation of the notion of "identifiability" should be interpreted less stringently. On the other hand, let us consider the John Smith who is a resident of Montreal and subscribes to the Montreal Gazette, a general interest newspaper. Even though there may be three John Smiths who fall into this category, the disclosure of this information would not present a considerably high risk of harm and therefore the data should not be considered as warranting stringent protection.

In light of the approach proposed in this article, the "identifiability" criteria has to be interpreted more softly if the data is otherwise sensitive (in the sense that its disclosure is potentially harmful since it is of "intimate" nature, not already "available" and it is "identifiable"). Therefore, the more "intimate" and the less "available" the information, the less important this "accuracy" factor (i.e., accuracy in identifying a unique individual) will actually play in the evaluation of the information. On the contrary, if the information is not of a very "intimate" nature, or it is "intimate" but it is already "available," then this "accuracy" factor will be more important to get to the point of qualifying the data as personal.

---

381 Article 29 Working Party Opinion 4/2007, *supra* note 31, at 4.

382 *Id.* at 13.

2. Applying the Approach to Recent Privacy Breaches or Activities

Certain DPLs provide for some measure of subjectivity when it comes to disclosing data protection practices to individuals,[383] and there is also a lot of subjectivity surrounding the notion of "consent" since the form of consent sought by the organizations may vary, depending upon the circumstances and the type and the sensitivity of the information.[384] Under the approach proposed, if the risk of harm were medium to high, perhaps a more stringent type of disclosure and consent would be required. If the risk is present but is on the "low" side, then perhaps a disclosure and an opt-out type consent would be sufficient (for instance, a simple privacy policy posted on a website) or the information should not be regulated under DPLs.

DPLs also usually provide for subjectivity in the assessment of the security measures that have to be implemented by an organization to protect the personal information that it is handling.[385] In determining what kind of security measures to

[383] *See* Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, Schedule 1 (s. 5), principle 4.3.2. (Can.) (stating that, in Canada, organizations shall make a "reasonable effort" to ensure that the individual is advised of the purposes for which the information will be used or disclosed, which must be communicated in such a manner that the individual "can reasonably understand" such purpose.); *see also* Personal Information Protection Act, S.A. 2003, c. P-6.5, Part 2, Division 2, s. 8 (3) (Can.) (showing that the Alberta DPL has a similar reasonableness provision since an organization may collect, use or disclose personal information about an individual if it provides notice, "in a form that the individual can reasonably be expected to understand."); Personal Information Protection Act, S.B.C. 2003, c. 63, Part 3, s. 8 (3) (a)–(e) (listing a similar requirement).

[384] *See* Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, Schedule 1 (s. 5), principle 4.3.44. (Can.) (specifying that the form of consent sought by the organizations may vary, "depending upon the circumstances and the type of information" and "the sensitivity of the information."). Furthermore, in obtaining consent, "the reasonable expectations of the individual are also relevant." *Id.* at Schedule 1 (s. 5), principle 4.3.5; *see also* Personal Information Protection Act, S.A. 2003, c. P-6.5, Part 2, Division 2, s. 8 (2)(b) (Can.) (relating how an individual is deemed to consent to the collection, use or disclosure of personal information if "it is reasonable" that a person would voluntarily provide that information.); Personal Information Protection Act, S.B.C. 2003, c. 63, Part 3, s. 8 (1) (a),(b) (Can.) (emphasizing that an individual is deemed to consent to the collection, use or disclosure of personal information if at the time of consent, the purpose would be considered to be "obvious to a reasonable person.").

[385] *See* Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, art. 2(1), s. 8(1)(a), (b), principle 4.7 (Can.) (providing that personal information shall be protected by security safeguards "appropriate to the

adopt and whether these measures are "reasonable," "necessary" or "appropriate" in accordance with the relevant DPL, the first step for the organization is to determine the extent of the *risk of harm* to individuals upon the occurrence of a security breach (or a disclosure of personal information).[386] Many have already determined that the "nature" of the information is relevant in assessing the risk of harm upon disclosure and that information which is not very "intimate" in nature such as "the name, address, or membership of a local drama group" does not need to be the subject of very robust standards of security.[387] Using the approach proposed, organizations will have to account for the "intimate" nature of the information in establishing the proper measures to adopt, while also assessing the other relevant criteria discussed in this section (whether the information is already "available" and whether it is "identifiable").[388]

There are various "reasonable" tests under certain DPLs. For example, under PIPEDA, an organization may only collect, use or disclose personal information "for purposes that a reasonable person would consider appropriate in the circumstances."[389] Other DPLs in Canada (Alberta and B.C., Quebec) and in Europe have similar reasonableness tests.[390] I maintain that a first step

sensitivity of the information," and that the nature of the safeguards will vary depending on "the sensitivity of the information that has been collected, the amount, distribution, and format of the information, and the method of storage."); *see also* An Act Respecting the Protection of Personal Information in the Private Sector, R.S.Q. 1993, c. P-39.1, s.10 (Can.); Personal Information Protection Act, S.A. 2003, c. P-6.5, Part 3, Division 2, s. 34; and Personal Information Protection Act, S.B.C. 2003, c. 63, Part 9, s. 34 (Can.) (showing similar and very subjective security requirements.); Directive 95/46, *supra* note 7, at Article 17 states that "appropriate" technical and organizational measures be taken in order to maintain the security of the data, and that such measures shall ensure a level of security "appropriate to the risks represented by the processing and the nature of the data to be protected."

386 Directive 95/46, *supra* note 7, at Article 17.

387 *Security Measures*, *supra* note 246.

388 *Id.*

389 Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, Schedule 1 (s. 5), s.5(3) (Can.).

390 In meeting its responsibilities under the Alberta DPL or the B.C. DPL, an organization must act "in a reasonable manner," and must develop and follow policies and practices "that are reasonable for the organization" to meet its obligations. *See* Personal Information Protection Act, S.A. 2003, c. P-6.5, (Can.); Personal Information Protection Act, S.B.C. 2003, c. 63, Schedule1 (s. 5) s.5(3) (Can.). The golden standard is as follows: "what a reasonable person would consider appropriate in the circumstances"; quite a subjective criterion. See Personal Information Protection Act, S.A. 2003, c. P-6.5, s. 2 (Can.); Personal Information Protection Act, S.B.C. 2003, c. 63, Schedule1 (s. 5) s5(3) (Can.). In

in determining whether a given data handling activity is "reasonable," is to determine if this activity will create a risk of harm to the individual concerned (since the lower is this risk, the more chances that the disclosure be considered as reasonable).

I will now illustrate how the outcome of certain situations, case law or privacy scandals, which can be assimilated to disclosures of information, would have been different if the approach proposed in this article had in fact been used.

a. High Risk of Harm: Launch of Buzz and AOL breach

The disclosure of personal information may create the highest risk of harm if the data is of an "intimate" nature, is not already "available" (or has, to a certain extent, been kept confidential) and is otherwise highly "identifiable" to a unique individual.[391]

To illustrate this thought, let's recall the privacy concerns that took place when Google released its "Buzz" service, a social-messaging system built into the Gmail service.[392] A major concern was that the earliest versions of the service revealed a list of the individuals the Gmail user e-mailed most frequently; which was found to be a privacy breach.[393] This type of information could lead to various unpleasant scenarios: for instance "a wife discovering that her husband emails and chats with an old [flame] . . ." or a boss discovering that his employee exchanges emails with executives at a competitor.[394] Under the approach proposed, before launching a service such as Buzz, the organization would have had to first acknowledge the potential

Quebec, an organization can only establish a file on an individual for a "serious and legitimate reason." An Act Respecting the Protection of Personal Information in the Private Sector, R.S.Q. 1993, c. P-39.1, s. 2 (Can.) at Division II s. 4; Directive 95/46/EC states that "any processing of personal data must be lawful and fair to the individuals concerned," and personal data must be collected for "legitimate" purposes. Under such "reasonableness," "legitimacy," or "fairness" tests, it is the organization handling the data that will make the judgment call of what is "reasonable," "legitimate," or "fair," which is a very subjective assessment.

391 *See* discussion *supra* Part III.B.1.

392 Todd Jackson, *Introducing Google Buzz*, GOOGLE OFFICIAL BLOG (Feb. 9, 2010), http://googleblog.blogspot.com.

393 Nicholas Carlson, *WARNING: Google Buzz Has A Huge Privacy Flaw*, BUS. INSIDER (Feb. 10, 2010; 4:49 PM), http://www.businessinsider.com/warning-google-buzz-has-a-huge-privacy-flaw-2010-2#ixzz1XIxQ9N8V; Robert McMillan, *Google Buzz Criticized for Disclosing Gmail Contacts*, PC WORLD (Feb. 10, 2010, 5:30 PM), http://www.pcworld.com/article/189081/google_buzz_criticized_for_disclosing_gmail_contacts.html.

394 Carlson, *supra* note 393.

disclosure of the names or email addresses of the individuals with which a Gmail user communicates most frequently.[395] Using the approach proposed in this section, since the information disclosed through the Buzz service was most likely of an "intimate" nature[396] and it was clearly linked with Gmail users (and therefore "identifiable"), unless one could demonstrate and prove that this information (the fact that Gmail user X communicated with individual Y most frequently) was already public (the "availability" test),[397] then it should not have disclosed this information upon the launching of the Buzz service. To make this disclosure worse, the disclosure of personal information actually impacted not only the Gmail users, but also the individuals with which Gmail users communicated with the most frequently.[398]

Another example to illustrate the outcome of the approach is with the privacy breach which took place on August 4, 2006, when AOL Research published (publicly disclosed for research purposes) a compressed text file on one of its websites containing twenty million search keywords which had been punched into AOL's search engine for over 650,000 anonymous AOL users over a 3-month period further discussed in Part II.A.2.[399] Using the approach proposed, the analysis conducted by AOL Research would take into account that while not every single profile was "identifying," given the volume of the information made available (millions of search keywords punched for over 650,000 AOL users

---

[395] *See supra* Part III.A.1 (proposed approaches to disclosing personal information).

[396] Personal communications are usually considered as "intimate" information. In the U.S., the *Electronic Communications Privacy Act of 1986* was enacted to extend government restrictions on wiretaps from telephone calls to include transmissions of electronic data by computer. 18 U.S.C. §§ 2510–2522 (1968). Title I of the ECPA protects "wire, oral, and electronic communications" while in transit, while Title II of the ECPA, the *Stored Communications Act* protects "communications while it is in electronic storage" (on computers). 18 U.S.C.A. §§ 2701–2712 (1986). The Civil Code of Quebec (1991) at article 36 (2) states that "intentionally intercepting or using [someone's] private communications" is considered an invasion of privacy. Civil Code of Québec, S.Q. 1991, c. 64 art. 36(2) (Can.). In Europe, the *Security Measures for Personal Data: A Guide to the New Data Protection Rules* provides that "[o]rganisations dealing with personal data of a private or sensitive nature—such as people's . . . private telecom[munications] naturally need to have very robust standards of security in place." *Security Measures*, *supra* note 246.

[397] *See supra* Part III.A.1.

[398] Carlson, *supra* note 393.

[399] *See* discussion *supra* Part II.A.2.

over a 3-month period), the "potential" to identify some of the users was present.[400] Although it wasn't clear if the information was "identifiable," the fact that the information was clearly of an "intimate" nature,[401] coupled with the fact that there was no evidence that this data was already "available" to the public, the notion of "identifiable" was to be interpreted more softly. This risk, evaluated using the proposed test, would have refrained AOL from disclosing this research data to the public.

Since the notion of "identifiability" is never foolproof, upon having a certain volume of information, all it takes is for the information released (disclosed) to come in contact with one single piece of information to make this bundle of information "identifiable" as illustrated by van den Hoven (this example was initially used by Gavison):

> Consider the famous anecdote about the priest who was asked, at a party, whether he had heard any exceptional stories during confessionals. 'In fact', the priest replied, 'my first confessor is a good example, since he confessed to murder'. A few minutes later, an elegant man joined the group, saw the priest, and greeted him warmly. When he was asked how he knew the priest, the man replied: "Why, I had the honour of being his first confessor.[402]

While the priest initially did not disclose "identifiable" information according to the standard legal definition found in most DPLs, the information disclosed coming in contact with another piece of information completely changed the picture.[403] Using the approach proposed, the fact that the data was of a very "intimate" nature and not widely "available," should have been enough for the priest to limit the disclosing of this confession in order to limit the risk of subjective harm triggered by disclosure, which was medium to high in this situation.

Information which is of a very "intimate" nature, and not "available," may still be potentially harmful upon being disclosed

---

400 *Id.*

401 Paul Ohm suggests that search engine data are even more sensitive than health data. *See* Ohm, *supra* note 34, at 1775–76 ("We reveal even more than health information to search engines, supplying them with our sensitive thoughts, ideas, and behavior, mixed in of course with torrents of the mundane and unthreatening."); *see also* Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object,* 52 STAN. L. REV. 1373, 1426 (2000) (stating how a loss of privacy will constrain and stifle individuality in society).

402 van den Hoven, *supra* note 1, at 309 (citation omitted).

403 *Id.*

even if it is absolutely not "identifiable."[404] For example, in *Northwestern Memorial Hospital* v. *Ashcroft*,[405] Posner comments on the fact that a privacy breach may still occur even if a person cannot be identified by name on the Internet: "Imagine if nude pictures of a woman, uploaded to the Internet without her consent though without identifying her by name, were downloaded in a foreign country by people who will never meet her. She would still feel that her privacy had been invaded."[406] Usually though, the *risk of harm* in such cases will be lesser than if the information was "identifiable" (in this case, if the woman was identified on the picture).

b. Low Risk of Harm: Note2be

The *risk of harm* relating to the disclosure of information becomes lower if it has to do with information, which may be "identifiable," but is not of an "intimate" nature, and is already "available" to a certain extent.[407]

In the Note2Be case law, the French court found that the processing of the name, workplace and rating by students of their teachers were found to be illicit and the website (similar to www.ratemyteacher.com or www.ratemyprofessor.com) was in part shut down.[408] Using the approach suggested here, one could claim that the disclosure test would have concluded that there was no *risk of harm* in the disclosure at stake. While the data was highly "identifiable" (name of teacher, place of work), it was also already "available" (if not publicly available) and not of "intimate" nature.[409] Under certain DPLs such as PIPEDA, "business contact information" (name of employee and their address of place of work) is even excluded from the definition of personal information.[410] Some authors, such as Trudel and

404 *See* Nw. Mem'l Hosp. v. Ashcroft, 362 F.3d 923, 929 (7th Cir. 2004) (stating that despite the identity of a patient not being discernable from a redacted medical, there nonetheless is a risk of invasion of privacy).

405 *Id.* at 923.

406 *Id.* at 929.

407 *Contra*, Tribunal de Grande Instance [TGI] [Ordinary Court of Original Jurisdiction] Paris, Mar. 3, 2008, 08-51650 (Fr.).

408 *Id.*; *see also* Press Release, CNiL, La CNIL se Prononce: Le Site Note2be.com est Illégitime au Regard de la loi Informatique et Libertés (Mar. 6, 2008) (Fr.), *available at* http://archive.is/96Ykc.

409 Unless we take the position that the notations provided by students were data of "intimate" nature.

410 Personal Information Protection Act, S.A. 2003, c. P-6.5, part 1, para. 2

Gautrais, have even raised the fact that it was not clear in this Note2be case whether the information at stake should in fact have been qualifying as *personal information* given its low sensitivity.[411]

Information, which is not of "intimate" nature, is already "available" and is not "identifiable," is on the lowest section in the risk of subjective harm upon being disclosed. This kind of information should not be regulated by DPLs or at least, the information should be able to circulate without having to obtain the relevant individual's prior consent.

There is information that, once disclosed, may be harmful not because of the fact that this data may create some type of embarrassment, but because of the way that it may be "used" by third parties. This may include financial information, which, if released (by banks or e-commerce websites), may be used to create harm such as fraud or identity theft. This could also include location data, which may be used by a stalker to physically harm another individual. In another U.S. example, an Internet site known as the "Nuremberg Files" posted information about doctors working in abortion clinics, including names, photos, Social Security numbers, home addresses, descriptions of their cars, and information about their families.[412] The doctors sued and at trial they testified as to how their lives became riddled with fear, how some wore bulletproof vests and wigs in public.[413] This is a clear illustration how sometimes a disclosure may trigger a more objective kind of harm, which mostly relates to the fear of this information being "used."

As Solove suggests: "Privacy . . . involves more than avoiding disclosure; it also involves the individual's ability to ensure that personal information is used for the purposes she desires."[414] The

(Can.).

411 Pierre Trudel and Vincent Gautrais are raising the fact that the French court never discussed whether the data at stake (name of teacher, name of school and notations) actually qualified as *personal information*. GAUTRAIS & TRUDEL, *supra* note 209, at 119.

412 Doctors who were killed had a black line drawn through their names. Names of wounded doctors were shaded in gray. *See* Planned Parenthood of the Columbia/Williamette, Inc. v. Am. Coal. of Life Activists, 244 F.3d 1007, 1013 (9th Cir. 2001).

413 *Id.* While the doctors were victorious in their suit and the site was shut down, the appellate court reversed on First Amendment grounds. Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy,* 53 STAN. L. REV. 1393, 1426 (2001).

414 Solove, *supra* note 46, at 1108.

risk of harm triggered by the "use" of information is addressed in the next section.

B. Objective Harm Associated with Personal Information

The last data handling activity regulated by DPLs is the use of personal information.[415] Reidenberg aptly observes: "the creation of special protection is also understood as requiring attention not only to whether information identifies particular aspects of a person's life that are sensitive, but how data will actually be used."[416]

Calo explains that while at the collection or disclosure levels, the corresponding harm may be subjective in nature,[417] the consequence of a third party using data would be much more concrete and in many cases, would have financial implications.[418] According to Calo, the objective category of privacy harm would be the unanticipated or forced use of personal information against a given person:

> The second category is 'objective' in the sense of being external to the person harmed. This set of harms involves the forced or unanticipated use of information about a person against that person. Objective privacy harms can occur when personal information is used to justify an adverse action against a person, as when the government leverages data mining of sensitive

---

[415] This particular activity (or the term "using") is not defined in the Canadian or French DPLs analyzed. In Europe, the activity of "processing" the information includes the "use" of personal information. As a matter of fact, EC Directive 95/46, at Article 2(b) defines "processing of personal data" as "any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction." Directive 95/46, *supra* note 7, at 38.

[416] REIDENBERG & SCHWARTZ, *supra* note 146, at 9.

[417] Calo, *supra* note 269, at 1147–48 ("Subjective privacy harms are injuries individuals experience from being observed. But why does the belief that one is being observed cause discomfort or apprehension? In some instances, the response seems to be reflexive or physical. The presence of another person, real or imagined, creates a state of 'psychological arousal' that can be harmful if excessive and unwanted.").

[418] *See, e.g.*, *In re* TJX Cos. Retail Sec. Breach Litig., v. TJX Cos., Inc., 564 F.3d 489, 491 (1st Cir. 2009) (describing the financial injury to former costumers of TJX Companies when its computer system was hacked and costumer credit and debit card information was stolen).

personal information to block a citizen from air travel, or when one neighbor forms a negative judgment about another based on gossip. Objective harms can also occur when such information is used to commit a crime, such as identity theft or murder.[419]

It is often the use of information that leads to a more tangible kind of harm. For example, if the criminal record of a bank employee is disclosed to his co-workers, this employee may feel embarrassed and humiliated (subjective harm resulting from the disclosure). Once the information is used by the bank to dismiss the employee, the resulting harm will be objective in nature (in this case, a financial or economical harm).

Documents from the early 1970s produced in the context of the adoption of DPLs already raised the concern of having organizations use the information of individuals in a way which would be detrimental to them.[420] In 1973, the *Report of the Secretary's Advisory Committee on Automated Personal Data Systems* (U.S.) mentioned that privacy was directly affected by the kind of "uses" made of personal information.[421] In the late 1970s, in the U.K., while discussing the adoption of a DPL or some type of regulation incorporating the FIPs, the Lindop Committee was already suggesting that individuals should be able to know if their data was to be used as the basis of "an adverse decision against them,"[422] and that "outdated data" should be discarded especially when "used for making decisions which affect the data subject."[423]

Solove argues that the use of personal information in databases presents a different set of problems than does government surveillance[424] and, therefore, the Big Brother metaphor fails to capture the most important dimension of the database problem.[425] He uses the metaphor of Franz Kafka's *The Trial*, to illustrate the problem (or the harm) resulting from databases and the activity of "using" personal information.[426] In

419 Calo, *supra* note 269, at 1143.

420 WARE, *supra* note 220.

421 *Id.*

422 LINDOP, *supra* note 12, at 49.

423 *Id.* at 51.

424 Daniel J. Solove, *I've Got Nothing to Hide" and Other Misunderstandings of Privacy,* 44 SAN DIEGO L. REV. 745, 756 (2007).

425 Solove, *supra* note 413, at 1399.

426 *Id.* at 1429 ("In sum, the privacy problem created by the use of databases stems from an often careless and unconcerned bureaucratic process—one that has little judgment or accountability—and is driven by ends other than the

The Trial, an unscrupulous bureaucracy uses personal information to take important decisions, while denying the relevant people the ability to participate in how their information is being used. Solove states that this problem is derived from information processing (which he defines as the storage, use and analysis of data) rather than simply information collection.[427] According to him, this sort of information processing (or use of information) would affect power relationships between people and the institutions of the modern state.[428] The individual would be frustrated by a "sense of helplessness" and "powerlessness." Social structure would also be affected by altering the kinds of relationships people have with the institutions that make important decisions about their lives.[429]

A broad range of harms can be inflicted on data subjects emerging out of the use of their personal information. van den Hoven believes that the first type of moral reason for thinking about constraining the flow of personal information is concerned with the prevention of information-based harm, which includes financial harm such as theft or identity fraud.[430] When discussing the type of harm that may result from the use of personal information, RAND Corporation (U.K., 2009) also refers to an economic harm such as "financial damages suffered as a consequence of identity theft, loss of earnings."[431] The Canadian breach notification guidelines and provisions discuss the fact that individuals should be notified in case of a security breach triggering a loss of employment, business or professional opportunities, financial loss, identity theft, negative effects on the credit record and damage to or loss of property.[432] Theft is another type of economic harm, which may take place upon the use of personal information by thieves (e.g., home address,

protection of people's dignity. We are not heading toward a world of Big Brother or one composed of Little Brothers, but toward a more mindless process—of bureaucratic indifference, arbitrary errors, and dehumanization—a world that is beginning to resemble Kafka's vision in The Trial.").

[427] *See generally* Solove, *supra* note 284, at 490–91 (describing the "information processing" process).

[428] Solove, *Privacy*, *supra* note 413, at 1455.

[429] *Id.* at 1456.

[430] van den Hoven, *supra* note 1, at 311.

[431] ROBINSON ET AL., *supra* note 55, at 48.

[432] *See* PIPA, INFORMATION SHEET 11: NOTIFICATION OF A SECURITY BREACH (2010), *available at* http://servicealberta.ca/pipa/documents/infosheet11.pdf (discussing the Alberta Notification of a Security Breach procedures).

whereabouts of the home owner). [433]

The second type of harm is one that van den Hoven refers to as "Informational Inequality" (or discrimination).[434] According to him, this type of moral reason to justify constraints on our actions with identity-relevant information is concerned with *equality and fairness*.[435] As early as the 1970s, misuse of data and the resulting discrimination was of paramount importance; evidence of this can be found in the documents leading to the adoption of Convention 108.[436]

Information may be used to discriminate, remove a benefit, or tarnish a reputation and an individual may be subject to some type of discrimination, which could lead him to being refused for a job, refused for credit, mortgage or a loan, etc. Many have voiced their concerns about consumer profiling, as it may be a tool used to facilitate the practice of discrimination.[437] With the onslaught of new Internet technologies, online profiling activities are taking on a range of different forms. One discriminatory practice taking place online is known as "adaptive pricing" or "dynamic pricing."[438] Amazon was suspected of using such practices, using cookies to identify the visiting consumers.[439] In the U.K., the OFT has also expressed its concern over price discrimination, especially if consumers are left in the dark.[440]

---

[433] *Guidelines: Key Steps for Organizations in Responding to Privacy Breaches*, OPCC (Aug. 2007), http://www.priv.gc.ca/information/guide/2007/gl_070801_02_e.asp ("What is the context of the personal information involved? For example, a list of customers on a newspaper carrier's route may not be sensitive. However, the same information about customers who have requested service interruption while on vacation may be more sensitive.").

[434] van den Hoven, *supra* note 1, at 312.

[435] *Id.*

[436] *See* EC Resolution (73) 22, *supra* note 12, at para. 19; *see also* EC Resolution (74) 29, *supra* note 12, at Annex 3 (referring to electronic data processing that "may lead to unfair discrimination").

[437] *2010 Consumer Privacy Consultations*, *supra* note 335, at 10–11; *see* Poullet et al., *supra* note 58, at 24.

[438] Anthony Danna & Oscar H. Gandy, Jr., *All That Glitters is Not Gold: Digging Beneath the Surface of Data Mining*, 40 J. BUS. ETHICS 373, 380–81 (2002). Some refer to this growing problem as first-degree price discrimination, a practice where organizations attempt to perfectly exploit the differences in price sensitivity between consumers.

[439] Poullet et al., *supra* note 58, at 29.

[440] The U.K.'s Office of Fair Trading is conducting two market studies into websites using behavioral data to set customized pricing, where prices are individually tailored using information collected about the user's behavior. Julia Kollewe & Richard Wray, *Office of Fair Trading to Probe Use of Personal Data by Online Retailers*, THE GUARDIAN (Oct. 15, 2009; 5:58 PM),

Chris Jay Hoofnagle and Kerry E. Smith warn that information flows can be used to eliminate certain customers.[441] They claim that financial institutions may analyze and use information that they collect about their customers in order to target them for the purchase of products and services and that the data may potentially be used "to deny consumers choice or to steer them towards choices not in their best interest."[442] Classifying people in such a way that their chances of obtaining certain goods, services or employment are diminished may also illustrate this type of harm.[443]

A third type of objective harm is a physical one.[444] For example, individuals may become a victim of a crime against their person, in the event that their information (home or work address) are used by criminals such as stalkers and rapists.[445] The harm in question can be severe; a perfect example is the murder of actress Rebecca Schaeffer in 1989.[446] In the U.S. case *Remsburg v. Docusearch*,[447] a stalker killed a woman after obtaining her work

http://www.theguardian.com/business/2009/oct/15/retail-pricing-tactics-oft-investigation.

441 Chris Jay Hoofnagle & Kerry E. Smith, *Debunking the Commercial Profilers' Claims: A Skeptical Analysis of the Benefits of Personal Information Flows* 20 (Univ. of Cali., Berkley–School of Law, Berkley Center for Law & Technology, Working Paper 2003), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=504622; *see also id.* at 13–17 (examples of price increasing based on consumer profiles).

442 *Id.* at 19.

443 *See, e.g.*, van den Hoven, *supra* note 1, at 312 ("Being classified as Muslim in many Western countries implies a reduced chance of getting a job."); Tal Z. Zarsky, "*Mine Your Own Business!*": *Making the Case for the Implications of Data Mining of Personal Information in the Forum of Public Opinion*, 5 YALE J.L. & TECH. 1, 22 (2003) ("After [analyzing data], vendors have the ability to discriminate between consumers based on this profile.").

444 *See generally* van den Hoven, *supra* note 1, at 311 (noting that stalkers and rapists have used the Internet to choose victims). These types of uses (physical harms), together with fraud and identity theft, are of a criminal nature, and they are governed by criminal laws. When certain objective harms resulting from the use of personal information are found to be very significant for individuals, they are often governed by laws, other than DPLs, which address these harms specifically. Still, acknowledging that certain disclosures may be harmful because criminals may use the information is relevant when assessing the risk of objective harm (or in assessing if there is a risk upon disclosing this information).

445 *Id.*

446 It was discovered that her assailant located her home address through the records of the Department of Motor Vehicles. *See* Margan v. Niles, 250 F. Supp.2d 63, 68 (N.D.N.Y. 2003).

447 Remsburg v. Docusearch, Inc., 149 N.H. 148, 152–53, 816 A.2d 1001, 1005–06 (2003).

address from a data broker. Canadian breach notification provisions include "physical harm" in the definition of "significant harm."[448]

1. Risk of Objective Harm: Identifiability Replaced by Negative Impact

In light of the objective harm, whether financial, discriminatory, or physical, linked to the use of personal information, there are two central outcomes that I will elaborate on. First, I maintain that the only relevant criteria when assessing whether the use of certain information should be governed by DPLs is whether the use of the information will potentially create an objective harm to the individual concerned (instead of whether the individual is "identifiable"). Once it is determined that the use of information triggers an objective harm, then the information used should be "accurate" and "relevant" to the intended use. As a matter of fact, DPLs usually provide for a data quality principle under which personal information shall be as accurate, complete, and up-to-date as is necessary for the purposes of its ultimate use.[449] DPLs also usually prohibit the collection of information that is not "necessary" for the intended use.[450]

Second, I maintain that the relevant criteria when establishing the *risk of harm* generated by the use of data (objective harm) are

448 PIPA, *supra* note 432, at 2.

449 *See, e.g.*, Personal Information Protection Act, S.A. 2003, c. P-6.5 art. 33 (Can.); An Act Respecting the Protection of Personal Information in the Private Sector, R.S.Q. 1993, c. P-39.1, s. 11 (Can.); Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, art. 2(1) (Can.); Personal Information Protection Act, S.B.C. 2003, c. 63, 33(a) (Can.); Directive 95/46, *supra* note 7, at 40.

450 Under PIPEDA, organizations shall collect only information "necessary" for the purposes identified and the data collected shall not be routinely updated "unless such a process is necessary to fulfill the purposes for which the information was collected." *See* Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, art. 39 (Can.). In Quebec, there is a similar principle and only the "information necessary" for the object of the file can be collected. *See* An Act Respecting the Protection of Personal Information in the Private Sector, R.S.Q. 1993, c. P-39.1, s. 5 (Can.). In Alberta and B.C., there is a more general "reasonableness test": an organization may collect personal information only for purposes that are reasonable. Personal Information Protection Act, S.A. 2003, c. P-6.5, art. 11 (Can.); Personal Information Protection Act, S.B.C. 2003, c. 63, 11 (Can.). Under Directive 95/46/EC, only relevant and non-excessive data may be processed. *See* Directive 95/46, *supra* note 7, at 40.

quite different than those relevant in the context of a disclosure (subjective harm). The criteria of "identifiability," "intimate" nature, and "availability" of the data, are pivotal when assessing the risk of harm at the "disclosure" level, and not relevant to assess whether there is an objective harm.[451]

While certain individuals may feel "uncomfortable" with the idea that certain data of an *intimate* nature may be used to take a decision which may have an impact on them, if the data is "relevant" for the purpose used, and is of "quality" in light of such purpose, individuals may have a hard time arguing that a certain use is illegal because harmful to them. For example, an applicant for a position with a pharmaceutical company where the individual will have access to narcotics, may be required to submit personal health records in order to demonstrate the lack of prior addiction to narcotics. This information will be a definitive factor in the hiring process, potentially triggering an objective harm for the individual if the individual is not hired. A health record may be considered as information of an "intimate" nature as well as "identifiable" information, but I argue that these criteria are not relevant when assessing the risk of objective harm. Instead, I maintain that the test to assess if there is an objective harm should instead focus on whether the data (health record) will potentially be used against the individual, in which case the information (health record) must be "relevant" and "accurate" for the purpose of assessing the applicant's candidacy.

The "availability" of the information is relevant when assessing the risk of harm at the disclosure level, however, it is not relevant when assessing the risk of harm at the "use" level. In this Information Age, and with new technology and tools on the Internet, there is a considerable amount of information already at our fingertips. For example, a pharmaceutical company evaluating the employment application of an individual who will have access to narcotics may want to verify certain information pertaining to the applicant's credentials with information available online. The bank may access an old resume made available online on LinkedIn which may not pass the "accuracy"

[451] The picture is flipped when we are assessing the risk of harm at the "disclosure" level. In the event that the data is "disclosed," the criteria of "relevancy" and data "quality" (which are important to take into account in the presence of an objective harm) are much less important to assess the risk of subjective harm.

test, because it may not be up-to-date. The company may also access certain compromising pictures of the applicant partying through Facebook, but this employer may have a hard time actually using these pictures unless they pass the "relevancy" test. Bottom line, it is not because information is publicly available, that it can be used unconditionally, as the data also has to be relevant and accurate for the intended use as discussed earlier.

Information usually has to be able to "identify" an individual to qualify as *personal* under DPLs.[452] I argue that this criterion of "identifiability" is much less relevant when assessing if there is an objective harm upon information being used, and that this metric ("identifiable") should instead be replaced by the following: whether the information used may have a "negative impact" (objective harm) on the individual. As a matter of fact, information may in certain cases be "used" by organizations for various purposes which may have no impact whatsoever on an individual, a very indirect and limited impact, or even a positive one. I argue that in such cases, the information should not be governed by DPLs.[453]

According to older as well as more recent documents (including DPLs), the central concern behind regulating the use of information has to do with the awareness of potentially negative impacts on the data subjects (objective harm).[454] A number of provisions or principles lead us to this conclusion.

DPLs were to apply to information "used" in such a way, which would have an impact on the individuals.[455] For instance, the Lindop Report (U.K. 1978) mentioned that: "The objective of discarding outdated data clearly applies principally to data bases used for making decisions which affect the data subject . . . ."[456] Many recent DPLs provide that the information should be accurate, especially if it will be used in such a way, which will

---

[452] *See* Stratford & Stratford, *supra* note 3, at 17–18.

[453] Or at least, that this information should be able to circulate without having to obtain the individual's prior consent.

[454] *See* Personal Information Protection Act, S.A. 2003, c. P-6.5, art. 34.1 (Can.); *see also* Personal Information Protection Act, S.B.C. 2003, c. 63, 21(1)(c) (Can.).

[455] *See Data Protection*, OUT-LAW, http://www.out-law.co/page-413 (last updated Feb. 2008) (stating that DPLs strike a balance between the rights of individuals to privacy and the ability of organizations to utilize such data for business purposes).

[456] LINDOP, *supra* note 12, at 51.

create a negative impact on the individual.[457] For example, PIPEDA provides that organizations should avoid that "inappropriate information . . . be used to make a decision about the individual."[458] Under the Civil Code of Quebec, any person may examine and cause the rectification of a file kept on him by another person "with a view to making a decision in his regard or to informing a third person."[459] Under the B.C. DPL, an organization must make a reasonable effort to ensure that personal information collected by or on behalf of an organization is accurate and complete, "if the personal information is likely to be used by the organization to make a decision that affects the individual to whom the personal information relates."[460] The Directive 95/46/EC on the subject matter, has a similar provision: any individual is entitled to interrogate the data controller of his personal data in order to obtain information allowing him to know and to object to the reasoning involved in the automatic processing, "in the case of . . . automated decisions . . ." that have legal effects on the data subject."[461]

These provisions were clearly meant to ensure that when personal information is used in assessments or decisions that may have a negative impact on an individual (what I refer to as an objective harm), the data in question should at least be accurate.[462] It is interesting to note that under the Directive 95/46/EC, the decision has to either produce "legal effects" for, or "significantly affects," an individual.[463] This means that there is an argument to be made that perhaps an organization using personal information, which triggers a non-significant impact for an individual should not be regulated in all instances by DPLs and a positive impact, even less.

It is not always clear whether new types of data are covered under the definition of *personal information*, as discussed in Part

---

457 *See* Personal Information Protection Act, S.A. 2003, c. P-6.5, art. 33 (Can.); *see also* Personal Information Protection Act, S.B.C. 2003, c. 63, 33 (Can.) (Alberta's and British Columbia's DPLs both explicitly state that organizations must make reasonable efforts to ensure the accuracy of any personal information used, collected or disclosed).

458 Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, principle 4.6.1(Can.)

459 Civil Code of Québec, S.Q. 1991, c. 64 art. 38 (Can.).

460 Personal Information Protection Act, S.B.C. 2003, c. 63, 33(a) (Can.).

461 Directive 95/46, *supra* note 7, at 42.

462 *Id.*

463 Directive 95/46, *supra* note 7, at 43.

II.A.3.[464] The Article 29 Working Party commented on this very issue as it relates to RFID tags, noting that data relates to an individual, "if such information is used to determine or influence the way in which that person is treated or evaluated."[465] This further confirms that information (including new types of data) should only be covered by DPLs if their use creates an impact on individuals.

As far as DPLs are concerned, protecting against the use (or misuse) of personal information has everything to with protecting against the risk of objective harm. Clearly, however, this risk only becomes a factor when the information is used *to the detriment* of the data subject in question. The parliamentary debates leading to the adoption of the Quebec DPL in 1993 confirm that this particular DPL was initially to focus on regulating uses which would have a "negative" impact on individuals.[466] The "negative" criterion was eventually abandoned in the final wording of the law, since there was a concern that organizations would argue that certain "uses" were not negative, they were only potentially so.[467] In Europe, it is interesting to note that the 2002 Proposals for Amendment to Directive 95/46/EC suggested redefining the scope of the provision pertaining to the use of the information in terms of acts of data processing that include "any kind of discriminatory practice."[468]

The purpose of DPLs regulating the activity of using personal information was not to address situations or uses having a positive impact for the individual, as illustrated by van den Hoven:

> They do not mind if their library search data are used to provide them with better *library* services, but they do mind if these data are used to criticize their taste and character. They would also object to these informational cross-contamination when they would benefit from them, as when the librarian would advise them a book on low-fat meals on the basis of knowledge of their medical records and cholesterol values, or when a doctor asks questions on the basis of the information that one has borrowed a book from the

[464] *See* discussion, *supra* Part II.A.3.

[465] Article 29 Working Party Working Doc. on RFID Tech., *supra* note 75, at 8.

[466] *See* Commission Permanente de la Culture, Les travaux parlementaires, 34th légis., c. 13 at 23 (1993).

[467] Therefore, the word "negative" was removed to avoid any uncertainty pertaining to this issue. *See id.*

[468] *See* ROBINSON ET AL., *supra* note 55, at 28.

public library about AIDS.[469]

Certain DPLs even authorize the use of personal information by organizations, without obtaining prior consent, if such use is in the interest of the individuals concerned.[470] For instance, the public sector Quebec DPL (*An Act Respecting Access to Documents Held by Public Bodies and the Protection of Personal Information*)[471] states that a public body may use personal information for a new purpose without the consent of the individual if the information is clearly used for the benefit of the person to whom it relates.[472] Under the Alberta DPL, an organization may use personal information without prior consent, if "a reasonable person would consider that the use of the information is clearly in the interests of the individual and consent of the individual cannot be obtained in a timely way or the individual would not reasonably be expected to withhold consent."[473] The British Columbia DPL has a similar requirement.[474] In Europe, personal data may be processed (or used) if it is deemed necessary for the performance of a contract to which the data subject is party or in order to protect the vital interests of the data subject.[475]

The fact that information can be used with no impact for individuals is also addressed in certain documents or DPLs.[476] For example, the Lindop Report (U.K. 1978) mentioned that: "several witnesses told us that users should not be prevented from retaining personal records for statistical, research and archival purposes . . . ."[477] As early as 1972, when DPLs were in their infancy, it was clear that a great deal of personal information would be useful to provide statistics to assist

469 van den Hoven, *supra* note 1, at 314.

470 An Act Respecting Access To Documents Held By Public Bodies And The Protection of Personal Information, R.S.Q. c. A-2.1, s. 37 div. 65.1 (Can.); Personal Information Protection Act, S.A. 2003, c. P-6.5, § 17(a) (Can.).

471 An Act Respecting Access to Documents Held by Public Bodies and the Protection of Personal Information, R.S.Q., c. A-2.1 (Can.).

472 *Id.* at c. A-2.1, s. 37 div. 65.1.

473 Personal Information Protection Act, S.A. 2003, c. P-6.5, § 17(a) (Can).

474 *Id.* at § 15(1)(a).

475 See Directive 95/46, supra note 7, at 40 (discussing the only situations in which personal data may be processed).

476 *See, e.g.*, *Personal Data Protection Act: Overview*, PERS. DATA PROT. COMM'N SING. (June 10, 2013), http://www.pdpc.gov.sg/personal-data-protection-act. (explaining Singapore's DPL).

477 LINDOP, *supra* note 12, at 43.

planning and other research.[478] Since researchers (or information gatherers in general) rarely needed to know the identity of their data subjects, the anonymization of data was seen as a natural solution.[479] Now that it is less and less clear at what point data is in fact anonymized,[480] I argue that we should focus on protecting information that may trigger a risk of harm upon being used or disclosed.

More recently, Canadian and European DPLs provide for certain exceptions for processing or using data for historical, statistical or scientific purposes, provided that certain appropriate safeguards are complied with.[481] These types of provisions may support the argument that only information used to impact an individual negatively (objective harm) should be governed by DPLs.

In Part II.A.2., I discuss how the definition of personal information and the notion of "identifiable" can trigger an under-reaching outcome or can be viewed as obsolete.[482] More specifically, with the advent of certain new technologies, decisions can be made that will exert a palpable impact on the owner of an online profile, without even needing to identify the individual behind the profile (identifying meaning by face, name or address). Therefore, at the "use" level, the metric of whether data "identifies" the individual should be replaced by whether the use of the information may create an objective harm on the individual. My approach is in line with van den Hoven's views, who believes that "the referential reading of 'personal data', 'identity' and 'identifiability' of the [European DPLs may] lead[]

[478] REPORT OF THE COMM. ON PRIVACY, *supra* note 287.

[479] *Id.* at 183 ("A great deal of personal information is acquired to provide statistics to assist planning and other research, or is acquired for some other purpose and subsequently adapted to a form suitable for such ends. Planners and researchers, however, rarely need to know identities of individuals. Therefore, in computerised systems handling information for statistical purposes, adequate provision should be made in their design and programs for separating identities from the rest of the data.").

[480] *See* discussion *supra* Part II.A.3.

[481] *See supra* notes 243–45 and accompanying discussion; *see also* Directive 95/46/EC, *supra* note 7, at 42 ("Subject to adequate legal safeguards, in particular that the data are not used for taking measures or decisions regarding any particular individual, Member States may, where there is clearly no risk of breaching the privacy of the data subject, restrict by a legislative measure the rights provided for in Article 12 when data are processed solely for purposes of scientific research or are kept in personal form for a period which does not exceed the period necessary for the sole purpose of creating statistics.").

[482] *See* discussion *supra* Part II.A.2; GRATTON, *supra* note 15, at 108.

to unduly [harsh] constraints on the use of *personal data*" and that "[a]s a result, attributively used descriptions could go unprotected."[483] According to him, one way to ensure that information that should be protected actually is (and therefore avoiding an under-inclusive interpretation of the definition of *personal information*) would be to focus on "Identity Relevant Information."[484] This argument is very telling in the context of the "use" of information to the point where I maintain that the notion of "identity" should perhaps not even be taken into account when evaluating a piece of data or data sets that are being "used." Instead of defining the object of protection in terms of referentially used descriptions, van den Hoven articulates the view that we need to define the object of protection in terms of the broader notion of "identity relevant information":

> 'The owner of a blue Ford living in a postal code area 2345' could have more than one individual satisfying the description, and the user of these descriptions may not have a particular individual in mind; he just thinks about the owner of a blue Ford 'whoever he is.' 'The owner of a blue Ford,' however, could also be used referentially, when we have a particular person in mind or in attendance. 'The man sipping his whisky' (pointing out to the person at a party) is used referentially, and is *about* the person the speaker mistakenly thought was drinking whisky, even when it turns out he is having apple juice instead of whisky, and there is, strictly speaking, no one over there sipping his whisky.[485]

Following the proposed approach, data would fall under the definition of *personal information*, regardless of whether or not it is identifying, if it could potentially trigger an objective harm for an individual upon being used; for instance, if the use has a focus on a specific individual.[486] The following example illustrates my point:

> One may open a mental or another type of file on a person under the label 'the murderer of Kennedy', in the same way crime investigators do, in the hope to find out more information about this person who ever he is, or turns out to be. These initially nondescript identifications may eventually lead to a physical

---

[483] van den Hoven, *supra* note 1, at 310 (emphasis added).

[484] *Id.* More specifically, van den Hoven suggests, that, "given the prominence and importance of identity management technology, RFID technology, profiling and data mining, and genetic profiling, we need to have a new look at the dominant referential interpretation of personal data." *Id.*

[485] *Id.* at 309–10.

[486] *Id.* at 310.

> encounter (i.e., arrest or interrogation) later. The history of a particular criminal investigation is at the same time the history of filling the file with identity-relevant information.[487]

In the above situation, since the information collected under "the murderer of Kennedy" is done in the hopes of eventually being able to arrest or file criminal charges against the right person (i.e., using this information in such a way which may trigger a negative impact to the individual concerned), it should be treated and considered as *personal information*.[488]

Schwartz and Solove, in a recent article, point out the distinction between "identified" and "identifiable" individual applied in the context of behavioral marketing and discuss the fact that if an individual can reasonably be capable of being "singled out" from others, then we should consider that the information at stake qualifies as *personal information*.[489] On January 25, 2012, the EC published its long-awaited legislative package to reform EU data protection rules.[490] On October 5, 2012, the Article 29 Working Party issued its Opinion indicating that the proposed reform did not "fundamentally change the notion of personal data as [currently] defined in the Directive 95/46/EC" and it suggested clarifying in Recital 23 and Article 4 of the proposed reform that personal data should also cover "any information allowing a natural person to be singled out and treated differently."[491] The proposed framework would ensure that organizations taking decisions with regards to profile information (in certain cases, using new types of data), regardless of whether these profile are "identifiable" or not, would have to comply with DPLs, if these decisions may trigger

---

[487] *Id.*

[488] *Id.* This means that some of the obligations provided by DPLs should apply to this data, such as ensuring that the data is "accurate" before using it, and also "relevant" for the intended use.

[489] Schwartz & Solove, *supra* note 8, at 1887–88 ("*Identified* information is present when a person's identity has been ascertained, or when there is a substantial risk of identification of a specific individual. In contrast, *identifiable* information exists when such a specific identification, while possible, is not significantly probable.").

[490] *Commission Proposes a Comprehensive Reform of the Data Protection Rules*, EUR. COMM'N (Jan. 25, 2012), http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm.

[491] *See* Article 29 Working Party Opinion 08/2012, *supra* note 82, at 5. The Working Party also recommends changing Recital 24 to explicitly consider IP addresses and cookies as personal data. *Id.* at 5–6.

an objective harm for the individuals involved.[492] Therefore, according to the Article 29 Working Party, as long as an individual behind the profile is "singled out," and this person's characteristics or behavior are used to "influence that person" (what I call a "negative impact" or an objective harm), then the information should qualify as *personal information*.[493]

Organizations may be using information relating to a small group of individuals, for example, individuals that are using the same device. As discussed in Part II.A.3, one issue that requires more attention is when a small group of individuals are sharing the same device: Does the negative impact have to be linked to a *unique* individual?[494] I maintain that the extent of the objective harm should be taken into account when evaluating the data. More specifically, the more impactful or negative the risk of objective harm for the individual upon the information being used, the less important the notion of "identifying" (or having the device link to) a "unique" individual versus a small group of individuals should actually be.[495] In certain situations, a group of individuals may be discriminated against by a given organization using their personal information. For example, an insurer may wish to refuse all clients living in a certain geographical area. Although, in this case, there would be an objective harm resulting from the use of personal information, the underlying issue may be outside the scope of DPLs, since a group of individuals are discriminated against (instead of a unique individual).[496]

In the event that the use of personal information has no impact on an individual or exerts a positive one (i.e., is beneficial), I maintain that organizations would not need to ensure compliance with DPLs on the notice and choice aspects. Nevertheless, they

---

492 *See* ERIKA MCCALLISTER ET AL., NAT'L INST. OF STANDARDS & TECH., GUIDE TO PROTECTING THE CONFIDENTIALITY OF PERSONALLY IDENTIFIABLE INFORMATION (PII) at B-6 (2010) (suggesting that the individual harms include discrimination, financial or physical harm).

493 Article 29 Working Party Opinion 08/2012, *supra* note 82, at 5.

494 *See* discussion *supra* Part II.A.3. *See also* Dolin, *supra* note 55, at 149.

495 INFO. COMM'R OFFICE, *supra* note 245, at 9. For example, if insurance services are being refused to an online profile, the fact that this profile is linked to a device that may be used by more than one individual should not be taken into account and the profile information used should qualify as *personal information* since the use is triggering a negative impact (objective harm) on one individual behind the profile.

496 *See* Schwartz & Solove, *supra* note 8, at 1869 ("Privacy rights should attach when data pertain to particular people.).

would need to ensure that the data is stored securely in order to avoid any subjective harm resulting from disclosures, as the case may be.[497] On the other hand, in the event that the use of personal information has a negative impact on the individual, then the information used should qualify as *personal information* and full notice, access, and correction rights (data accuracy), should be granted to the affected individuals.[498]

Certain industries will use personal information in harmful ways towards individuals and discriminate, and this is normal and acceptable to society to a certain extent. For example banks will refuse to give credit (loan, mortgage, etc.) to those who don't have a good financial track record. Employers will refuse to hire individuals who don't have the requisite credentials for a given job, etc. Ron A. Dolin discusses how discrimination is, to a certain extent, necessary to ensure the viability of our financial system.[499] DPLs were meant to ensure that individuals, if discriminated against (objective kind of harm), for instance when being refused employment, insurance coverage, or a loan or credit, that this discrimination (or the financial harm which they sustain) be based on data meeting the "relevancy" and "accuracy" criteria; meaning data which is up to date and accurate.[500]

2. Applying the Approach to New Types of Data

Under the proposed approach, if information is used in such a way that there is no impact for the individual concerned or that the impact is positive, then the information should not be governed by the relevant DPL (i.e., no disclosure and consent necessary, no need to provide access to this information, etc.). On the other hand, if information is used in a way, which may have a negative impact on the individual (objective harm), then the data should be governed by the relevant DPL (and therefore, notice and consent would be necessary). If the data will be used in a

[497] *See id.* at 1881 ("Data security for identifiable information, as for identified information, should be commensurate with the nature of the information and the likely risks of disclosure.").

[498] While I realize that this may be challenging for organizations that wish to apply impactful decisions on profiles, I believe that the approach proposed would comply with the purpose behind regulating the activity of "using" personal information in DPLs.

[499] Dolin, *supra* note 55, at 161 (using the example of an individual with adverse credit and the benefit of that information being accessible to credit companies in order to protect and maintain an efficient financial system).

[500] For details on these criteria, *see* GRATTON, *supra* note 15, at 363, 383.

negative impactful way, it will also have to comply with the data "accuracy" and the "relevancy" tests.[501] I will illustrate how the proposed approach would work in practice, by applying the approach to new types of data.

a. IP addresses, Log files, Cookies

Organizations active in the online space are collecting new types of data and using new types of collection tools and using the data collected for various purposes.[502] Information collected using IP addresses or cookies may be used in such a way as to create no impact for individuals (or a positive impact). For example, the information may be collected and used to improve user experience on the website; for instance to remember what is in the user's shopping cart or to remember the language of preference etc. I maintain that the purpose behind the use will determine if the information used qualifies as *personal information* and is therefore governed by DPLs.

If the profile is used by the organization to determine which web pages to present to the user (in the right language, etc.), then it may not have a negative impact for the individual. If the information is used to present personalized advertising (behavioral marketing) which takes into account prior searches or purchase history from this specific user (instead of random advertising), then it could also be arguable that there would be no objective harm and that the information would not be subject to DPLs, subject to the concerns which related to certain information being disclosed.[503] If certain information will be used to categorize this individual and the individual will be refused certain services based on his or her profile, then the use would trigger an objective harm (since the impact would be a negative one). The user should therefore promptly be made aware of the information use, in order to be in a position to consent to such

[501] *See supra* notes 450–51. (elaborating on the various provisions found in Canadian and European DPLs which usually require that information *used* be "accurate and "relevant" for its intended purpose). *See also* GRATTON, *supra* note 15, at 363, 383.

[502] GRATTON, *supra* note 15, at 41, 42, 46.

[503] *See id.* at 317 (discussing how, in certain situations, behavioral marketing techniques may be assimilated to a disclosure of personal information, if advertising using otherwise unavailable information of intimate nature is displayed to an anonymous user behind an IP address, for instance, if this computer is shared amongst a small group of people such as family members or co-workers).

collection and use and be in a position to determine whether the information used by the organization complies with the data "quality" or "relevancy" tests.[504]

b. Search Queries

Search engines allow web users to find information pertaining to the topic that they are searching.[505] Search queries may be useful for various potential uses other than simply processing the individual's search request.[506] Some of these uses may have a positive impact for the individual, a negative impact (objective harm), or no impact whatsoever.

First, search queries may be used in such a way that may create a negative impact (objective harm) for the individual. For example, an insurance company could have the intention to raise insurance premiums by 5% upon finding out that a certain individual has researched a number of books on a particular type of cancer. Since this would constitute a use triggering an objective harm for the individual, the data used in this context would be governed by the relevant DPL.[507]

Search engines may collect and use search query data for purposes, which may benefit individuals.[508] A registered search history may be used to reduce irrelevant advertising; it can help differentiate ambiguous terms on an individual basis (e.g., jaguar—car vs. cat); help with personalized spell corrections and term substitution; indicate which languages someone has used (football vs. soccer); and aid in determining appropriate levels of filtering for profanity (sexual content, etc.).[509] Google mentions that it is collecting search queries for various purposes, including keeping their services secure,[510] their users safe from *malware* or

[504] *See supra* notes 450–51 (elaborating on the various provisions found in Canadian and European DPLs which usually require that information *used* be "accurate and "relevant" for its intended purpose).

[505] *See* Dolin, *supra* note 55, at 137 (listing the kinds of searches people might use search engines to conduct).

[506] *Id.* at 142.

[507] This means that the insurance company would have to inform the user of this use, the consent of the individual prior to such use would be necessary and the data used would have to comply with the "accuracy" and the "relevancy" tests. *See supra* notes 426–27.

[508] Dolin, *supra* note 55, at 142.

[509] *Id.*

[510] *See* Matt Cutts, *Using Data to Fight Webspam*, GOOGLE OFFICIAL BLOG (June 27, 2008), http://googleblog.blogspot.com/2008/06/using-data-to-fight-webspam.html (discussing Google's use of search logs to combat webspam).

phishing attacks,[511] and to detect and prevent advertising "click fraud."[512] All of these uses could be considered as having a potentially positive impact for an individual, as they would improve the user's experience, increase the effectiveness of his or her web search and protect him or her against certain unwanted viruses or content.[513] These types of uses being harmless to individuals, I maintain that search query information strictly used for these purposes should not be governed by the relevant DPLs.[514]

Search query data may also be used for purposes that may have no impact for individuals. For instance, this kind of data may be used by search engines to improve their search algorithms and the quality of their search services.[515] In such cases, search engine data would be used to provide knowledge for the organization, and potentially, such uses would have no direct impact for individuals. An example illustrating how search engine data may be used is with Google Flu Trends,[516] a service provided by Google since 2008, which furthers early detection of influenza epidemics throughout the world by monitoring health-seeking behavior, specifically the online web search queries that millions of individuals submit to the Google search engine each day.[517] In 2009, Hal Varian, Google's chief economist, published a paper showing that Google searches can also be used to predict a

511 Niels Provos, *Using Log Data to Help Keep You Safe*, GOOGLE PUB. POL'Y BLOG (Mar. 13, 2008), http://googleblog.blogspot.com/2008/03/using-log-data-to-help-keep-you-safe.html.

512 *See* Article 29 Working Party Opinion 1/2008, *supra* note 138, at 15.

513 Unless we consider that personalized advertising may instead create a negative impact for the individual as it may reduce the choices offered to the individual or be used to discriminate against the individual or profile. *See* GRATTON, *supra* note 15, at 343–45.

514 Although if the information collected by the search engine creates a risk of subjective harm upon being disclosed, for instance in the context of a security breach, then the information collected would be considered as *personal information*. This translates into individuals having to be made aware that the search engine is collecting and storing this information (which is of an "intimate" nature, not "available" and "identifiable"), and agreeing to such collection and storage, etc.

515 Hal Varian, *Why Data Matters*, GOOGLE OFFICIAL BLOG (Mar. 4, 2008) http://googleblog.blogspot.com; Shuman Ghosemajumder, *Using Data to Help Prevent Fraud*, GOOGLE OFFICIAL BLOG (Mar. 18, 2008) http://googleblog.blogspot.com; *see also* Provos, *supra* note 511.

516 *Explore Flu Trends Around the World*, GOOGLE.ORG, http://www.google.org/flutrends (last visited Sept. 17, 2013).

517 Dolin, *supra* note 55, at 143.

bevy of economic data, including retail sales[518] and unemployment claims.[519] I maintain that the use of search queries for the purpose of providing this kind of service has no negative impact for individuals (potentially no impact and perhaps even a positive one) and that therefore, search queries used for this purpose should not be subject to the DPL's requirements of obtaining the prior consent of individuals.[520]

c. Location Information

Location information may be collected using various methods, and for different purposes.[521] Organizations managing trucks or taxis may use location-tracking technology to track their vehicles strictly for fleet management purposes. In such cases, since this use triggers no impact for individuals, this information should not be considered *personal information* governed by DPLs.[522] An organization may also track the location of a vehicle for security purposes (knowing that a certain truck containing valuable merchandise is at a given location). Information used for this purpose would also not be governed by DPLs, since it may have no impact for the individual (or even potentially a positive impact, i.e., for the security of the driver). In the event that the location data is used to evaluate an employee (such as the driver of a taxi or a truck) and to potentially reprimand this employee, then the information would be subject to the relevant DPL as this use of information may trigger a risk of objective harm. This means that employees would have to be informed of this collection and use, and consent to it, and the location information would have to be "accurate" and "relevant" for the intended use.

[518] Hal Varian & Hyunyoung Choi, *Predicting the Present with Google Trends*, GOOGLE RESEARCH BLOG (Apr. 2, 2009), http://googleresearch.blogspot.com.

[519] Hal Varian & Hyunyoung Choi, *Predicting Initial Claims for Unemployment Benefits,* GOOGLE RESEARCH BLOG (Apr. 22, 2009), http://googleresearch.blogspot.com.

[520] Unless the information may trigger a risk of subjective harm upon being disclosed. *See* discussion *supra* Part III.A.1.

[521] ÉLOÏSE GRATTON, INTERNET AND WIRELESS PRIVACY: A LEGAL GUIDE TO GLOBAL BUSINESS PRACTICES 24–29 (CCH Canada, 2003).

[522] Again, if the data, upon being "disclosed," may create a risk of subjective harm, then it would be governed by the relevant DPL and notice would have to be provided to the relevant individuals, consent would have to be obtained, etc.

IV. CONCLUSION

The ultimate purpose of DPLs was to protect individuals against a *risk of harm*, which may take place upon their information being collected, used, or disclosed. Therefore, in the approach proposed, I maintain that information should only qualify as *personal* if the information upon being collected, disclosed or used creates such risk. Since each data handling activity triggers different sets of concerns, I have analyzed them separately, to come to the realization that while certain data handling activities such as the collection and disclosure of information trigger a more subjective kind of harm to individuals, the use of this information usually triggers a more objective kind of harm. In Part III,[523] I propose a decision tree which may be used in determining which data is or should be covered by DPLs and more specifically, what are the risks associated with the collection, use or disclosure of data.

I maintain that information collected creates problems often through its use or disclosure. The collection "per se" or the means by which personal information is gathered is an activity that is not as efficiently regulated by DPLs. I argue that we should focus on the risks of harm, which may take place at the "disclosure" and "use" levels. Therefore, upon information being collected, the analysis which should take place in order to determine whether the information collected is *personal*, is whether the information collected may create a *risk of harm* upon being disclosed (for instance in the context of a security breach) or upon being used, in which case it will qualify as *personal information*. This translates in data *collected* only having to be disclosed to individuals (and their consent having to be obtained) if the data creates a risk of harm at the "disclosure" or "use" levels.

When the data is to be *disclosed*, to determine if the data qualifies as *personal information*, there should be an assessment of whether the disclosure will create a *risk of subjective harm* to the individual. This subjective harm can be assimilated to a feeling of being embarrassed or uncomfortable upon the information being disclosed. I argue that the notion of "identifiable" individual should be interpreted in light of the overall sensitivity of the information, therefore in light of two additional criteria. More specifically, the test would include

[523] *See* discussion *supra* Part III.

whether the data is "identifiable" (the more identifiable to a unique individual, the higher the risk of harm), whether this data is of "intimate" nature (the more intimate, the higher the risk of harm), and the extent of its "availability" (the less it was previously available prior to the disclosure or the more available it may become post disclosure, the higher the risk of harm). If the data to be disclosed result in a very low risk of harm to the individual (the data is not of "intimate" nature, it is not "identifiable" to a unique individual or small group of people, and it is already very widely or publicly "available"), then the information should not qualify as *personal information*, and the disclosure of the data would therefore fall outside of the scope of DPLs.

When the data is to be *used*, to determine if the data qualifies as *personal information*, the test to follow would indicate to determine if the data used will have an impact on the individual and if so, a negative one. If there is no impact for an individual or the impact is positive, then I maintain that the data should not qualify as *personal information* and it can be used without further restrictions, as this data was not meant to be covered by DPLs. If there is a negative impact (or what I refer to as an objective harm, such as a financial harm, physical harm or some type of discrimination), then the information would qualify as *personal information* and it would have to be "accurate" and "relevant" for the intended use. If the information is not accurate and relevant, then the data should simply not be used for the purpose intended. The fact that the information *can or can't identify a unique individual* does not need to be taken into account at the use level and may usually not need to be included in the assessment test at that point.

A first objective of the present work is to come to a common understanding of the notion of *personal information*, the situations in which DPLs should be applied, and the way they should be applied. Working on a common interpretation of the definition of *personal information* is tantamount to defining what falls inside or outside the scope of DPLs. Another objective is to assist organizations in determining which data handling activities are harmful in order to guide them when there are looking to comply with the various "reasonable" tests under DPLs. For example, under PIPEDA, an organization may only collect, use or disclose personal information "for purposes that a reasonable person would consider appropriate in the

circumstances."[524] The proposed approach may be useful to determine whether a given data handling activity is "reasonable," depending if this activity will create a risk of harm to the individual concerned. The approach proposed may also be useful when determining what type of privacy notice should be made to the individual and what type of consent should be obtained: for instance if the risk of harm is present but is on the "low" side, then perhaps a disclosure in the form of a privacy policy made available on the website and an opt-out type consent would be sufficient. DPLs usually also provide for subjectivity in the assessment of the security measures that have to be implemented by an organization. The approach may therefore be of assistance when determining what kind of security measures are "reasonable," "necessary," or "appropriate" in accordance with the relevant DPL.

A corollary of this work is to provide guidance to lawmakers, policymakers, privacy commissioners, courts, organizations handling personal information and individuals assessing whether certain information are or should be governed by the relevant DPLs, depending on whether the data handling activity at stake creates a *risk of harm* for an individual. The proposed approach is meant to provide for a useful framework under which DPLs remain efficient in light of modern Internet technologies.

[524] Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, art. 3 (Can.).