Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework

Zuiderveen Borgesius, F.; van Eechoud, M.; Gray, J.

Publication date
2015

Document Version
Final published version

UNIVERSITY OF AMSTERDAM

OPEN DATA, PRIVACY, AND FAIR INFORMATION PRINCIPLES: TOWARDS A BALANCING FRAMEWORK

Frederik Zuiderveen Borgesius

Mireille van Eechoud

Jonathan Gray

Amsterdam Law School Legal Studies Research Paper No. 2015-46

Institute for Information Law Research Paper No. 2015-04

OPEN DATA, PRIVACY, AND FAIR INFORMATION PRINCIPLES: TOWARDS A BALANCING FRAMEWORK[1]

Frederik Zuiderveen Borgesius, Jonathan Gray, Mireille van Eechoud [2]

Working draft

Please only refer to the published version, in the Berkeley Technology Law Journal

http://btlj.org

Open data are held to contribute to a wide variety of social and political goals, including strengthening transparency, public participation and democratic accountability, promoting economic growth and innovation, and enabling greater public sector efficiency and cost savings. However, releasing government data that contain personal information may threaten privacy and related rights and interests. In this paper we ask how these privacy interests can be respected, without unduly hampering benefits from disclosing public sector information. We propose a balancing framework to help public authorities address this question in different contexts. The framework takes into

1 Acknowledgements: We thank Simon Hania, Dariusz Kloza; Stefan Kulk, Maja Lubarda, Richard Rogers, Javier Ruiz, Nico van Eijk, Ben Worthy, and Bendert Zevenbergen for participating in the *Workshop Reconciling Fair Information Principles and Open Data Policies*, 6 February 2015, Institute for Information Law, Amsterdam. We also thank the participants of the symposium *Open Data: Addressing Privacy, Security, and Civil Rights Challenges,* 17 April 2015, Berkeley Center for Law & Technology, in particular Cathy O'Neil and David Flaherty. The thought-provoking discussions during both events helped to shape our ideas for this paper. Furthermore, Matthijs Koot, Bendert Zevenbergen, and the editors of the Berkeley Technology Law Journal deserve our gratitude for comments on earlier versions of this paper. We also express our gratitude to the members of the advisory board for the project that led to this paper: Simon Hania, VP Privacy & Security, TomTom International B.V.; Dr. Jaap-Henk Hoepman, Associate Professor of Privacy Enhancing Protocols and Privacy by Design, University of Nijmegen; Dr. Aleecia McDonald, non-residential fellow, Center for Internet & Society, Stanford University; Prof. B. Roessler, Professor of Ethics and its History, University of Amsterdam; Javier Ruiz Diaz, Policy Director, Open Rights Group; Prof. N.A.N.M. van Eijk, Professor of Media and Telecommunications Law, University of Amsterdam; Dr. Ben Worthy, lecturer in Politics at Birkbeck University of London, independent reporter for the UK's IRM of the Open Government Partnership (UK). We thank Sarah Eskens, Rachel Wouda, and Dirk Henderickx for research assistance. All errors are the authors' own. Financial support for this project came from the Berkeley Center for Law & Technology and Microsoft.

2 Frederik Zuiderveen Borgesius is post-doctoral researcher at the Institute for Information Law, University of Amsterdam Law School, Jonathan Gray is Research Associate at the Digital Methods Initiative, University of Amsterdam and Director of Policy and Research at Open Knowledge, Mireille van Eechoud is professor of Information Law at the Institute for Information Law, University of Amsterdam Law School (The Netherlands).

account different levels of privacy risks for different types of data. It also separates decisions about access and re-use, and highlights a range of different disclosure routes. A circumstance catalogue lists factors that might be considered when assessing whether, under which conditions, and how a dataset can be released. While open data remains an important route for the publication of government information, we conclude that it is not the only route, and there must be clear and robust public interest arguments in order to justify the disclosure of personal information as open data.

TABLE OF CONTENTS

I. INTRODUCTION

Open government data refers to data released by public sector bodies, in a manner that is legally and technically re-usable. The G8 Open Data Charter states: "free access to, and subsequent re-use of, open data are of significant value to society and the economy."[3] Open data are commonly held by its advocates to mean data that "can be freely used, modified, and shared by anyone for any purpose."[4] However, releasing public sector datasets that include personal information, or data that can be re-identified, may threaten privacy and related rights.

In this paper, we examine the tension between public sector open data policy and the Fair Information Principles (FIPs). The FIPs lie at the core of most data privacy laws around the world, including those in the E.U. and the U.S. The FIPs give guidelines to balance privacy-related interests and other interests, such as those of business and the public sector. The paper focuses on the following question: from the perspective of the Fair Information Principles, how can privacy and related interests be respected, without unduly hampering benefits from disclosing public sector information?

We rely mostly on desk research, using the usual sources for legal scholarship, such as legislation, soft law, policy documents, and literature. We use descriptive and analytical legal research to determine the main legal tensions between open data policy and the FIPs. Parts of the paper are more normative: we give recommendations to strike a balance that respects privacy and related interests, and not unduly hampers the benefits of open data.

---

3 UNITED KINGDOM CABINET OFFICE, G8 OPEN DATA CHARTER AND TECHNICAL ANNEX (2013).
4 *See* OPEN DEFINITION VERSION 2.0, http://opendefinition.org/ (last visited May 1, 2015). The Open Knowledge Foundation's first open definition dates from 2005, *see* OPEN DEFINITION, http://opendefinition.org/od/1.0/ (last visited May 1, 2015).

We enriched our research results with insights from a workshop, where we tested hypotheses and discussed the promises and pitfalls of privacy and open data. Conference participants came from academia, industry, civil society organizations, and data protection authorities, and were all working on issues in open data and privacy.[5] Discussions during the *Open Data: Addressing Privacy, Security, and Civil Rights Challenges* symposium of the Berkeley Center for Law & Technology also provided valuable insights.[6]

Furthermore, we conducted an empirical study into concerns that various stakeholders, in civil society, the public sector, research, and business, express about the interactions between privacy and open data. The study draws on document collections and digital traces from the web to map the debates about privacy and open data. The empirical study follows the "digital methods" approach, pioneered by Rogers and his colleagues at the Digital Methods Initiative.[7]

While each national legal system has its own traditions and characteristics, this paper focuses on common problems that arise in many jurisdictions. After all, as the Open Government Partnership testified, governments around the world create open data policies and must cope with privacy concerns.[8] Hence, we do not examine what sets jurisdictions apart, but instead discuss shared problems. For instance, we do not address specific requirements that follow from the First Amendment in the U.S.,[9] or from the fundamental right to data protection in the E.U.[10] Therefore, the paper's recommendations come with a caveat: they cannot be directly implemented in national legal systems.

The paper is structured as follows. Part II describes open data goals and privacy problems regarding open data. We clustered the objectives associated with open data into three categories: (i) innovation and economic growth, (ii) political accountability and democratic participation, and (iii) public sector efficiency. We identified three kinds of concerns about releasing personal information as open data: (i) the chilling effects on people interacting with the public sector, (ii) a lack of

---

5 Workshop 'Reconciling Fair Information Principles and Open Data Policies,' 6 February 2015, Institute for Information Law, Amsterdam.

6 *See* BERKELEY CENTER FOR LAW AND TECHNOLOGY,

https://www.law.berkeley.edu/centers/berkeley-center-for-law-technology/past-events/april-2015-the-19th-annual-bcltbtlj-symposium-open-data-addressing-privacy-security-and-civil-rights-challenges/ (last visited May 1, 2015.

7 *See generally* RICHARD ROGERS, DIGITAL METHODS (2013).

8 The OGP is an international platform for reform, to make "governments more open, accountable, and responsive to citizens." Participating states submit action plans in which they make commitments, inter alia on datasets to be made available as open data. Compliance and progress mechanisms are in place. Membership has grown to 65 countries in the five years since the OGPs inception. *See* OPEN GOVERNMENT PARTNERSHIP, http://www.opengovpartnership.org/ (last visited May 1, 2015).

9 *See* Daniel J. Solove, *Access and Aggregation: Public Records, Privacy and the Constitution*, 86 Minn.L.Rev. 1137 (2002), 1201.

10 On EU data protection law and public sector information re-use policy, *see* Cristina dos Santos et al.,*On Privacy and Personal Data Protection*, 6 (3) MASARYK UNIV. J.L. & TECH. (2012), 337, *available at* https://journals.muni.cz/mujlt/article/view/2613/2177 (last visited 15 May 2015); *see also* Mireille van Eechoud et al., LAPSI Position Paper on Access to Data, LAPSI (Dec. 12, 2014), http://www.lapsi-project.eu/sites/lapsi-project.eu/files/LAPSI_D2%202.pdf (last visited May 1, 2015).

individual control over personal information, and (iii) the use of open data for social sorting or discriminatory practices.

Part III discusses rules regarding access to information held by public sector. Freedom of information laws provide inspiration on how to strike a balance between privacy and transparency in the open data context. Part IV discusses the governance of personal information, focusing on the Fair Information Principles (FIPs). In this section we also discuss the main challenges in reconciling open data policy and the FIPs. From a FIPs perspective, the main problem with open data is that unrestricted re-use of personal data breaches the purpose specification principle. But we argue that there are possible compromise measures to balance privacy and open data interests.

We propose a balancing framework to accommodate privacy concerns and open data goals. Part V outlines the first element of the balancing framework, and distinguishes four data categories with different levels of privacy risks: (i) raw personal data, (ii) pseudonymized data, (iii) anonymized data, and (iv) non-personal data. Different modes of access and re-use control are the second element of the balancing framework. In many cases, disclosing data with access or re-use restrictions, rather than as fully open data, strikes a balance between open data goals and privacy (Part VI). As a third element of the balancing framework we provide a circumstance catalogue, a list of circumstances to consider when deciding whether or not a dataset should be disclosed, and under which conditions (Part VII).

Part VIII concludes that releasing personal information as fully open data is generally not appropriate. But sometimes a compromise can be found by disclosing data with access or re-use restrictions.

II. OPEN DATA AND PRIVACY

Open data are held to contribute to a wide variety of social and political goals. However, releasing data as open data may threaten privacy, for instance, if the open data contain personal information. Below we describe open data goals and privacy problems regarding open data. We clustered the objectives associated with open data into three categories: (1) innovation and economic growth, (2) political accountability and democratic participation, and (3) public sector efficiency. We also clustered privacy concerns in the area of open data into three categories: (1) the chilling effects on people interacting with the public sector, (2) a lack of individual control over personal information, and (3) the use of open data for social sorting or discriminatory practices.

A. OPEN DATA INTERESTS

Definitions of open data from technologists and civil society actors focus on enabling redistribution and re-use, and on limiting legal and technical barriers to re-use. For example, the summary of the "Open Definition" from Open Knowledge reads: "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)."[11] The full definition stipulates conditions that include legal openness,

11 OPEN DEFINITION, *supra* note 3.

bulk downloadability, and machine-readability.[12] Similar definitions are used in the *8 Principles of Open Government Data*,[13] the Sunlight Foundation's *Ten Principles for Opening Up Government Information*,[14] and the World Wide Web Consortium's *Five Stars of Linked Open Data.*[15]

Technical obstacles for re-using data include non-machine readable formats, proprietary formats, technological protection mechanisms,[16] and Digital Rights Management software. Legal restrictions on re-use include intellectual property rights, such as copyright and database rights.[17] When open data advocates say that "anyone can freely access, use, modify, and share [data] for any purpose,"[18] they are often referring to removing these specific kinds of legal and technical restrictions.

This conception of open data that focuses on limiting legal and technical restrictions for re-use has carried into public policy. Over the past decade, open data developed from being a niche idea at the margins of open source software, scientific research and hacker communities, into an idea with traction among public policymakers.[19] For example, the 2013 G8 Open Data Charter mentions that open data should be "machine readable," available in bulk, available in formats for which the specification is "available to anyone for free," and under open licenses such that "no restrictions or charges are placed on the re-use of the information for non-commercial or commercial purposes."[20] A similar focus on removing technical restrictions to re-use can be found in open data guidelines of the Organisation for Economic Co-operation and Development,[21] the UK Government,[22] and U.S. President Barack Obama.[23]

---

12 *Id.*

13 OPEN DATA WORKING GROUP, *The 8 Principles of Open Government Data*, OPENGOVDATA.ORG (Dec. 08, 2007), http://opengovdata.org/.

14 SUNLIGHT FOUNDATION, *Ten Principles for Opening Up Government Information,* (Aug. 11, 2010), http://sunlightfoundation.com/policy/documents/ten-open-data-principles/.

15 Tim Berners-Lee, *Linked Data*, W3.ORG (Jun. 18, 2009), http://www.w3.org/DesignIssues/LinkedData.html.

16 Technological protection mechanisms (TPMs) and digital rights management information are protected against circumvention and interference in their own right, separate from, e.g., copyright in the underlying work (database, software or other works). *See* Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979), art. 11-12 WIPO Copyright Treaty, December 20, 1996, http://www.wipo.int/wipolex/en/details.jsp?id=12214.

17 There is controversy about the role of intellectual property rights in implementing public sector open data, but this controversy is beyond the scope of this paper.

18 OPEN DEFINITION, *supra* note 3.

19 Jonathan Gray, Conference Paper, *Towards a Genealogy of Open Data,* General Conference of the European Consortium for Political Research in Glasgow, SOCIAL SCIENCE RESEARCH NETWORK (Sep. 3, 2014), *available at* http://dx.doi.org/10.2139/ssrn.2605828.

20 UNITED KINGDOM CABINET OFFICE, *supra* note 3.

21 Barbara Ubaldi, (2013), *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives* (OECD Working Papers on Public Governance No. 22, 2013), http://dx.doi.org/10.1787/5k46bj4f03s7-en. *See also.* Organisation for Economic Co-operation and Development [OECD], *OECD Recommendations of the Council for Enhanced Access and More Effective Use of Public Sector Information*, OECD Doc. C(2008)36 (2008), https://www.oecd.org/sti/44384673.pdf.

22 *Public Data Principles*, DATA.GOV.UK (Apr. 10, 2012, 5:43 PM), http://data.gov.uk/library/public-data-principles.

23 *See* Exec. Order No. 13,642, *Making Open and Machine Readable the New Default for Government Information*, 78 Fed. Reg. 28,111 (May 9, 2013), *available at* https://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government- [*hereinafter* Exec. Order, *Open and Machine Readable*].

Open data are held to contribute to a wide variety of social and political goals.[24] For ease of discussion in this paper, we have clustered the many objectives associated with open data into the following three areas: (i) innovation and economic growth, (ii) political accountability and democratic participation, and (iii) public sector efficiency. First we look at fostering innovation and economic growth.

1. Innovation and Economic Growth

Most official open data initiatives highlight the potential of enabling the re-use of public sector information to create new businesses and innovative services and products. Open data policies are increasingly becoming the preferred route to unlock the value of public sector information. This is evident from the European Commission's Guidelines on the Public Sector Information Directive.[25] The 2013 Executive order which aims to make Open and Machine Readable the New Default for Government Information views (federal) government information as a national asset and recognizes the importance of enabling widespread re-use for "economic growth and job creation."[26] The U.S. 2013 executive order on Open Data Policy adds: "making information resources accessible, discoverable, and usable by the public can help fuel entrepreneurship, innovation, and scientific discovery."[27] Similarly, the G8 Open Data charter claims open data are "a catalyst for innovation in the private sector, supporting the creation of new markets, businesses, and jobs."[28] The World Bank also recognizes this potential of open data.[29]

Information services built on public sector data are diverse. Financial services providers use official statistics as input.[30] Companies in the meteorological sector use weather data to provide highly specialized services, e.g. forecasts for off-shore oil industries.[31] Planning permissions, zoning data and housing data are combined with other sources to produce advice for customers such as real

---

24 *See, e.g.*, Gray, *supra* note 19.

25 *Commission Notice: Guidelines on Recommended Standard Licenses, Datasets and Charging for the Re-Use of Documents*, OJ 2014 C 240/1–10.

26 Exec. Order, *Open and Machine Readable, supra note 22*. The Order is one of several that follow up on open government policy announced in the Memorandum for the Heads of Executive Departments and Agencies on Transparency and Open Government, 74 Fed. Reg. 15 (Jan. 21, 2009).

27 Memorandum M-13-13, Open Data Policy – Managing Information as an Asset (May 9, 2013).

28 *See* Charter on open data signed by G8 leaders to promote transparency, innovation and accountability on 18 June 2013.
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf (last visited May 1, 2015).

29 WORLD BANK, OPEN DATA FOR ECONOMIC GROWTH, 5 (2014), *available at* https://openknowledge.worldbank.org/handle/10986/19997 License: CC BY 3.0 IGO.

30 See the examples of government information re-use in Assessment of the Re-use of Public Sector Information, report for the European Commission by MICUS, Brussels 2008 *and in* Makx Dekkers, Femke Polman, Robbin te Velde, Marc de Vries, Measuring European Public Sector Information Resources (MEPSIR), report for the European Commission, Brussels 2006, *37*.

31 *E.g.* the private company Meteogroup, see http://www.meteogroup.com/en/gb/sectors/marine.html (last visited May 15, 2015).

estate developers.[32] Postal codes are widely used as identifiers.[33] School and health inspection data serve as input for apps that help inform parents or patient choice.[34] Public transport timetable data when combined with geolocation data enable real-time and customized travel advice.[35] There are many other kinds of commercial exploitation of open data, often involving the combination of data from different public and private sources to deliver information products or services. The emphasis on economic benefits of re-using data held by public sector bodies predates open data policies. For example, in 1989, the E.U. sought to stimulate commercial exploitation of public sector data by the private sector.[36] The E.U. Public Sector Information Directive of 2003 also focused on public sector information as raw material for creating services and products.[37] The Directive obliged a wide range of public sector bodies to allow commercial and non-commercial re-use of their information assets, but not necessarily as open data.[38] Under the directive, conditions may be imposed, costs charged, and data may be made available in non-structured form. The U.S. Office for the Management of Budget also recognized federal information as a "commodity in the marketplace."[39]

Many studies have been commissioned to assess the value of public sector information; these studies suggest impressive figures, but range widely.[40] For example, the U.S. Department of

---

32 *E.g.* in Europe, the company Landmark provides such services and took the city of Amsterdam to court for the price it charged for re-use of city data. *See B&W Amsterdam v Landmark, Afd. Bestuursrechtspraak Raad van State* (29 Apr. 2009) *comment M.M.M. van Eechoud,* AMI, 6, 233 (2009).

33 For this reason the G8 Open Data Charter lists postal codes as 'high value' data, to be made available with priority. 33 *G8 Open Data Charter and Technical Annex*, *supra* note 3.

34 U.S. DEPARTMENT OF COMMERCE, FOSTERING INNOVATION, CREATING JOBS, DRIVING BETTER DECISIONS: THE VALUE OF GOVERNMENT DATA (2014);*id.*; MCKINSEY & CO. OPEN DATA: UNLOCKING INNOVATION AND PERFORMANCE WITH LIQUID INFORMATION 11 (2013).

35 *See* MCKINSEY & CO..*supra* note 33 at 6.

36 Guidelines For Improving the Synergy Between the Public and Private Sectors in the Information Market. DG XIII, European Commission, Brussels (1989).

37 Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, OJ 2003, L 345/90, as revised by Directive 2013/37/EC, OJ 2013, L175/1.

38 The *obligation* to allow re-use was introduced in the 2013 revision Directive 2013/37/EC, OJ 2013, L175/1, Member States must implement the revised directive by July 2015. The Directive builds on public access regimes in Member States; it does not regulate access directly.

39 Office for the Management of the Budget, Circular A-130 (revised). First issued in 1985, the Circular fostered (among many things) a larger role for the private sector in dissemination of government information and creating added-value (electronic) services. With subsequent revisions (1993-1996) under the Clinton administration the focus moved to release of electronic information by federal agencies directly to the public. For an overview of early policy development, *see* United States Congress Office of Technology Assessment, Informing the Nation: Federal Information Dissemination in an Electronic Age, USPO (1988).

40 For a recent example of a study on the economic value of public sector information at EU level, see MARC DE VRIES ET AL., PRICING OF PUBLIC SECTOR INFORMATION. MODELS OF SUPPLY AND CHARGING FOR PUBLIC SECTOR INFORMATION, FINAL REPORT (Oct. 2011), Study for the European Commission, DG Information Society. Brussels: Deloitte Consulting. Recent examples of studies about the value of open data and public sector information at national level include: U.S. DEPARTMENT OF COMMERCE, *supra* note 34; DELOITTE, MARKET ASSESSMENT OF PUBLIC SECTOR INFORMATION, STUDY FOR UK DEPARTMENT FOR BUSINESS, INNOVATION AND SKILLS (2013); UK OFFICE OF FAIR TRADING, THE COMMERCIAL USE OF PUBLIC INFORMATION (2006). For examples of subnational level studies, *see* J PREISCHE, DIGITALES GOLD. NUTZEN UND WERTSCHÖPFUNG DURCH OPEN DATA FÜR BERLIN. TECHNOLOGIE STIFTUNG (2014); GREGOR EIBL & BRIGITTE LUTZ, MONEY FOR NOTHING – DATA FOR FREE: HARD FACTS ABOUT THE

Commerce looked at the size of private sector revenues from "government data-intensive business activities" for the U.S. and arrived at a crude estimate in the range of 24 to 221 billion USD per year.[41] And a 2000 study for the European Commission estimated that for the then 15 E.U. Member States the part of the combined national income attributable to industries and activities built on exploiting public sector information ranged between €28 billion and €134 billion. Some have judged these estimates as far too optimistic.[42] Generally, researchers recognize there is a lack of hard data on which to base estimates.[43] Nevertheless, policymakers see fostering innovation and economic growth as an important goal of open data.

2. Political Accountability and Democratic Participation

A second goal pursued through open data policy is fostering political accountability and democratic participation. Current proactive disclosure policies cover a broad range of information: from basic information about a public authority's responsibility, organization, and procedures, to granular data about public spending and subsidies awarded.[44]

In the open data context, statements about the perceived benefits of open data for democracy are frequent. The G8 Open Data Charter mentions good governance and anti-corruption,[45] and argues that more public data on the use of natural resources and distribution of revenues, on land management and on development spending would promote accountability and good governance.[46] The World Bank makes a similar case, arguing that open data "supports democratic societies" and "encourages greater citizen participation in government affairs".[47] The French government's open data policy is driven by the idea that "opening and sharing data is the way for modern government to organize itself so that it is accountable, opens dialogue and trusts the collective intelligence of its citizens."[48] The Obama administration posits that making information available proactively online in

---

ECONOMIC POWER OF OPEN GOVERNMENT DATA, in CEDEM13: CONFERENCE FOR E-DEMOCRACY AN OPEN GOVERNMENT 289-302. Donau-Universitat Krems ed., 2013.

41 U.S. DEPARTMENT OF COMMERCE, *supra* note 34.

42 Robbin te Velde, *Public Sector Information: Why Bother?*, *in* THE SOCIO-ECONOMIC EFFECTS OF PUBLIC SECTOR INFORMATION ON DIGITAL NETWORKS: TOWARD A BETTER UNDERSTANDING OF DIFFERENT ACCESS AND REUSE POLICIES: WORKSHOP SUMMARY 25-28 (P. Uhlir ed., 2009).

43 *See* Mireille van Eechoud, *Calculating and Monitoring the Benefits of Public Sector Information Re-use*, *in* ZUGANG UND VERWERTUNG ÖFFENTLICHER INFORMATIONEN (Thomas Dreier et al. eds., forthcoming 2015).

44 See, for example, UNITED KINGDOM CABINET OFFICE, *Cabinet Office Organogram*, available at: http://data.gov.uk/organogram/cabinet-office; UNITED KINGDOM CABINET OFFICE, *Senior Officials "High Earners" Salaries*, available at: http://data.gov.uk/dataset/uk-civil-service-high-earners; OPEN KNOWLEDGE, *Where Does Europe's Money Go? A Guide to EU Budget Data Sources*, available at: http://blog.okfn.org/2015/07/02/where-does-europes-money-go/ (last accessed on 18th July 2015).

45 G8 OPEN DATA CHARTER, , *supra* note 3, ¶ 4-5

46 *Id.*

47 See, for example, WORLD BANK, Open Data Toolkit, available at: http://opendatatoolkit.worldbank.org/en/starting.html.

48 Translated from: "L'ouverture et le partage des données, c'est la manière, pour un Etat moderne, de s'organiser afin de rendre des comptes, d'ouvrir le dialogue, et de faire confiance à l'intelligence collective des citoyens". *Vademecum sur l'ouverture et le partage des données publiques*, Sécretariat Général pour la Modernisation de la Fonction Publique, Paris 2013, p. 5.

open formats increases accountability and promotes informed participation by the public.[49] A basic consideration of policy for the management of U.S. Federal information is that public disclosure of government information is essential to the operation of a democracy.[50] Similarly, the E.U. Public Sector Information Directive says that publishing documents held by the public sector "is a fundamental instrument for extending the right to knowledge, which is a basic principle of democracy."[51]

The idea of open government is tied to the ideal of transparency of governments' decisions and activities. Transparency is widely regarded as a precondition for the effective exercise of political rights and freedoms, and for ensuring accountable public authorities.[52] Access to information is a key aspect of democratic institutions that are based on representation, delegation, and accountability. Assessing, debating, and sanctioning public sector behavior requires accurate information.[53] In sum, the proactive disclosure of government data to the public for the purposes of political transparency, accountability and participation is becoming a central tenet in democratic governance.

3. Public Sector Efficiency and Service Delivery

A third set of pro open data arguments focuses on efficiency: open data should help to save resources and improve public services. For instance, the European Commission says open data will improve health services and traffic management, and help tackle environmental challenges, for instance through monitoring energy consumption.[54]

At the national level, an increasingly popular strategy is to publish performance data of publicly funded organizations.[55] Disclosing inspection and other data is alleged to improve performance of recipients of tax monies, like schools (test scores) and hospitals (deaths, waiting times).[56] Citizens in their capacity as customers are presumed to make better-informed choices when provided with such performance data.[57] Other initiatives serve to improve compliance and to assist in better

---

49 OMB Memorandum M-10-06 (Open Government Directive).

50 *See* OMB Circular A-130, rev. Transmittal 2, Management of Federal Information Resources. July 1994 (94 FR 18007), as well as the current version (Transmittal Memorandum No. 4, November 28, 2000 (65 FR 77677). The Circular has a residual role: it does not affect disclosure duties or rights to information under FOIA.

51 Recital 16 of the PSI Directive (Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, OJ 2003, L 345/90, as revised by Directive 2013/37/EC, OJ 2013, L175/1) states: "Making public all generally available documents held by the public sector — concerning not only the political process but also the legal and administrative process — is a fundamental instrument for extending the right to knowledge, which is a basic principle of democracy. This objective is applicable to institutions at every level, be it local, national or international."

52 On transparency *see* CHRISTOPHER HOOD, & DAVID HEALD, TRANSPARENCY: THE KEY TO BETTER GOVERNANCE? (2006); MARK BOVENS ET AL., THE OXFORD HANDBOOK OF PUBLIC ACCOUNTABILITY (2014).

53 Like transparency, accountability is a multifaceted concept. For a discussion of dimensions in relation to democracy, *see* Gijs Jan Brandsma & Thomas Schillemans, *The Accountability Cube: Measuring Accountability*, J. PUB. ADMIN. RESEARCH THEORY, *available at* mus034. doi:10.1093/jopart/mus034 (last accessed Sep. 18, 2012).

54 E.C. Communication on Open Data 3 2011.

55 Mireille van Eechoud, DE LOKROEP VAN OPEN DATA 9 (inaugural lecture University of Amsterdam 2014)

56 *Id.* at 9.

57 MCKINSEY & CO., *supra* note 35, 83-85

policymaking or prioritizing enforcement, for instance in the area of food safety standards or building safety.[58] Some open government data initiatives propose a more active role for the public: as an army of armchair auditors who can help identify possible savings.[59]

Furthermore, open data are expected to help public sector bodies carry out their tasks. Many users of open data portals are from the public sector.[60] Efficiency gains made when more transparency about information resources leads to less duplication of information collection, and hence more shared use of resources, are said to improve public sector services.[61] Furthermore, public sector bodies are expected to improve their services when they have more information at their disposal.[62] Efficient use of information resources is not a new concern of governments. For several decades information management policies have been argued to increase government efficiency.[63]

In the empirical mapping study, we found that different arguments for open data obtain varying levels of attention amongst different actors in different forms of digital media.[64] For example, in English language mainstream media outlets arguments and examples about the economic growth and technological innovation potential of open data received more attention than those related to public participation or democratic accountability. On social media platforms such as Twitter, distinct groups of actors were interested in different sets of topics around open data such that, for example, some were interested in startups and smart cities, and others were interested in transparency and open government.[65]

In sum, open data policies serve diverse interests. For the purposes of this paper, these can be clustered into: (i) innovation and economic growth, (ii) political accountability and democratic participation, and (iii) public sector efficiency.

B. PRIVACY INTERESTS

At the global level, the right to privacy is protected under, for instance, the United Nations Declaration of Human Rights[66] and the International Covenant on Civil and Political Rights.[67] In the

---

58 *See, e.g.*, Michael Flowers, *Beyond Open Data: The Data-Driven City,* in BEYOND TRANSPARENCY 185 (Brett Goldstein & Lauren Dyson,eds., 2013),

59 *See* Ben Worthy, *David Cameron's Transparency Revolution? The Impact of Open Data in the UK.* 1–24 (Univ. London, Nov. 29, 2013), available at: http://doi.org/10.2139/ssrn.2361428.

60 *See* WORLD BANK, *supra* note 29.

61 McKinsey, *supra* note 35, 57-58 makes the case for the energy sector-

62 On the advantages of combining existing data to yield useful information for e.g. disaster relief efforts or environmental pollution: Alan Feuer, *Mayor Bloomberg's Geek Squad*, N.Y. TIMES, 23 March, 2013, http://www.nytimes.com/2013/03/24/nyregion/mayor-bloombergs-geek-squad.html.

63 U.S. Congress, Office of Technology Assessment. INFORMING THE NATION: FEDERAL INFORMATION DISSEMINATION IN AN ELECTRONIC AGE (1988).

64 JONATHAN GRAY ET AL., MAPPING THE POLITICS OF OPEN DATA ON DIGITAL MEDIA (forthcoming).

65 *Id.*

66 Universal Declaration of Human Rights art. 12, G.A. Res. 217A (III), U.N. Doc. A/810 at 71 (1948).

67 International Covenant on Civil and Political Rights art. 17, Dec. 16, 1966, S. Treaty Doc. No. 95-20, 6 I.L.M. 368 (1967), 999 U.N.T.S. 171.

U.S., the Fourth Amendment and other laws protect privacy.[68] In Europe, the European Convention on Human Rights,[69] the European Union Charter of Fundamental Rights,[70] national constitutions, and other laws protect privacy.[71]

Public sector bodies hold an enormous amount of personal information, and this amount will likely grow. For instance, so-called "smart cities" may provide the public sector information about people such as up-to-date location data of cars, and detailed electricity metering data.[72] And, as public sector bodies offer more services online, they will obtain even more information about people.[73] Sometimes citizens volunteer personal information, for example when they use public services. But public authorities can also collect information through third parties, like educational and health care institutions.[74] And authorities can compel citizens to provide personal information. This element of force heightens privacy concerns.

We distinguish three broad categories of privacy concerns regarding open data: (i) the chilling effects on people in their interaction with the public sector, (ii) a lack of individual control over personal information, and (iii) the use of open data as input for social sorting and discriminatory practices.[75]

1. Chilling Effects

First, a chilling effect can occur if people interacting with public bodies fear that their information will be stored, or will be made public.[76] For example, people might be less inclined to contact public sector agencies if they doubt that their personal data will remain confidential.[77]

---

68 *See* WILLIAM CUDDIHY, THE FOURTH AMENDMENT: ORIGINS AND ORIGINAL MEANING 602-1791 (2009); DANIEL SOLOVE & PAUL SCHWARTZ, INFORMATION PRIVACY LAW 260-335 (5th ed., 2014).

69 Convention for the Protection of Human Rights and Fundamental Freedoms, art. 8, Nov. 4, 1950, 213 U.N.T.S. 222.

70 Charter of Fundamental Rights of the European Union of the European Parliament, art. 7-8, 2010 O.J. C 83/02, at 1.

71 *See, e.g.,* Dutch constitution, art. 10. Furthermore, each E.U. Member State has a national Data Protection Act implementing the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995 O.J. (L 281).

72 A smart city has been defined "as one that has digital technology embedded across all city functions," SMART CITY COUNCIL, http://smartcitiescouncil.com/smart-cities-information-center/definitions-and-overviews. *See*, ROBERT G. HOLLANDS. "Will the Real Smart City Please Stand Up? Intelligent, Progressive or Entrepreneurial?" 12 City 303-20 (2008).

73 Teresa Scassa, *Privacy and Open Government*, 6 FUTURE INTERNET 397-98 (2014).

74 *See* Solove, *supra* note 9, at 1142-1150 for an overview of federal, state and local record collection in the United States.

75 The three categories are based on: FUTURE FREDERIK ZUIDERVEEN BORGESIUS, IMPROVING PRIVACY PROTECTION IN THE AREA OF BEHAVIOURAL TARGETING 53-63 (2015). That study does not concern open data. In this paper, we adapt the categories to the open data context.

76 *See* KIERON O'HARA, TRANSPARENT GOVERNMENT, NOT TRANSPARENT CITIZENS: A REPORT ON PRIVACY AND TRANSPARENCY FOR THE CABINET OFFICE 24 (2011) www.gov.uk/government/uploads/system/uploads/attachment_data/file/61279/transparency-and-privacy-review-annex-a.pdf.

77 Jeff Jonas & Jim Harper, *Open Government: The Privacy Imperative*, *in* PRIVACY-ENHANCING PRACTICES, 321-30, 323 (2010).

People often provide personal information when engaging with public sector bodies. Public sector bodies often require information, for example, when people apply for a planning permission or business license, attempt to comply with health and safety standards, or submit tax claims or grant applications. The collection, use and exchange of personal information are part of the normal fabric of public sector activity. Many public services cannot be delivered without these activities.

People might refrain from contacting the public sector if they fear their personal information will not be kept confidential. Especially people with questions about diseases, pregnancies, drugs, financial troubles, or suicidal thoughts might refrain from asking help. Jonas and Harper illustrate the importance of communicating with the public sector without disclosing too much personal information with an example regarding a migrant.[78] Say Alice is a migrant who thinks her residence permit contains errors. If she thinks that visiting the immigration website will bring her to the attention of immigration law enforcement, she might forego looking for information. "If she cannot communicate this information anonymously, she almost certainly will not ask questions or volunteer information, denying herself help she might deserve while denying policymakers relevant information."[79] If Alice thought her data would be disclosed to others in and outside government, such a chilling effect might be greater.

By itself the chilling effect already harms the individual who refrains from an activity she might otherwise engage in. But if somebody does not seek help because of a chilling effect, for instance if someone does not seek information regarding a disease, he or she may also experience more tangible harms. People forgoing treatment of infectious diseases could harm society as a whole.

Uncertainty about what happens with one's personal information can ultimately adversely impact the quality of public services. As Scassa notes, with open data "there is a risk not only to individual privacy, but also to the relationship of trust that is meant to exist between citizens and their government."[80] Government statistics offices have realized for a long time that confidentiality of census answers is important – otherwise people might not give honest answers anymore. Trust in public authorities could diminish if people do not believe that their personal data will remain confidential.[81] In sum, open data policy could lead to a chilling effect on people communicating with the public sector, which is a privacy problem.

2. Lack of Control over Personal Information

A second privacy concern is that people lack control over their personal information if that information is released as open data. Publicly releasing personal information as open data can be especially troublesome because open data policy in its most liberal form implies that unlimited numbers of re-users can use the data for any purpose.

---

78 *Id.*
79 *Id.* at 317.
80 Scassa, *supra* note 73, at 408.
81 *See e.g.* U.S. Government Accountability Office, GAO-01-126SP, Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information 18 (2001).

Many privacy definitions focus on individual control over personal information. For instance, Westin defined privacy in 1967 as "the claim of individuals, groups or institutions to determine when, how, and to what extent information about them is communicated to others."[82] Many scholars use similar privacy definitions.[83] The privacy as control perspective is apparent in legal practice. For instance, the U.S. Supreme Court has described privacy as "the individual's control of information concerning his or her person."[84] The German Supreme Court says a person has, in principle, the right "to determine for himself whether his personal data should be divulged or utilized."[85] Privacy as control has deeply influenced the Fair Information Principles *(See infra* section IV).[86] The privacy as control perspective does not capture all the subtleties of privacy. Nevertheless, a loss of individual control over personal information is widely seen as a privacy problem.[87]

A lack of individual control over personal information can lead to subjective and objective privacy harm. Objective harm is, in Calo's words, "the unanticipated or coerced use of information concerning a person against that person."[88] The Eightmaps website provides an example of objective harm resulting from data released by the public sector.[89] Proposition 8 was a 2008 proposal to amend the California constitution with a referendum to ban gay marriage.[90] California law requires that campaign donations be published.[91] An anonymous website publisher took information regarding donors who supported Proposition 8, and overlaid that information on Google maps.[92] The map showed information such as the donor's name, approximate location, and the amount donated. Some of the donors received death threats, or were the victim of boycotts.[93] The

---

82 Alan F. Westin, *Privacy and Freedom* (1967), Westin 7 (1970) (reprint of 1967).

83 *See* Charles Fried, *Privacy*, 77 Yale L.J. 482 (1968) (discussing that 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."). *See also* AR Miller, *The Assault on Privacy: Computers, Data Banks, and Dossiers* 25 (1971) at 25 (describing privacy as "the ability to control the circulation of information relating to him").

84 U.S. DOJ v. Reporters Comm., 489 U.S. 749, 763 (1988).

85 Bundesverfassungsgericht 25 March 1982, BGBl.I 369 (1982), (Volks-, Berufs-, Wohnungs- und Arbeitsstättenzählung (Volkszählungsgesetz)), translation by Riedel, E.H., 5 HUM. RTS. L.J. 1984, at, 94, 101, ¶ II.

86 *See e.g.* Colin J. Bennett, *Regulating Privacy: Data Protection and Public Policy in Europe and the United States* 14 (1992).

87 *See e.g.* S Gürses, *Multilateral Privacy Requirements Analysis in Online Social Networks (PhD thesis University of Leuven)* (KU Leuven (academic version) 2010); Helen Nissenbaum, *Privacy In Context: Technology, Policy, and the Integrity of Social Life* (2010); Daniel J. Solove, *A Taxonomy of Privacy* 154 U. PA. L. REV. 477 (2006).

88 M. Ryan Calo, *The Boundaries of Privacy Harm* 86 Ind. L.J.1131,1133 (2011).

89 *Eightmaps.com and Too Much Information*, DALLAS MORNING NEWS (2009), http://dallasmorningviewsblog.dallasnews.com/2009/01/eightmapscom-an.html/. See also *See also* Michael Shin, *Show Me the Money! The Geography of Contributions to California's Proposition 8* 1 CALIF. J. POL. POL'Y 10 (2009). See generally on privacy-invasive online map services: Mark Burdon, *Privacy Invasive Geo-Mashups: Privacy 2.0 and the Limits of First Generation Information Privacy Laws* U.ILL.JL TECH.& POL'Y 1 (2010).

90 CALIFORNIA, *Prop 8: Eliminates Right of Same-sex Couples to Marry* (2008), available at: http://vigarchive.sos.ca.gov/2008/general/title-sum/prop8-title-sum.htm

91 Deborah G. Johnson, Priscilla M. Regan, Kent Wayland, *Campaign Disclosure, Privacy and Transparency*, 19 Wm. & Mary Bill Rts. J. 959, 972 (2011).

92 *Id.*

93 *Id. See also Prop 8 Donor Web Site Shows Disclosure Law Is 2-Edged Sword* N.Y. TIMES, (2009), http://www.nytimes.com/2009/02/08/business/08stream.html

dissemination of correct information can already produce objective harms, but the potential of harm arising from the public release of inaccurate or false data is at least as big.

The feeling of having no control over one's personal information is a "subjective harm," described by Calo as "the perception of loss of control that results in fear or discomfort."[94] Many people are uncomfortable with organizations processing large amounts of information about them. Furthermore, there is often information asymmetry between the individual and the organization that uses personal information. People may know that information about them is collected and stored, but may not know how this will be used. If people do not know who holds data about them, they cannot exercise control over those data.[95] Releasing data to an undetermined number of re-users aggravates the lack of control.

Furthermore, data privacy rules that apply to the public sector are often stricter than those that apply to the private sector.[96] However, if the public sector releases personal data as open data, that is, with no restrictions, the private sector can subsequently use those data, subject to more lenient (statutory) rules.[97] Hence, releasing personal data as open data reduces privacy protection, Furthermore, the more datasets governments disclose, the richer the possibilities for re-identification. In sum, releasing personal information as open data causes a lack of individual control over personal information.

3. Social Sorting and Discrimination

A third privacy-related concern is that open data could be used as input for social sorting and discriminatory practices.[98] For instance, if the public sector released personal data, data brokers would likely be among the main re-users.[99] Data brokers are "companies that collect consumers'

94 Calo, *supra* note 88 at 1143.

95 *See generally* on information asymmetry in the privacy area: Alessandro Acquisti & Jens Grossklags,*What Can Behavioral Economics Teach Us About Privacy? in* DIGITAL PRIVACY: THEORY, TECHNOLOGIES AND PRACTICES (Allesandro Acquisti et al. eds., 2007); Zuiderveen Borgesius, *supra* note 75, at 201-2015. According to Solove, the feeling of lost control resembles Kafka's THE TRIAL (DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE,38 (2004)). He suggests the main problem is "not knowing what is happening, having no say or ability to exercise meaningful control over the process" (*id.*, at 38).

96 For instance, in the U.S. the 1974 Privacy Act does not apply to the private sector. In the E.U., firms more easily meet the required legal basis test for personal data processing than public sector bodies do (*see* Directive 95/46/EC, *supra* note 71. Article 7(f)) applies to firms; article 7(e) applies to the public sector).

97 Scassa, *supra* note 73, at 405:, *id.* at 402.

98 *See* Solon Barocas and Andrew Selbst "Big Data's Disparate Impact" (2014) available at http://ssrn.com/abstract=2477899; WHITE HOUSE (John Podesta et al.). BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES (May 2014) *available at* www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf; Scassa, *supra* note 73 at 407.

99 Thomas P. Keenan, *Are They Making Our Privates Public? Emerging Risks of Governmental Open Data Initiatives,* Privacy And Identity Management For Life, IFIP AICT 375, 1 (2012), p. 11. See also Solove, *supra* note 9, at 1148-50.

personal information and resell or share that information with others."[100] The information can be used, for instance, for direct marketing, credit scoring, or screening job applicants.[101]

Many find data brokers' activities unfair and privacy-invasive.[102] As the Federal Trade Commission notes, personal information could be used for unfair discrimination. For instance, a company might use the information that there is a "Smoker in Household" to conclude that people in that household should not be offered insurance.[103] In surveillance studies, such practices are called "social sorting." As Lyon explains, social sorting involves "obtain[ing] personal and group data in order to classify people and populations according to varying criteria, to determine who should be targeted for special treatment, suspicion, eligibility, inclusion, access, and so on."[104] Social sorting is not inherently bad or good.[105]

For social sorting, data brokers can also use open data that do not include personal information. For instance, the average housing price in a certain zip code is not personal information. But that average price could be matched with somebody's address to estimate the value of his or her house. Hence, non-personal information can be used to enrich digital dossiers about people.

The following is another example of a social sorting effect resulting from open data. Suppose a city council releases crime statistics. A vendor of GPS car systems can overlay its own maps with the crime data in order to designate high-crime areas. The car GPS system can then route the driver around those areas. The practice could be seen as unfair for the people and businesses in that newly invented no-go area. In the no-go areas, insurance premiums might rise, and real estate prices and shop profits might drop.

In sum, potential privacy problems regarding open data include chilling effects on people communicating with the public sector, a lack of individual control over personal information, and discriminatory practices enabled by the released data. Hence, especially when datasets contain personal data, public sector bodies should give due consideration to the risks of disclosing data. We discuss below how to strike a balance between open data policy and privacy. But first we turn to the rules and guidelines that govern the disclosure of public sector information.

---

100 *See also* FEDERAL TRADE COMMISSION, DATA BROKERS. A CALL FOR TRANSPARENCY AND ACCOUNTABILITY 1 (May 2014) <www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf> (last accessed 26 May 2015).

101 *See* Scassa, *supra* note 73 at 407.

102 *See generally* Chris Jay Hoofnagle, *Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement* 29 N.C. J. INT'L L. & COM. REG. 595 (2003); JOSEPH TUROW, NICHE ENVY: MARKETING DISCRIMINATION IN THE DIGITAL AGE (2006); JOSEPH TUROW, THE DAILY YOU: HOW THE NEW ADVERTISING INDUSTRY IS DEFINING YOUR IDENTITY AND YOUR WORTH (2011).

103 FEDERAL TRADE COMMISSION, *supra* note 100, at 55-56.

104 David Lyon, *Surveillance As Social Sorting: Computer Codes and Mobile Bodies in* SURVEILLANCE AS SOCIAL SORTING: PRIVACY, RISK AND AUTOMATED DISCRIMINATION 20 (David Lyon ed., 2002).

105 David Lyon, Kevin Haggerty & Kirstie Ball, 'Introducing Surveillance Studies' in, ROUTLEDGE HANDBOOK OF SURVEILLANCE STUDIES 3 (David Lyon. Kevin Haggerty & Kirstie Ball eds., 2012).

III. GOVERNANCE OF PUBLIC SECTOR INFORMATION

In this section we discusses governance frameworks regarding access to public sector information. We discuss norms that govern open data, and norms that govern access to public sector information more generally. Freedom of information laws provide inspiration on how to strike a balance between privacy and transparency in the open data context.

A. OPEN DATA NORMS

Obligations for public authorities to release information as open data tend not to be encoded in hard law. Rather, open data policy is often promoted through administrative hierarchies, whereby the policy objectives, targets, and instructions range from superficial and permissive to detailed and strict.[106] Open data policymaking is partly shaped through political commitment in international forums such as the G8 and the Open Government Partnership.[107]

Open data initiatives rely on norms that regulate access to information. After all, open government data are, by definition, publicly available data. A myriad of such norms exists at the national level. The most generic disclosure duties arise under freedom of information acts, which typically cover the executive branch. Constitutional and administrative norms that help cement basic checks and balances also have implications for access to information, mandating for example that legislative texts are published,[108] and that the public has access to court decisions.[109]

Additionally, many countries have dedicated laws that govern information production for specific purposes, such as (national) statistics to aid policy development and monitoring,[110] land registries to facilitate secure property transactions, business registers,[111] or earth observation data

106 For example, the 2013 Obama order breathes ambition and decisiveness, and the elaboration by the Office for Management and Budget in a Open Data Policy Memorandum (Memorandum M-13-13*, supra* note 27) contains specific duties for departments to i.a. create lists of available data sets ('Public data listing'), engage with usergroups to prioritize release, see https://project-open-data.cio.gov/implementation-guide/. The E.U.'s Public Sector Information Directive shows a preference for the release of data in open formats, and also demands that Member States make practical arrangements "that help re-users in their search for documents available for re-use", e.g. in the form of asset registers. The E.C. Guidelines clearly favour pro-active release of data as open.

107 The members of G8 have through the 2013 Open Data Charter (*see supra* note 3) committed to drafting national open data action plans. The same mechanism is used by the Open Government Partnership. Note 8 supra.

108 *E.g.* article 10-11 Constitution du 4 octobre1958 (French constitution), JORF (Official Journal of the French Republic) No 0238 of 5 Oct. 1958; article 82 Bundesgesetz (German constitution), Bundesgesetzblatt III- 100-1.

109 *E.g.* Art. 6 European Convention on Human Rights (on the right to a fair trial) prescribes that court decisions are to be made public. This will usually be through delivery in court but may be achieved by other means as well, see Council of Europe, Guide to Article 6 (2013), 49-50. That a right to information is no guarantee for easy and affordable access is witnessed by the electronic access system for federal courts; *see* Vera Eidelman and Amul Kalia, *Right to Know: The PACER Mess And How to Clean It,* ELECTRONIC FRONTIER FOUNDATION (September 2, 2014). https://www.eff.org/deeplinks/2014/09/right-know-pacer-mess-and-how-clean-it.

110 *E.g.* the Wet op het Centraal Bureau voor de Statistiek, Staatsblad (Official Journal) 2003, 551 (Act on the Central Bureau of Statistics, the Netherlands); Statistics Act, R.S.C., 1985, c. S-19 (Canada); Statistics and Registration Service Act 2007, 2007 c. 18 (United Kingdom).

111 E.g. Handelsregisterwet, Staatsblad (State Journal) 2007, 153 (Act on Trade Register, The Netherlands); Companies Act 2006, 2006 c. 46 (United Kingdom); Handelsgesetzbuch § 8, Bundesgesetzblatt (Federal Official Journal) 4100-1 (Act on Trade Register, Germany).

produced for environmental and agricultural management.[112] Such specific laws will often lay down modalities for access. For example, confidentiality of identifiable information is of fundamental interest for the production of reliable and useful statistics. Hence, a basic principle in instruments that govern the production and dissemination of statistics is that personal information supplied for statistical purposes will not be disclosed or used for other (administrative) purposes.[113] In the interest of research, some statistics offices organize secure environments, where researchers can access micro-data under strict conditions. While no international legal right to (re)use public sector information exists, access to government information is increasingly recognized as a human right.[114]

B. ACCESS TO INFORMATION NORMS

Several international courts see access rights as part of, or closely connected to, the right to freedom of expression.[115] However, access rights are also recognized in case law of the European Court of Human Rights in the context of the right to private life.[116] By contrast, access rights may be conceived of as stand-alone constitutional rights.[117]

---

112 Mireille van Eechoud, *Commercialization of public sector information. Delineating the issues*, in THE FUTURE OF THE PUBLIC DOMAIN - IDENTIFYING THE COMMONS IN INFORMATION LAW 281-83 (Lucie Guibault & Bernt Hugenholtz (eds.), (2006),

113 See the Fundamental Principles of Official Statistics (Conference of European Statisticians, 1991), since updated and endorsed by the UN General Assembly (resolution 68/261 of 29 January 2014). Principle 6 reads: "Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes." Examples at national level: U.S. Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) (Pub. L. 107-347), title V; 116 Stat. 2962, Dec. 17, 2002); Dutch Statistics Act (Wet op het Centraal bureau voor de statistiek), Nov. 20, 2003).

114 For an extensive analysis of different human rights based conceptualizations of access to government information, *see* CHERYL A. BISHOP, ACCESS TO INFORMATION AS A HUMAN RIGHT (2011).

115 *See e.g.* Claude-Reyes et al. v. Chile, Inter-Am. Ct. H.R. Judgment of September 9, (2006) (¶ 77): a right to access of government information is guaranteed under Article 13 (Freedom of Thought and Expression) of the American Convention on Human Rights; States have a positive obligation to provide access, subject only to access restrictions are proportionate and for reasons permitted by the Convention. Youth Initiative for Human Rights v. Serbia, no. 48135/06 ECHR 2013: refusal to grant access to government information to a public watchdog violates the right to freedom of expression (Art. 10 ECHR). In TASZ v. Hungary, no. 37374/05, ¶ **35,** ECHR 2009,) the Court conceded it "has recently advanced towards a broader interpretation of the notion of "freedom to receive information" (*see* Sdruženi Jihočeské Matky v. Czech Republic (dec.), no. 19101/03, 10 July 2006) and thereby towards the recognition of a right of access to information." Previously it had rejected the claim that Article 10 ECHR includes a right to access government information, or a positive obligation for states to collect and disseminate information, *see, e.g.,* Guerra v. Italy, no. 116/1996/735/932, ECHR1998-I.

116 The ECtHR recognized a duty to impart information for the government as part of the right to respect for private life (Art. 8 ECHR) on various occasions: where it concerned access to fostercare records (Gaskin v. UK, no. 10454/83, ECHR 1989) and with respect to information about environmental pollution (threathening citizens' health; Guerra v. Italy, no. 14967/89, EcHR 1998-I; Onderyildiz v. Turkey, ECHR,18 Jun. 2002). In these cases applicants had a special interest.

117 *E.g.,* Art. 42 of the Charter of Fundamental Rights of the E.U. provides that any citizen of the Union has a right of access to documents held by E.U. institutions. Charter of Fundamental Rights of the European Union of the European Parliament, *supra* note 70. For an in depth analysis of access rights of a wider openess agenda, see Alberto Alemanno, *Unpacking the Principle of Openness in EU Law: Transparency, Participation and Democracy* (July 30, 2013). HEC Paris Research Paper No. LAW-2013-1003. Available at http://ssrn.com/abstract=2303644.

The Tromsø Convention of the Council of Europe concerns access to government information,[118] but it is unlikely that enough member states will ratify this convention for it to enter into force any time soon.[119] Much more successful is the U.N. Aarhus Convention of 1998, with nearly fifty contracting states.[120] The Aarhus Convention provides for a right of access to environmental information as part of every citizen's right to an adequate environment and duty to safeguard the environment for future generations.[121]

A fundamental right of access does not necessarily imply that authorities must actively disclose information to the general public in electronic form without use-restrictions. But open government agendas do steer policy in that direction. At the global level, the Open Government Partnership promotes proactive disclosure in re-usable formats.[122]

In various human rights domains, proactive disclosure is also advocated. The U.N. rapporteur on Human Rights typifies the right to access government information as "one of the central components of the right to freedom of opinion and expression."[123] To give effect to the right of access to information under article nineteen of the United Nations International Covenant on Civil and Political Rights and the Universal Declaration of Human Rights, "parties should proactively put in the public domain Government information of public interest" and "make every effort to ensure easy, prompt, effective and practical access to such information."[124] In 2006 the Inter-American Court of Human Rights held that States have a positive obligation to legislate freedom of information laws or take other measures that ensure access to government information.[125]

The adoption rate of freedom of information laws has accelerated on all continents over the past decade. Today nearly a hundred countries have enacted freedom of information laws.[126] Some

---

118 Council of Europe Convention on Access to Official Documents (Tromsø Convention, 18 Jun. 2009), available at http://conventions.coe.int/Treaty/EN/Treaties/Html/205.htm.

119 Mireille van Eechoud & Katleen Janssen, Rights of Access to Public Sector Information, 6 (3) MASARYK UNIV. J.L. & TECH. (2012), 471 at 486, available at https://journals.muni.cz/mujlt/article/view/2621/2185

120 Aarhus Convention of 1998 - Status January 2015.

121 UNECE (United Nations Economic Commission for Europe) Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters 1998 (entry into force 30 October 2001), https://treaties.un.org/doc/Publication/UNTS/Volume%202161/v2161.pdf.

122 *See, e.g.*, the Open Government Partnership declaration, available at http://www.opengovpartnership.org/about/open-government-declaration (last visited 1 May 2015)

123 U.N. General Assembly, 68th Session, 4 Sep. 2013, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, doc. A/68/362, 3. *See also* Resolution 12/12 adopted by the UN Human Rights Council, Right to the truth, 12 Oct. 2009, "Emphasizing that the public and individuals are entitled to have access, to the fullest extent practicable, to information regarding the actions and decision-making processes of their Government, within the framework of each State's domestic legal system"; Inter American Commission on Human Rights, *The Right to Truth in the Americas* (2014).

124 Human Rights Committee, 102nd session, Geneva, 11-29 July 2011. *General comment No. 34 on Article 19: Freedoms of opinion and expression (CCPR/C/GC/34)*, at ¶ 19.

125 Claude-Reyes, Inter-Am. Ct. H.R. Judgment of September 9, (2006) (¶ 77, 102).

126 *See* Global Right to Information Rating *available at* http://www.rti-rating.org/.

freedom of information laws contain provisions on proactive disclosure of information.[127] These tend to be vague and rather limited in scope. Traditionally access laws focus on disclosure of information on request by a member of the public. Access laws detail how requests can be made and how decisions must be reached.[128] A basic principle in freedom of information acts is that citizens do not have to motivate why they want access; the public interest in disclosure is considered a given.[129] A right to access information does not necessarily imply that the information can subsequently be used freely.[130]

Generally, freedom of information laws do not prescribe how data must be made available (for example in an open format, machine readable, with a certain frequency).[131] Usually, information disclosed under freedom of information laws is not required to be legally or technically open.[132] It is, however, a common feature that public bodies must, wherever possible, respect the mode of supply preferred by the requesting party, if the documents are available in such form or easily so produced.[133] Freedom of information laws usually contain privacy provisions, as discussed next.

C. ACCESS TO INFORMATION NORMS AND PRIVACY

Machine readable, bulk-downloadable open data complicate a problem that was already a difficult one in the pre-digital era. Since at least the 1970s, countries have grappled with the problem of balancing privacy protection and public sector transparency.[134] Generic freedom of information laws typically aim to accommodate privacy interests, for example by reserving access to personal information to parties with particular interests, or by only making records available in secure reading rooms.

---

127 For an analysis of the drivers of pro-active disclosure of government information and its growing enactment in binding norms, *see* H. Darbishire, PROACTIVE TRANSPARENCY: THE FUTURE OF THE RIGHT TO INFORMATION? A REVIEW OF STANDARDS, CHALLENGES, AND OPPORTUNITIES. World Bank Institute, Washington D.C. (n.d.).

128 Jonathan Gray & Helen Darbishire, *Beyond Access: Open Government Data & the Right to (Re)Use Public Information* (January 7, 2011). ACCESS INFO EUROPE AND OPEN KNOWLEDGE, *available at* SSRN: http://ssrn.com/abstract=2586400; Mireille van Eechoud et al (2014), *LAPSI Good practices collection on access to data*, LAPSI PROJECT, http://www.lapsi-project.eu/sites/lapsi-project.eu/files/LAPSI_D2.1_GoodPracticesAccess%28final%29.pdf.

129 Gray & Darbishire, *supra* note [132]; van Eechoud et al., *supra* note [132].

130 For instance, before implementation of the EU Public Sector Information Directive, the Belgian federal freedom of information act stipulated that no commercial use was allowed of information obtained under the act. See Art. 10 Wet van 11 april 1994 betreffende de openbaarheid van bestuur, Belgisch Staatsblad (Belgian State Journal) 30 Jun. 1994 (deleted by Act N. 2007-1600 of 7 March 2007.)

131 See the analysis of over forty freedom of information acts by Gray & Darbishire, *supra* note [132]; van Eechoud et al., *supra* note [132].

132 Jonathan Gray & Helen Darbishire, *Beyond Access: Open Government Data & the Right to (Re)Use Public Information* (January 7, 2011). ACCESS INFO EUROPE AND OPEN KNOWLEDGE, Available at SSRN: http://ssrn.com/abstract=2586400.

133 *See e.g.* Aarhus Convention of 1998, *supra* note 120, Art. 4.

134 For instance, in 1973 Sweden adopted its data privacy law partly to ensure that the generous Swedish regime for access to official documents, which dates back to 1776, would not unduly interfere with privacy. *See* GLORIA GONZÁLEZ FUSTER, THE EMERGENCE OF PERSONAL DATA PROTECTION AS A FUNDAMENTAL RIGHT OF THE EU 59 (2014).

Two balancing models regarding privacy and transparency can be distinguished in freedom of information laws. First, sometimes privacy is an absolute limitation to disclosure. That is, the legislator has done the balancing ex-ante. For example, the Dutch Freedom of Information Act provides that certain types of sensitive personal data (for example data concerning medical matters or religion) may never be disclosed.[135]

Second, sometimes freedom of information laws include a relative privacy exemption, to be weighed against the public interest in disclosure on a case-by-case basis.[136] U.S. freedom of information law exempts disclosure of personal, medical and similar files.[137] The test is whether disclosure "would constitute a clearly unwarranted invasion of personal privacy."[138] Personal information gathered as part of law enforcement is also exempt, if disclosure "could reasonably be expected to constitute an unwarranted invasion of personal privacy."[139] If privacy interests prevent disclosure, it is common for freedom of information laws to demand that exempted information is redacted so that the remainder can be released, even if cleaning documents is labor intensive.[140]

The Regulation that governs access to documents from EU institutions (Council of Ministers, Parliament and Commission) stipulates that access to a document shall be refused if "disclosure would undermine the protection of privacy… in particular in accordance with Community legislation regarding the protection of personal data".[141] The Obama Freedom of Information Memorandum states: "[i]n the face of doubt, openness prevails."[142]

At global human rights forums, the presumption is that the public interest in access to public sector information (as part of the freedom of expression) trumps privacy and other interests Human rights rapporteurs for the United Nations argue that access to information should be granted unless disclosure would cause serious harm to a protected interest such as privacy that outweighs the interest in disclosure.[143] The rapporteurs also stress the importance of proactive disclosure

---

135 Art. 10(1)d Wet openbaarheid van bestuur (Dutch Freedom of Information Act).

136 The Dutch Freedom of Information Act provides such a relative ground for non-disclosure, where the public's right to know does not outweigh a person's interest to have his or her private sphere protected. *Id.* Art. 10(2)e.

137 Exemption 6 of Electronic Freedom of Information Act of 1966, Pub. L. No. 104-231, 110 Stat. 3048 (codified at 5 U.S.C. § 552 (b)(6) (2012).

138 5 U.S.C. § 552 (b)(6).

139 5 U.S.C. § 552 (a)(2)(7).

140 E.g. Art. 4(4) Aarhus Convention exempts the release of personal data (if confidential under domestic law); Article 4(6) obliges states to redact the documents. Aarhus Convention of 1998, *supra* note 120.

141 Art. 4(1) Regulation 1049/2001. The way the institutions have interpreted this limitation is controversial; the European Ombudsman and the European Data Protection Supervisor signal overzealous interpretation of the rules on data protection as a threat to transparency. How the scales tip thus depends as much on the prevailing culture of transparency (or secrecy) as on the black letter. *See* H. Kranenborg, Toegang tot documenten en bescherming van persoonsgegevens in de Europese Unie, (Wolters Kluwer 2007), 188-194 [Access to documents and data protection in the European Union].

142 *See* Electronic Freedom of Information Act of 1966, *supra* note 137.

143 *See e.g.* the 19 Dec. 2006 Joint Declaration *International Mechanisms for Promoting Freedom of Expression* of UN Special Rapporteur on Freedom of Opinion and Expression, the OSCE Representative on Freedom of the Media, the OAS Special Rapporteur on Freedom of Expression and the ACHPR (African Commission on Human and Peoples' Rights) Special Rapporteur on Freedom of Expression. In similar vein: *see* the 2004 Joint Declaration *International Mechanisms for*

obligations, and posit that "access to information law should, to the extent of any inconsistency, prevail over other legislation."[144]

Particularly for the disclosure in the interest of political accountability and public debate, judgments in which the right to freedom of expression and the right to privacy are balanced can give guidance. The European Court of Human Rights recognizes the importance of proactive release of data on the internet as a means to ensure effective transparency and accountability. In the Wyptych case, the Court rejected the claim by an elected local councilor who argued that by requiring him to disclose information on his financial interests online, the Polish legislator infringed his right to privacy under article 8 of the European Convention on Human Rights. The Court said: "[t]he general public has a legitimate interest in ascertaining that local politics are transparent and Internet access to the declarations makes access to such information effective and easy. Without such access, the obligation would have no practical importance or genuine incidence on the degree to which the public is informed about the political process."[145]

Earlier, the European Court of Human Rights held that the privacy interests of politicians and higher public officials must yield to access rights.[146] The Court considered "that it would be fatal for freedom of expression in the sphere of politics if public figures could censor the press and public debate in the name of their personality rights, alleging that their opinions on public matters are related to their person and therefore constitute private data which cannot be disclosed without consent".[147]

In conclusion, regulation and case law regarding freedom of public sector information can provide inspiration on how to strike the balance between privacy and transparency in the open data context. Apart from that, there are more general principles to balance privacy-related interests and other interests. We turn to those Fair Information Principles now.

IV. GOVERNANCE OF PERSONAL INFORMATION

The Fair Information Principles (FIPs) provide a framework to balance privacy and other interests. Below we give an introduction to the FIPs, and to the OECD Privacy Guidelines, which include a version of the FIPs. We also discuss the main challenges when reconciling the FIPs and open data policy.

Promoting Freedom of Expression of UN Special Rapporteur on Freedom of Opinion and Expression, the OSCE Representative on Freedom of the Media, and the OAS Special Rapporteur on Freedom of Expression.

144 *See e.g.* the 2004 Joint Declaration, *supra* note 143.

145 Wypych v. Poland, no. 2428/05, ECHR 2005 (Admissability decision).

146 *TASZ v. Hungary* , no. 37374/05, _¶ **37**ECHR 2009.

147 *TASZ v. Hungary* , no. 37374/05, ¶ 37, ECHR2009.

A. FAIR INFORMATION PRINCIPLES (FIPS)

1. Background of the FIPs

The Fair Information Principles (FIPs),[148] or the Fair Information Practice Principles (FIPPs),[149] are ingrained in most data privacy laws and guidelines around the world. For example, the FIPs can be recognized in the 1973 report *Records, Computers, and the Rights of Citizens*, by the U.S. Department of Health, Education, and Welfare,[150] the Privacy Act,[151] and the Fair Credit Reporting Act.[152] The Federal Trade Commission and the White House have recently called for FIPs-based privacy regulation for the private sector.[153]

About a hundred countries in the world have a data privacy law including a version of the FIPs.[154] The FIPs can also be recognized in the United Nations Guidelines for the Regulation of Computerized Personal Data Files 1990,[155] and the APEC Privacy Framework of the Asia-Pacific Economic Cooperation (2005).[156] The E.U. Data Protection Directive (1995) contains one of the world's most stringent implementations of the FIPs.[157] European legal scholars tend to speak of data protection principles rather than of FIPs, but both sets of principles are similar.[158] Different countries, however, implement the FIPs differently. The FIPs give guidelines to balance privacy-related interests and other interests, such as those of business and the public sector.[159]

---

148 *See* NEIL RICHARDS, INTELLECTUAL PRIVACY: RETHINKING CIVIL LIBERTIES IN THE DIGITAL AGE 162 (2014) *See also* Gellmann, who speaks of "Fair Information Practices" (Gellman, Robert. "Fair Information Practices: A Basic History" (Version 2.02, November 11, 2013, continuously Updated). http://bobgellman.com/rg-docs/rg-FIPShistory.pdf).

149 *See* http://www.dhs.gov/xlibrary/assets/privacy/dhsprivacy_fippsfactsheet.pdf.

150 U.S. DEP'T OF HEALTH, EDUC. & WELFARE, RECORDS, COMPUTERS, AND THE RIGHTS OF CITIZENS i, xx (1973), www.justice.gov/opcl/docs/rec-com-rights.pdf.

151 Privacy Act of 1974, Pub. L. No. 93–579, 88 Stat. 1896 (codified at 5 U.S.C. § 552a (2012)).

152 Fair Credit Reporting Act of 1970, Pub. L. No. 91-508, 84 Stat. 1128 (codified as amended at 15 U.S.C. §§ 1681-1681x (2012)).

153 WHITE HOUSE, *Consumer Data Privacy in a Networked World* (Feb. 2012), www.whitehouse.gov/sites/default/files/privacy-final.pdf; FEDERAL TRADE COMMISSION, Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers' (March 2012) www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

154 Graham Greenleaf, *Sheherezade and the 101 Data Privacy Laws: Origins, Significance and Global Trajectories*, 23 J.L. Inf. & Sci. 4 (2014); Graham Greenleaf, *Global Tables of Data Privacy Laws and Bills*, SOCIAL SCIENCE RESEARCH NETWORK (3rd ed. June 1 2013), http://ssrn.com/abstract=2280875.

155 Guidelines for the Regulation of Computerized Personal Data Files, G.A. RES. 45/95, U.N. DOC. A/RES/45/95 (Dec. 14, 1990).

156 ASIA-PACIFIC ECONOMIC COOPERATION, *Privacy Framework*, http://www.apec.org/Groups/Committee-on-Trade-and-Investment/~/media/Files/Groups/ECSG/05_ecsg_privacyframewk.ashx (2005).

157 Directive 95/46/EC, *supra* note 71.

158 The core of E.U. data protection law can be found in article 6 of the Data Protection Directive (Directive 95/46/EC, *supra* note 71, Article 6).

159 Paul de Hert & Serge Gutwirth, *Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of Power*, *in* PRIVACY AND THE CRIMINAL LAW 91 (Erik Claes. Antony Duff & Serge Gutwirth eds., 2006); *see also* RICHARDS, *supra* note 148, at 162; *see also* Marc Rotenberg, Fair Information Practices and the Architecture of Privacy

2. OECD Guidelines

An influential version of the FIPs can be found in the Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, of the Organisation for Economic Co-operation and Development (OECD).[160] The OECD was established in 1960, by 18 European countries, the U.S. and Canada.[161] Now, the OECD has 34 member countries, including Mexico, Chile, Korea and Japan.[162] The OECD's self-stated mission is to "to promote policies that will improve the economic and social well-being of people around the world."[163]

One of the main reasons for the OECD to adopt the Guidelines was that several European data privacy laws from the 1970s restricted the export of personal data to countries that offered inadequate legal protection to personal data. Some, the United States in particular, worried that European countries would use data privacy law as a trade barrier.[164] The chairman of the expert group that wrote the 1980 OECD Guidelines summarized, "the OECD concern was that the response of European nations (and European regional institutions) to the challenges of TBDF [transborder data flows] for privacy might potentially erect legal and economic barriers against which it was essential to provide effective exceptions."[165] Therefore, OECD member states negotiated about more international cooperation, leading to the adoption of the Privacy Guidelines in 1980.[166]

The OECD Guidelines have a dual goal: they aim to protect privacy and individual liberties, and to foster the free flow of information between OECD member countries.[167] Many legal data privacy instruments have a similar dual goal.[168] In this paper we focus on protecting privacy and individual liberties, rather than on transborder data flows.[169]

(What Larry Doesn't Get). 2001 STAN. TECH. L. REV. 1, 1-4; Ann Cavoukian, *Evolving FIPPs:* Proactive Approaches to Privacy*, Not Privacy Paternalism*, *in* 20 REFORMING EUROPEAN DATA PROTECTION LAW 293 (Serge Gutwirth, Ronald Leenes & Paul de Hert eds., 2015).

160 OECD, *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm (last visited June 22, 2015). The OECD Privacy Guidelines call these principles: "Basic Principles of National Application."

161 Robert Wolfe, *From Reconstructing Europe to Constructing Globalization: The OECD in Historical Perspective, in* THE OECD AND TRANSNATIONAL GOVERNANCE 25, 25-26 (Rianne Mahon & Stephen McBride, eds. 2008).

162 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *Members and Partners*, http://www.oecd.org/about/membersandpartners/ (last visited June 22, 2015).

163 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *About the OCED*, http://www.oecd.org/about/ (last visited June 22, 2015).

164 Nicholas Platten, *Background to and History of the Directive*, *in* EC DATA PROTECTION DIRECTIVE 15 (David Bainbridge ed. 1996); GONZÁLEZ FUSTER, *supra* note 134, at 77.

165 Michael Kirby, *The History, Achievement and Future of the 1980 OECD Guidelines on Privacy*, 20 J.L. INF. & SCI. 1, 6 (2009-10).

166 *Id.* at 7-10.

167 Preamble of the OECD Privacy Guidelines (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *supra* note 160.)

168 *See* González, *supra* note 134, at 130. For instance, the E.U. Data Protection Directive has a similar dual goal (*see* Directive 95/46/EC, *supra* note 71, Art. 1).

169 On transborder data flows, *see* CHRISTOPHER KUNER, TRANSBORDER DATA FLOWS AND DATA PRIVACY LAW (2013).

The Guidelines are not legally binding, they merely "recommend" that OECD member countries implement the Guidelines.[170] The Guidelines stress that they provide "minimum standards"[171] and do not "preven[t] the application of different protective measures to different categories of personal data, depending upon their nature and the context in which they are collected, stored, processed or disseminated."[172] The OECD Guidelines use flexible terms so that all of the member countries can agree with them, even though the United States and European countries have different legal traditions, especially regarding privacy and personal data.[173]

When the OECD Guidelines were adopted in 1980, only about one third of the member states had adopted a data privacy law. Now, almost every OECD member state has a data privacy law with the FIPs at its core.[174] The OECD Guidelines were updated in 2013, but the essence of the principles was retained.[175] The 2013 OECD Privacy Guidelines are listed below. The principles partly overlap, and should be read together.

Collection Limitation Principle

> There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.[176]

Data Quality Principle

> Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.[177]

Purpose Specification Principle

> The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.[178]

---

170 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *Recommendation of the Council concerning Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data (2013)*, [C(80)58/FINAL, as amended on 11 July 2013 by C(2013)79], 11, 12, http://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf.

171 *Id.* at 14 (Art. 6).

172 *Id.* at 13 (Art. 3(a)).

173 Kirby, *supra* note 165, at 10.

174 David Wright, Paul de Hert & Serge Gutwirth, *Are the OECD Guidelines at 30 Showing Their Age?*, 54(2) COMMUNICATIONS OF THE ACM 119, 122 (2011).

175 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *The OECD Privacy Framework* 1, 4, www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf (last visited June 22, 2015) ("[T]his revision leaves intact the original 'Basic Principles' in Part Two of the Guidelines.").

176 *Id.* ¶ 7 (Paragraph 7 of the Guidelines governs the protection of privacy and transborder flows of personal data).

177 *Id.* ¶ 8.

178 *Id.* ¶ 9.

Use Limitation Principle

Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with [the Purpose Specification Principle] except:

a) with the consent of the data subject; or

b) by the authority of law.[179]

Security Safeguards Principle

Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data.[180]

Openness Principle

There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.[181]

Individual Participation Principle

Individuals should have the right:

a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to them;

b) to have communicated to them, data relating to them

(i) within a reasonable time;

(ii) at a charge, if any, that is not excessive;

(iii) in a reasonable manner; and

(iv) in a form that is readily intelligible to them;

c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and

---

179 *Id.* ¶ 10.
180 *Id.* ¶ 11.
181 *Id.* ¶ 12.

> d) to challenge data relating to them and, if the challenge is successful to have the data erased, rectified, completed or amended.[182]

Accountability Principle

> A data controller should be accountable for complying with measures which give effect to the principles stated above.[183]

3. Scope of the OECD Guidelines

The OECD Guidelines apply to "personal data," which the Guidelines define as "any information relating to an identified or identifiable individual (data subject)."[184] But the Guidelines limit the scope of application considerably; they "apply to personal data, whether in the public or private sectors, which, *because of the manner in which they are processed, or because of their nature or the context in which they are used, pose a risk to privacy and individual liberties*" (emphasis added).[185]

The Guidelines thus follow a risk-based approach: they only apply to personal data processing if it threatens privacy or individual liberties. By contrast, E.U. data protection law generally applies to personal data processing, and requires that personal data be processed fairly, including when the data do not pose a prima facie risk for individual liberties.[186]

In this paper, we assume that personal data should always be handled in line with the FIPs.[187] Hence, we do not follow the risk-based approach suggested by the OECD Guidelines. We do consider the risk of personal data processing and the sensitivity of personal data, but we do so *within* the FIPs framework (*see infra* section V-VII).

The OECD Guidelines have been criticized, for instance, for implementing the FIPs too weakly. Clarke says the OECD Guidelines aim "to facilitate international business, not to protect privacy" (emphasis in original).[188] The OECD Guidelines "were motivated by the facilitation of international

---

182 *Id.* ¶ 13.
183 *Id.* ¶ 14.
184 *Id.* ¶ 1 (The OECD personal data definition is similar to the definition in E.U. data protection law (Directive 95/46/EC, *supra* note 71, Art. 2(a))).
185 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *supra* note 175, ¶ 2.
186 *See* Charter of Fundamental Rights of the European Union, *supra* note 70, Art. 8: "1. Everyone has the right to the protection of personal data concerning him or her. 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law." *See also* Google Spain SL, Google Inc. v Agencia Española de Protección de Datos (AEPD), Mario Costeja González, C-1131/12 [CJEU] (May 13 2014) ¶ 69; Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources, C-293/12, Kärntner Landesregierung, C-594/12 (joined cases) [CJEU] (April 8 2014) ¶ 36.
187 The idea that personal data should always be processed in line with the FIPs could be seen as a European approach.
188 Roger Clarke, *Research Use of Personal Data: Comments to a Panel Session of the National Scholarly Communications Forum on "Privacy: Balancing the Needs of Researchers and the Individual's Right to Privacy under the New Privacy Laws"* (Aug. ,. 2002), www.rogerclarke.com/DV/NSCF02.html; *see also* William Bonner & Mike Chiasson, *If fair information principles are the*

business; they were constrained by the need to leave existing legislation unaffected; and their formulation reflected the need for cross-cultural comprehensibility."[189]

For better or for worse, the FIPs are widely accepted as a starting point for data privacy law. Although the application of the FIPs varies considerably, they express a nearly worldwide consensus on minimum standards for fair personal data use. The next section describes the main challenges that arise when trying to reconcile the FIPs and open data policy.

B. FIPS AND OPEN DATA: CHALLENGES

To date, policymakers and academics have given limited attention to the question of how privacy norms might be reconciled with policies aimed at making government data available for a wide range of uses. Policymakers and civil society actors recognize the privacy implications of open data.[190] But detailed analyses of the tension between open data and privacy, and especially of open data and the FIPs, is scarce. In the empirical mapping study, we found that, while there were some mentions of open data and privacy together on various forms of digital media, many of these were fleeting or incidental, and few of them contained substantive discussion about how to achieve a balance between the two.[191]

Several open data guidelines from civil society mention privacy – albeit cursorily. For example, the "8 Principles of Open Government Data" state that "[r]easonable privacy, security and privilege restrictions may be allowed."[192] The Sunlight Foundation says that for a dataset to be open, "[a]ll raw information from [the] dataset should be released to the public, except to the extent necessary to comply with federal law regarding the release of personally identifiable information."[193]

In the open data context, governmental and intergovernmental bodies also mention protecting privacy, albeit in a cursory fashion. For example, the G8 Open Data Charter recognizes that "there is national and international legislation, in particular pertaining to intellectual property, personally-identifiable and sensitive information, which must be observed."[194] The 2008 OECD

answer, what was the question? An actor-network theory investigation of the modern constitution of privacy, 15 INFO. & ORG. 267, 284 (2005).

189 Roger Clarke, *Beyond the OECD Guidelines: Privacy Protection for the 21st Century* (Jan. 4, 2000), www.rogerclarke.com/DV/PP21C.html.

190 For example, the United Kingdom's Open Rights Group expressed concern over the U.K. government's plans to release anonymized health and education data (*see* OPEN RIGHTS GROUP, *Open Data Privacy*, https://www.openrightsgroup.org/campaigns/opendata/open-data-privacy (last visited June 22, 2015)). The Open Knowledge and the Open Rights Group convened a working group on open data, personal data and privacy. *See* http://personal-data.okfn.org/.

191 Gray, Rogers & Bounegru, *supra* note 64.

192 *8 Principles of Open Government Data*, PUBLIC RESOURCE.ORG (Dec. 8, 2007), https://public.resource.org/8_principles.html.

193 SUNLIGHT FOUNDATION, *supra* note 14.

194 *G8 Open Data Charter and Technical Annex*, *supra* note 3.

recommendation on public sector information urges that member countries should clearly define "grounds of refusal or limitations," including "personal privacy."[195]

Compared to the OECD's recommendation, the implementation guidance material for Obama's 2013 Executive Order contains a more substantive discussion of privacy and the Fair Information Principles. The Executive Order includes the suggestion to "strengthen measures to ensure that privacy and confidentiality are fully protected and that data are properly secured," and to "incorporate privacy analyses into each stage of the information's life cycle." [196] As well as demanding compliance with relevant laws such as the U.S. Privacy Act of 1974 and the E-Government Act of 2002, the Executive Order suggests that "agencies should implement information policies based upon Fair Information Practice Principles and NIST guidance on Security and Privacy Controls for Federal Information Systems and Organizations."[197]

In the European Union, some work has been done on reconciling privacy and open data, in a thematic network funded by the European Commission to reflect on legal aspects of public sector information ("LAPSI"). The LAPSI Working Group on privacy warns that full application of European data privacy rules will seriously hamper the ability of public sector bodies to disclose information for re-use purposes.[198] In the following section, we discuss the main challenges that occur when trying to reconcile the FIPs and open data policy, starting with the purpose specification principle.

1. Purpose Specification Principle

The main problem that occurs when trying to reconcile the FIPs and open data policy is that open data policy fosters unanticipated re-use and innovation – "serendipitous reuse" as Shadbolt et al. put it.[199] But secondary use of personal data brings privacy risks. In FIPs parlance, using personal information for unforeseen purposes may breach the purpose specification principle.

The purpose specification principle is a cornerstone of many data privacy laws in the world. It follows from the purpose principle that personal data should only be collected for a purpose that is specified in advance, and that those data should not be used for incompatible purposes. [200] The 1973 "Records, Computers, and the Rights of Citizens" report from the U.S. Department of Health, Education, and Welfare already contained a similar principle: "[t]here must be a way for an

195 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*, [C(2008)36], 1, 5, http://www.oecd.org/sti/44384673.pdf (last visited June 22, 2015).

196 Memorandum M-13-13*, supra* note 27, at 9.

197 *Id.*

198 dos Santos et al., *supra* note 10 at 348-349; *see also* van Eechoud et al., *supra* note 10.

199 Nigel Shadbolt, Wendy Hall & Tim Berners-Lee, *The Semantic Web Revisited*, 21(3) IEEE INTELLIGENT SYSTEMS 96 (2006); *See also* Wendy Hall et al., *Open data and charities*, NOMINETTRUST.ORG 1, 16 (July 2012), http://www.nominettrust.org.uk/sites/default/files/Open%20Data%20and%20Charities.pdf ("Open data, taking inspiration from other ideologies of openness such as open source and open access publishing, articulates the idea that data should be usable by anyone, not just the data owner (or 'data controller' in the language of the Data Protection Act).")

200 *See* the Purpose Specification Principle from the OECD Guidelines (*supra* section IV.A.2).

individual to prevent information about him that was obtained for one purpose from being used or made available for other purposes without his consent."[201] In the Charter of Fundamental Rights of the European Union, the purpose specification is included in the right to protection of personal data.[202]

The requirement that personal data may only be used for purposes that are "not incompatible" is somewhat vague. The Article 29 Working Party, an advisory body in which European national Data Protection Authorities cooperate,[203] has discussed the purpose specification in depth. To assess whether a new purpose is compatible with the collection purpose, says the Working Party, all circumstances must be considered. Relevant circumstances include the relation between the original and the new purpose, the collection context; the reasonable expectations of the data subject,[204] the personal data's sensitivity, the risks resulting from the new purpose, and the measures the controller has in place to mitigate risks.[205]

According to the Working Party, an example of a new purpose that is incompatible with the original processing purpose is contained in the following example. A public sector body publishes public servants' contact details on its website, to enable the public to contact them. [206] A re-user wants to merge the public servants' home addresses and phone numbers with the published contact details, to build an interactive map.[207] The re-use is not within the reasonable expectations of the civil servants, making. the purpose incompatible and thus not allowed.[208]

2. Security and Accountability Principles

The security principle requires appropriate security for personal data. Data controllers must protect data against unauthorized disclosure, access, or other use. When thoughtlessly releasing personal data, a public sector body breaches the security principle. After all, the public sector body would have no control over how the data are used – and neither would the data subjects. The mere fact that data subjects have no control over the use of their data is a subjective privacy harm. Moreover, anybody could access the data, including data brokers and identity thieves.

---

201 U.S. DEP'T OF HEALTH, EDUC. & WELFARE, RECORDS, *supra* note 150, at xx.

202 Charter of Fundamental Rights of the European Union, *supra* note 70, Art. 8(2).

203 *See generally* on the Working Party: Yves Poullet & Serge Gutwirth, *The contribution of the Article 29 Working Party to the construction of a harmonised European data protection system: an illustration of "reflexive governance"?*, *in* DÉFIS DU DROIT À LA PROTECTION DE LA VIE PRIVÉE [CHALLENGES OF PRIVACY AND DATA PROTECTION LAW] (María Verónica Perez Asinari & Pablo Palazzi eds., 2008). The Working Party's opinions are not legally binding, but they are influential in Europe. Judges and national Data Protection Authorities often follow the Working Party's interpretation.

204 In the United States and the European Union, the "reasonable expectation of privacy" is interpreted differently. The European Court of Human Rights says: "A person's reasonable expectations as to privacy is a significant though not necessarily conclusive factor" (ECtHR, Perry v. United Kingdom, No. 63737/00, (July 17, 2003) ¶ 37). *See* on the United States: SOLOVE & SCHWARTZ, *supra* note 68, at 288-335.

205 Article 29 Data Protection Working Party, Opinion 06/2013 on open data and public sector information ('PSI') reuse, 1021/00/EN WP 207, at 20 (June 5, 2013); *see also* Article 29 Data Protection Working Party, Opinion 03/2013 on purpose limitation, 00569/13/EN WP 203 (Apr. 2, 2013).

206 Article 29 Data Protection Working Party, Opinion 06/2013, *supra* note 205, at 20.

207 *Id.*

208 *Id.*

The accountability principle makes the data controller responsible for complying with the FIPs. The OECD Guidelines define the data controller as the party that "is competent to decide about the contents and use of personal data regardless of whether or not such data are collected, stored, processed or disseminated by that party or by an agent on its behalf."[209] A public sector body holding the personal data is usually the data controller. If a re-user obtains personal data from the public sector body, the re-user typically becomes a data controller as well.

3. Data Quality Principle

The data quality principle requires appropriate accuracy, completeness, and relevancy of personal data. One of the aims of the principle is to reduce the risk that organizations base decisions about people on incorrect data. Decisions based on incorrect data can have disastrous effects for a data subject.[210] The data quality principle is relevant to open data. Releasing incorrect personal data could have detrimental effect.[211] For example, imagine that a website about political campaign financing erroneously includes your name as a donor to a fringe extremist party.

4. Collection Limitation and Transparency Principle

The transparency principle, or openness principle, requires transparency regarding data processing, especially towards the data subject.[212] The transparency principle aims to prevent data controllers from abusing information asymmetry.

The transparency principle is prominent in data privacy laws, and can be recognized, for instance, in the proposed U.S. Consumer Privacy Bill of Rights,[213] the E.U. Data Protection Directive,[214] and the proposed E.U. Data Protection Regulation.[215] Some authors suggest that the transparency principle is the most important principle of the FIPs.[216] The transparency principle has old roots. The first principle of the U.S. Department of Health, Education, and Welfare report of 1973 states: "[t]here must be no personal-data record-keeping systems whose very existence is secret."[217] The second principle adds that "[t]here must be a way for an individual to find out what information about him is in a record and how it is used."[218]

---

209 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *supra* note 160, Art. 1. The OECD data controller concept is different from the E.U. concept of "data controller." In brief, under E.U. data protection law, the data controller is the party that determines the goals and means for personal data processing. A party that processes personal data on behalf of the controller is the "data processor" (Directive 95/46/EC, *supra* note 71, Art. 2(d), 2(e)).

210 *See*, for instance, Romet v. Netherlands, No. 7094/06, ECHR (2012).

211 *See* Scassa, *supra* note 73; Rotenberg, *supra* note 159.

212 To avoid confusion with the open character of open data, we will speak of the "transparency principle" rather than of the "openness principle."

213 White House, *supra* note 153, at 47 (Consumer Privacy Bill of Rights and transparency principle).

214 Directive 95/46/EC, *supra* note 71, Art. 10, 11.

215 European Commission, *Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)* COM(2012) 11 final, 2012/0011 (COD), Jan. 25, 2012, Art. 5(a).

216 *See, e.g.*, de Hert & Gutwirth, *supra* note 159; Zuiderveen Borgesius, *supra* note 75, at 99, 106-11.

217 U.S. DEP'T OF HEALTH, EDUC. & WELFARE, *supra* note 150, at 41.

218 *Id.*

The collection limitation principle requires that personal data, where appropriate, be collected with the data subject's knowledge or consent. The Article 29 Working Party recommends that a public sector body inform data subjects in advance whether the personal data they provide might be disclosed, for example due to freedom of information laws.[219]

5. Use Limitation and Individual Participation Principle

The individual participation principle aims to give people some control over the processing of their personal data. For instance, data subjects have the right, under certain circumstances, to rectify their data. The principle illustrates that the privacy as control perspective has influenced the FIPs.[220]

As previously stated, unrestricted re-use of personal data would breach the purpose specification principle – but the use limitation principle seems to offer a way out. The use limitation principle says that personal data should only be used in accordance with the purpose specification principle, except: "a) with the consent of the data subject; or b) by the authority of law."[221] Hence, personal data can be used for a new (prima facie incompatible) purpose if the data subject consents to the new use. Indeed, some have suggested that public sector bodies should obtain consent of the relevant individuals before releasing data as open data.[222]

However, relying on data subject consent for disclosing personal data as open data has some drawbacks. First, people are often in a dependent position vis-à-vis the public sector, and that position may make consent involuntary. Somebody interacting with the public sector might not feel free to withhold consent. Say Alice goes to a city council office for unemployment benefits. Alice really needs money, as she has missed five rent payments, and risks being evicted with her young child. Because she wants to be cooperative, Alice is unlikely to withhold consent to any request by the city council office. Under E.U. data privacy law, consent given under too much pressure is invalid, because consent must be "freely given."[223] For instance, if an employer asks an employee for consent, the consent might not be freely given because of the power imbalance.[224] And according to the European Court of Justice, people applying for passports cannot be deemed to have freely consented to have their fingerprints taken, because people need a passport.[225]

---

219 Article 29 Data Protection Working Party, Opinion 06/2013, *supra* note 205, at 9. Narayanan et al. suggest that people should be informed regarding re-identification risks. (Arvind Narayanan, Joanna Huey & Edward W. Felten, *A Precautionary Approach to Big Data Privacy* 1, 16 (Mar. 19, 2015), http://randomwalker.info/publications/precautionary.pdf.)

220 *See* Kirby, *supra* note 165, at 8 (citing Alan Westin as an influence on the OECD Guidelines).

221 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *supra* note 175, ¶ 10 (Use Limitation Principle).

222 *See, e.g.*, Bart van der Sloot, *On the fabrication of sausages, or of Open Government and Private Data* 3 JeDEM 1-16, 14 (2011).

223 ELENI KOSTA, CONSENT IN EUROPEAN DATA PROTECTION LAW 256 (2013).

224 Article 29 Data Protection Working Party, Opinion 15/2011 on the definition of consent, 01197/11/EN WP187, at 13-14 (July 13, 2011).

225 Schwartz v. Stadt Bochum, C-291/12, [CJEU] (Oct. 17, 2013) ¶ 32.

A second problem with data subject consent as a justification for disclosing personal data is that a request for consent can only be meaningful if it specifies a processing purpose.[226] A third problem is that behavioral studies cast doubt on individual consent as a privacy protection measure. For example, on the internet, people tend to click "I agree" to requests that they see on their screens without knowing what they are agreeing to.[227] Furthermore, it may be impractical for the public sector body to obtain the consent of thousands of individuals. In sum, obtaining data subjects' consent to release personal data is not a general solution to reconcile FIPs and open data policy.

To conclude, from a FIPs perspective, the main problem with open data is that open data can be used by anyone, for any purpose, without re-use restrictions. A complete lack of re-use restrictions clashes with the purpose specification principle.

V. TYPES OF DATA

Compromises are possible to balance privacy and open data interests. This balancing act may play out differently for different types of data. To help balance the different interests, we distinguish between four data categories, with different levels of privacy risks: (i) raw personal data, (ii) pseudonymized data, (iii) anonymized data, and (iv) non-personal data. We borrow the "raw personal data" category from Davies, and borrow the other three categories from the Article 29 Working Party.[228] We distinguish the four categories to structure the discussion, but the boundaries between them are not clear-cut (*see infra* Section V.E).

A. RAW PERSONAL DATA

With raw personal data, no attempt has been made to mitigate re-identification risks. Examples of raw personal data include names, social security numbers, and personal email addresses. Some open data advocates suggest that open data should never include raw personal data.[229] Indeed, while open data policy is important, releasing raw personal data without any re-use restrictions is usually neither desirable nor legally feasible.

However, in some circumstances raw personal data should be disclosed, because the public interests in disclosure outweigh the privacy interests. For instance, say a public registry of judges

---

226 Article 29 Data Protection Working Party, Opinion 15/2011, *supra* note 224, at 9.

227 *See, e.g.*, Acquisti & Grossklags, *supra* note 95; Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880 (2013); Solon Barocas & Helen Nissenbaum, *Big Data's End Run around Anonymity and Consent*, *in* PRIVACY, BIG DATA, AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT 44 (Julia Lane et al. eds., 2014); Zuiderveen Borgesius, *supra* note 75.

228 Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN WP 216, (Apr. 10, 2014). The "raw" category is inspired by work by Tim Davies, *Untangling the data debate: definitions and implications*, OPENDATAIMPACTS.NET (Mar. 23, 2012), http://www.opendataimpacts.net/2012/03/untangling-the-open-data-debate-definitions-and-implications/; *see also* Hall et al., *supra* note 199.

229 For instance, Hall et al. say that raw personal data "should never be directly published as openly licensed and accessible data without explicit consent of the individuals covered in the data." (Hall et al., *supra* note 199, at 14) Tim Berners-Lee and Nigel Shadbolt, two authors who promote open data, say, "[i]n the drive to free up data we have always argued that it is essential to respect individual privacy and national security." (Tim Berners-Lee & Nigel Shadbolt, *There's gold to be mined from all our data*, TIMES, LONDON, Dec. 31, 2011.)

reveals positions and jobs judges hold elsewhere, to uphold impartiality of the judiciary. If these data did not identify individual judges, disclosure would not offer sufficient transparency.[230] More generally, people must accept that their privacy diminishes if they take on certain functions in the public sector. For example, it is widely accepted that media can report on politicians, even when politicians might sometimes prefer that certain information remain confidential.[231]

But the fact that certain raw personal data should be disclosed does not imply that they should be disclosed as open data without re-use restrictions. Even if a law states that certain information must be made public, it does not necessarily follow that such information should be released fully openly. By 1972, some already argued that "the assumptions built into 19th century ideals of public records need revisiting in light of technology."[232] And as Scassa notes, "[i]n many cases, decisions around the public nature of the information were made in an era before the Internet."[233] Hence, personal data that are required to be public by law should not automatically be seen as data that can be released as fully open data.

To illustrate, in many countries court proceedings are mandated to be public. But if court proceedings can only be consulted by traveling to the courthouse and inspecting paper files, the personal information in those files is protected by "practical obscurity."[234] As the U.S. Supreme Court noted in 1989, "there is a vast difference between the public records that might be found after a diligent search of courthouse files, county archives, and local police stations throughout the country and a computerized summary located in a single clearinghouse of information."[235]

In sum, even if the law requires disclosing certain personal information as part of the public record, the public sector body should still assess whether this information should also be made available as open data on the web.[236] As we shall argue in Part VI, the question of whether or not data should be made available as open data is a further, additional question that follows the question of whether the data should be made publicly available at all.

B. PSEUDONYMIZED DATA

Pseudonymized data are personal data about an individual that are tied to a unique identifier other than a name. For instance, *William Carey Jones* could be referred to as person number *4.417.749*. Pseudonymization can be described as follows: "replacing one attribute (typically a unique

230 As Gary T. Marx puts it, sometimes "disclosure norms" trump "privacy norms." (Gary T. Marx, *Foreword: Privacy Is Not Quite Like the Weather*, *in* PRIVACY IMPACT ASSESSMENT v, viii (David Wright & Paul De Hert eds., 2012))
231 *See supra* Section III.C.
232 Chris Jay Hoofnagle, *The Origin of Fair Information Practices: Archive of the Meetings of the Secretary's Advisory Committee on Automated Personal Data Systems (SACAPDS)*, SOCIAL SCIENCE RESEARCH NETWORK 1, 12 (July 15, 2014), http://ssrn.com/abstract=2466418.
233 Scassa, *supra* note 73, at 403; *see also* Keenan, *supra* note 99, at 1.
234 Dep't of Justice v. Reporters Comm. for Freedom of the Press, 489 U.S. 749, 762 (1989).
235 *Id.* at 764.
236 *See generally* Amanda Conley, Anupam Datta, Helen Nissenbaum & Divya Sharma, *Sustaining Privacy and Open Justice in the Transition to Online Court Records: A Multidisciplinary Inquiry*, 71 MD. L. REV. 772 (2012).

attribute) in a record by another."[237] Merely substituting names with other unique identifiers is rarely enough to anonymize personal data, or to safeguard privacy.[238]

A well-known example of the limited effect of pseudonymization as an anonymization measure is the 2006 AOL data breach. AOL released pseudonymized data about users of its search engine by replacing the name of each searcher with a number.[239] However, journalists soon found out the real name of the person behind one of the pseudonymous search profiles and published an article entitled "A Face Is Exposed for AOL Searcher No. 4417749."[240] The journalists found the woman behind search profile 4417749 without using sophisticated re-identification techniques.[241] The search queries of user no. 4417749 suggested that the searcher was an elderly woman with a dog, from a specific town.[242] When the journalists visited her house, she confirmed that the searches were hers.[243]

The Article 29 Working Party suggests that pseudonymized data are a type of personal data, and are thus within the scope of European data protection law.[244] Some computer scientists have a similar view.[245] However, the Working Party's view has also been criticized for making the scope of personal data too broad.[246]

While pseudonymizing personal data rarely, if ever, makes people non-identifiable, pseudonymization can help to protect privacy interests, by making it a bit harder to recognize people by name.[247] For instance, Conley et al. suggest that pseudonymization can help to mitigate privacy concerns when court cases are published online.[248] If people's names are changed to "[party 1]" and "[party 2]" in judgments, it would be impossible to search within court records on the basis of a person's name. Pseudonymization also reduces the chance that somebody who looks at the data will recognize a person by name. However, it might still be possible to recognize people based on the facts of a case discussed in the judgment. Nevertheless, pseudonymization adds a thin layer of practical obscurity.[249]

---

237 Article 29 Data Protection Working Party, Opinion 05/2014, *supra* note 228, at 20.
238 Narayanan et al., *supra* note 219, at 2; PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE, at 38-39 (May 2014), http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.
239 Michael Barbaro & Tom Zeller Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES, Aug. 9, 2006.
240 *Id.*
241 *Id.*
242 *Id.*
243 *Id.*
244 Article 29 Data Protection Working Party, Opinion 05/2014, *supra* note 228, at 10.
245 Narayanan et al., *supra* note 219.
246 *See e.g.*, Khaled El Emam & Cecilia Álvarez, *A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques*, 5(1) INT'L DATA PRIVACY LAW 73 (2015).
247 Narayanan et al., *supra* note 219, at 2; PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 238, at 38-39.
248 Conley, Datta, Nissenbaum, & Sharma, *supra* note 236, at 842.
249 *See id.*

Different countries have different traditions. In the Netherlands, many court decisions are published online on a centralized website.[250] But if the litigating parties are individuals, their names are changed to neutral phrases such as "plaintiff" and "defendant."[251] In other countries, litigants' names are often included in court documents, even when published online.[252]

In sum, pseudonymization can help to reduce privacy risks – it is a useful but not sufficient security measure. Because pseudonymizing data is not enough to anonymize data, pseudonymous data must generally be treated as personal data.

C. ANONYMIZED DATA

Anonymized data are ex-personal data that are rendered anonymous in such a way that data subjects are no longer identifiable. Aggregated data are typically anonymous. For instance, the information that "*112,580 people live in Berkeley*," without additional information, does not identify an individual. Anonymization can be defined as "a technique applied to personal data in order to achieve irreversible de-identification."[253] Anonymized data are outside the scope of the FIPs, as the FIPs only apply to personal data.

The fact that personal data can be aggregated and thereby anonymized seems an appropriate way to strike a balance between privacy interests and open data interests.[254] For instance, statistics can often be disclosed as open data, as long as they are anonymized and aggregated.[255] To illustrate, a crime map could say that on a certain day, "between one and ten burglaries took place on or near Bancroft Way," rather than "one burglary took place at 11 Bancroft Way." Traffic data could say that "between one and ten cars drove on Bancroft Way between April 15 and April 20, 2015," rather than "one car drove on Bancroft Way on April 19, 2015 at 12:24 a.m."

---

250 *See generally* Laurens Mommers, *Access to Law in Europe*, *in* INNOVATING GOVERNMENT 383 (Simone van der Hof & Marga M. Groothuis eds., 2011); LEONIE VAN LENT, EXTERNE OPENBAARHEID IN HET STRAFPROCES (2008).

251 Our translations. Anonimiseringsrichtlijnen, RECHTSPRAAK.NL, http://www.rechtspraak.nl/Uitspraken/Anonimiseringsrichtlijnen/ (last visited July 2, 2015). Not all personal data are obfuscated; for example, the attorneys for a case are mentioned by name. The Spanish system is similar to the Dutch one (*see* James B. Jacobs & Elena Laurrauri, *Are criminal convictions a public matter? The USA and Spain*, 14(1) PUNISHMENT & SOC'Y 3 (2012), with further references to literature on other countries.)

252 *See*, for instance, about the United States: Nancy S. Marder, *From "Practical Obscurity" to Web Disclosure: A New Understanding of Public Information*, 59 SYRACUSE L. REV. 441, 444-47 (2009); Conley, Datta, Nissenbaum, & Sharma, *supra* note 236. For an overview of how to obtain criminal records in fifty-four countries, *see* KPMG, *Disclosure of Criminal Records in Overseas Jurisdictions, Summary of Findings* (March 2009), http://www.cpni.gov.uk/documents/publications/2009/2009-criminal_records_disclosure_intro_and_exe_summary_march09.pdf?epslanguage=en-gb.

253 Article 29 Data Protection Working Party, Opinion 05/2014, *supra* note 228, at 7.

254 *See* Article 29 Data Protection Working Parry, Opinion 06/2013, *supra* note 205, at 12; Hall et al., *supra* note 199, at 45; *see also* OFFICE OF THE AUSTL. INFO. COMM'R, *Information Policy agency resource 1: De-identification of data and information*, (Apr. 2014), http://www.oaic.gov.au/images/documents/information-policy/information-policy-resources/information-policy-agency-resources/information_policy_agency_resource_1.pdf.

255 *See* Francesco Molinari & Jesse Marsh, *Does Privacy have to do with Open Data? Some preliminary reflections – and answers*, CEDEM13 CONFERENCE FOR E-DEMOCRACY AND OPEN GOVERNMENT 303, 311 (Peter Parycek & Noella Edelmann eds., 2013).

But two caveats are in order. First, anonymizing data does not guarantee privacy and fairness.[256] For instance, the Dutch public reacted angrily when the police used aggregated information derived from data gathered by TomTom, a vendor of car navigation systems.[257] The police used the data to choose the best spots to install speeding cameras.[258] The Dutch Data Protection Authority examined whether TomTom's practices complied with E.U. data privacy law, and did not find major problems.[259] The data obtained by the police were properly anonymized through aggregation, and thus outside the scope of the FIPs.[260]

The TomTom example illustrates a broader problem; the FIPs apply to personal data – and only to personal data. But people can be treated unfairly, or feel like they are being treated unfairly, on the basis of information that is *based* on personal data concerning them, but that is not personal data anymore.[261] Moreover, as the aggregated information is outside the scope of the FIPs, the data subject rights that follow from the FIPs, such as access and correction rights, no longer apply. As Seda Gurses notes, anonymization can "disempower" the individual.[262] The FIPs and most data privacy laws around the world have this problem in common.[263] We will not attempt to solve the problem here. But we do note that sometimes a public sector body may want to decide not to release anonymized information, even if the information is outside the of FIPs' scope.

A second caveat is that anonymized data are often less interesting for re-users than raw personal data or pseudonymous data. As Zevenbergen et al. put it:

> The utility and privacy of data are generally directly and inversely related. For many datasets, it has proven difficult – if not impossible – to increase data subjects' privacy without concurrently decreasing the overall utility of the dataset. Small privacy gains are generally achieved by far-reaching decreases in data utility. A small increase in data utility often requires much more personal information to be revealed.[264]

---

256 Narayanan et al., *supra* note 219, at 3.

257 Charles Arthur, *TomTom satnav data used to set police speed traps*, GUARDIAN, Apr. 28, 2011.

258 *Id.*

259 Press Release, College Bescherming Persoonsgegevens, Following report by Dutch DPA, TomTom provides user with better information (Jan. 12, 2012), https://cbpweb.nl/en/news/following-report-dutch-dpa-tomtom-provides-user-better-information.

260 *See id.; See also* Harold Goddijn, *This is what we really do with your data*, TOMTOM.COM, http://www.tomtom.com/page/facts (last visited June 23, 2015).

261 *See generally* Lyon, *supra* note 104.

262 Seda Guerses, *The Spectre of Anonymity*, VOUS-ETES-ICI.NET 1, 5, http://vous-etes-ici.net/wp-content/uploads/2014/02/SedaAnonymityMute.pdf (last visited June 23, 2015).

263 *See generally* Mireille Hildebrandt & Serge Gutwirth eds., PROFILING THE EUROPEAN CITIZEN: CROSS-DISCIPLINARY PERSPECTIVES (2008); Barocas & Nissenbaum, *supra* note 227; Joris Van Hoboken & Frederik Zuiderveen Borgesius, *Scoping Electronic Communication Privacy Rules: Data, Services or Values*, (EuroCPR 2015 Working Paper).

264 Bendert Zevenbergen, Ian Brown, Joss Wright & David Erdos, *Ethical Privacy Guidelines for Mobile Connectivity Measurements*, OXFORD INTERNET INST. 1, 11 (Nov. 2013), http://www.oii.ox.ac.uk/research/Ethical_Privacy_Guidelines_for_Mobile_Connectivity_Measurements.pdf. *See* along similar lines, Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010). Slightly more optimistic is Felix T. Wu, *Defining Privacy and Utility in Data Sets*, 84 U. COLO. L. REV. 1117 (2013).

In sum, anonymized data can – in theory – safely be disclosed as open data, without re-use restrictions. However, in practice, irreversible anonymization is exceedingly difficult, and perhaps impossible.

D. NON-PERSONAL DATA

A fourth type of data is non-personal data. Many datasets do not contain, and have never contained, personal data. Examples include datasets regarding public transport times, weather conditions, sea tides, road maps, public sector budgets, and environmental pollution.[265] Such datasets have little to do with information about individuals, and do not fall under the purview of the FIPs.

The FIPs do not hinder releasing datasets with non-personal data. Hence, strict compliance with the FIPs does not necessarily interfere with releasing public sector information. Some suggest that "[m]ost open datasets have nothing personal to be protected in them (e.g.: digital maps, public budgets, air pollution measurements etc.)."[266]

But even for non-personal data, caveats must be mentioned. First, sometimes there may be non-privacy related arguments against releasing data. For instance, some information may have to remain confidential because of state security, such as information regarding critical infrastructure locations.[267] Second, as discussed in the next section, a dataset with non-personal data may, on closer inspection, include information about an individual.

In sum, we distinguish between four data categories with different risk levels: raw personal data, pseudonymous data, anonymized data, and non-personal data. However, the categories cannot be neatly distinguished in practice, as discussed next.

E. FUZZY BOUNDARIES

The borders between the four data categories are fuzzy. While many data privacy laws make a distinction between personal data and anonymized data, computer science suggests that the distinction is a matter of degree rather than kind.[268] Irreversible anonymization is difficult – perhaps impossible.[269] Apart from that, it is possible to distinguish sub-categories within the four categories. For instance, Zevenbergen et al. distinguish between three types of purportedly anonymized data, with different levels of re-identification risk.[270] And it is debatable whether data about an individual tied to his or her social security number should be seen as raw personal data, or as pseudonymized data.

---

265 Molinari & Marsh, *supra* note 254 236, at 311.
266 Molinari & Marsh, *supra* note 255, at 311; *see also*, Narayanan et al., *supra* note 219, at 21.
267 *See* Conley, Datta, Nissenbaum & Sharma, *supra* note 236, at 827.
268 *See e.g.*, Arvind Narayanan & Vitaly Shmatikov, *Myths and Fallacies of "Personally Identifiable Information"*, 53(6) COMM. ACM 24 (2010); Ohm, *supra* note 264; Matthijs R. Koot, *Measuring and Predicting Anonymity* (2012) (Ph.D thesis, University of Amsterdam), https://cyberwar.nl/d/PhD-thesis_Measuring-and-Predicting-Anonymity_2012.pdf.
269 Arvind Narayanan & Vitaly Shmatikov, supra note 268, 26.
270 Zevenbergen, Brown, Wright & Erdos, *supra* note 264, at 22.

Whether data are sufficiently anonymized is difficult to assess in advance. This is especially so, as more datasets may become available that enable "jigsaw identification."[271] The more data public sector bodies release, the higher the potential for combining data and thus creating information that can identify people.[272] The Obama administration recognizes this, and urges departments and agencies to perform a risk analysis.[273]

Even purportedly non-personal data can provide information about an individual. For example, a dataset with local air pollution levels contains non-personal data. However, if the dataset says that ZIP code 94720 is the most polluted, and the only business in that ZIP code is a one-man business, the pollution level in that ZIP code can say something about the business owner – namely that he or she is likely polluting. We do not suggest that privacy should enable business owners to escape responsibility for polluting. We merely want to illustrate that even datasets with non-personal data can provide information about an individual, for instance after linking datasets.[274]

In conclusion, we distinguish between four data categories with different risks levels: raw personal data, pseudonymous data, anonymized data, and non-personal data. The next section shows that open data should not be considered the only route when arguments for disclosure outweigh privacy interests. Options other than releasing data as open data are also available, such as disclosing data with access or re-use restrictions.

VI. TYPES OF DISCLOSURE

A maximalist approach to publishing public sector information as open data might imply that a public sector body should not impose any conditions on accessing or re-using public sector information. But a more moderate view is that public sector bodies should be allowed to impose conditions for access and re-use, if this is required to protect privacy interests.

We distinguish between three types of disclosure with different degrees of openness: (i) restricted access, (ii) restricted use, and (iii) open data. Restrictions on access and restrictions on re-use can be combined. Some forms of access and re-use restrictions do not comply with certain definitions of "open" data.[275] But sometimes disclosing data with restrictions is better than not disclosing at all.[276]

---

271 Narayanan et al., *supra* note 219, at 5-7. The phrase "jigsaw" identification is from O'Hara, *supra* note 76, at 40.

272 Narayanan et al., *supra* note 219, at 5-7; *see also* U.S. GEN. ACCOUNTABILITY OFFICE, *supra* note 81, at 107.

273 There, the risk is called the "mosaic effect" (Memorandum M-13-13, *supra* note 27, at 9-10).

274 For example, a health insurance company might use the data to calculate the health risks of the pollution, and might charge some people higher prices for coverage.

275 For example, if data are made available for non-commercial uses only, this runs counter to the open data principles set out in Section I ("open" implies that data can be used for any purpose). The same is true if only certain types of users are given access ("open" implies that data can be used by anyone). *See supra* Section II.A.

276 Scassa arrives at a similar conclusion about balancing privacy and public access: "As the experience of courts and tribunals shows, it may sometimes be necessary to place limits on the digital disclosure of some of the information in a 'public' record in order to achieve this balance." Scassa, *supra* note 73, at 404.

A. DISCLOSURE WITH ACCESS RESTRICTIONS

The first way to balance privacy and open data policy is by restricting access. To achieve a particular objective that underpins open data, it might not be necessary to allow everyone access, or to allow access to the raw data held by a public sector body. Data can be disclosed to particular groups for particular purposes, rather than to anybody for any purpose. Completely blocking data release on the one hand, and releasing data as open data on the other hand, can be seen as two extremes on a continuum. Disclosing data with restrictions is in-between those two extremes.

There are various ways to disclose data, which bring different risk levels. For example, Zevenbergen et al. distinguish open data from "restricted" disclosure, "managed access," "interactive methods," and "hybrid" methods.[277]

With "restricted" disclosure, data are only disclosed "to persons or organisations on request, refusing dissemination when the level of risk is considered too high."[278] Zevenbergen et al. suggest, for instance, that it is riskier to disclose data to a company than to academic researchers.[279] One problem with this type of disclosure is that it is hard to monitor what a receiving party does with the data.

With managed access, "[t]hird parties can query the dataset and conduct statistical (or other) analysis. Such an approach allows the researcher to ascertain exactly who accesses the datasets, while maintaining control over its dissemination."[280] For instance, researchers might have to visit the offices of the public sector body to inspect data.[281]

An example of an interactive method is "differential privacy."[282] As Zevenbergen et al. explain:

> Differential Privacy . . . only gives statistical answers to queries about an underlying dataset. To protect privacy even further, a certain amount of noise is added to the disclosed statistical data. In principle, differential privacy offers a lower risk for privacy, but there are certain limitations to this approach that need to be understood. For example, the uncertainty related by the addition of noise to the data can be exhausted, which means the dissemination must then stop. [283]

(emphasis omitted)

Hybrid approaches are also possible. For instance, parts of a dataset could be disclosed publicly, while other parts of the set could be kept confidential, or could be disclosed with strict access restrictions.[284]

---

277 Zevenbergen, Brown, Wright & Erdos, *supra* note 264, at 28-29.
278 *Id.* at 28.
279 *Id.* at 14-16. Similarly, Narayanan et al. say that restricted access "is a good solution" to enable scientific research without releasing data as fully open data. Narayanan et al., *supra* note 219, at 20.
280 Zevenbergen, Brown, Wright & Erdos, *supra* note 264, at 29.
281 *See id.* at 15.
282 Cynthia Dwork, *Differential Privacy*, *in* ENCYCLOPEDIA OF CRYPTOGRAPHY AND SECURITY 338, 338-40 (Henk C.A. van Tilborg & Sushil Jajodia eds., 2011).
283 Zevenbergen, Brown, Wright & Erdos, *supra* note 264, at 29.
284 *Id.*

In sum, sometimes a compromise between openness and privacy can be found by releasing data with access restrictions. Apart from access restrictions, it is also possible to restrict re-use, as discussed next.

B. DISCLOSURE WITH RE-USE RESTRICTIONS

Another way to strike a balance between privacy and open data policy is by applying restrictions on re-use of the disclosed data.[285] For instance, re-use restrictions can come in the form of licenses.[286] The license could require re-users to not re-identify data. Such measures have been used in practice. For example, on the website of the U.S. Healthcare Cost and Utilization Project, data users can purchase data sets – but if they purchase a dataset, they must sign an agreement that "expressly prohibits any attempt to identify individuals."[287]

The Article 29 Working Party suggests that the license should "prohibit license-holders from using the data to take any measure or decision with regard to the individuals concerned."[288] The license should also require "the license-holder to notify the licensor in case it is detected that individuals can be or have been re-identified."[289] As proper anonymization is difficult, in some situations, anonymized datasets should only be released under a license regime, rather than as fully open data. The higher the risk of de-anonymization, the more reason to tie a license to a dataset.

Access and re-use restrictions can also be combined. For instance, researchers could be required to visit the office of a public sector body to inspect a dataset: an access restriction. But at the same time, the researchers could be required to not try to re-identify people in the dataset: a re-use restriction.

C. DISCLOSURE AS OPEN DATA

The third access type is releasing data as fully open data: with no access or re-use restrictions. For instance, perhaps some personal data included in lobbying or company registers should be released as open data. Restricting access or re-use might make it too difficult to analyze the influence of lobbyists or to hold companies accountable.[290]

---

285 Solove makes a similar distinction between "access restrictions" and "use restrictions." Solove, *supra* note 9, at 1169-70.

286 Article 29 Data Protection Working Party, Opinion 06/2013, *supra* note 205, at 25-26.; *see also* Narayanan et al., *supra* note 219, at 18. A question that falls outside the scope of this paper is the legal basis for such licenses. In some countries, the public sector might have a type of intellectual property right on the dataset; in other countries the public sector body could invoke general contract law to impose a license on the dataset.

287 HEALTHCARE COST AND UTILIZATION PROJECT SID/SASD/SEDD APPLICATION KIT, *Data Use Agreement for the State Databases from the Healthcare Cost and Utilization Project Agency for Healthcare Research and Quality* 1, 24 (June 17, 2015), http://www.hcup-us.ahrq.gov/db/state/SIDSASDSEDD_Final.pdf.

288 Article 29 Data Protection Working Party, Opinion 06/2013, *supra* note 205, at 25.

289 *Id.*

290 *See, e.g.*, Jonathan Gray & Tim Davies, *Fighting Phantom Firms in the UK: From Opening Up Datasets to Reshaping Data Infrastructures?*, SOCIAL SCIENCE RESEARCH NETWORK (May 27, 2015), available at http://dx.doi.org/10.2139/ssrn.2610937; Transparency International UK, *How Open Data Can Help Tackle Corruption* (June 2015), available at http://www.transparency.org.uk/publications/15-publications/1287-how-open-data-can-help-tackle-corruption-policy-paper.

In conclusion, sometimes a balance can be struck between open data goals and privacy by disclosing data with access or re-use restrictions, rather than as fully open data. Hence, a public sector body must first assess whether a dataset should be disclosed at all. If it is decided that data should be disclosed, the next question is whether the data should be released with access or re-use restrictions, or as fully open data.

VII. A CIRCUMSTANCE CATALOGUE TO INFORM DISCLOSURE DECISIONS

The above suggests that public sector bodies should decide on a case-by-case basis whether, and under which conditions, a dataset should be disclosed.[291] Narayanan et al. note that "[e]ach dataset has its own risk-benefit tradeoff, in which the expected damage done by leaked information must be weighed against the expected benefit from improved analysis."[292] The researchers add that "[b]oth assessments are complicated by the unpredictable effects of combining the dataset with others, which may escalate both the losses and the gains."[293]

There is not one clear-cut rule to decide whether datasets including or based on personal data should be disclosed. The lack of a hard-and-fast rule is not surprising. As discussed in Section III, the problem of balancing privacy and open data interests can be seen as a modern version of the problem of balancing privacy and public sector transparency.

The objectives behind open data policies and corresponding public interests involved merit closer scrutiny; this allows for differentiation that is necessary for balancing the interests involved.[294] The general FIPs guidance suggesting a balance between privacy and other interests is not detailed enough in the case of open data. We propose that a circumstance catalogue can help to decide whether and how to release data.[295] The circumstance catalogue lists circumstances, or factors, that should be considered when assessing whether, and under which conditions, a dataset should be released, as well as different options for how it should be released. We provide a list as a starting point for a debate – the list is not meant to be exhaustive or final. The circumstance catalogue can be extended, for instance, by taking inspiration from case law, freedom of information law, and guidelines regarding open data and privacy.

We mention some rules of thumb regarding re-identification risks and releasing data. One rule of thumb is that raw personal data should generally not be released as fully open data, unless there is a compelling public interest argument for choosing this route for disclosure over other available options.[296] We argue that pseudonymous data must generally be treated as a type of personal data,

---

291 Many authors arrive at that conclusion. *See, e.g.*, Katleen Janssen & Sara Hugelier, *Open Data: A New Battle in an Old War Between Access and Privacy?*, *in* DIGITAL ENLIGHTENMENT YEARBOOK 2013 190, 199 (Mireille Hildebrant et al. eds., 2013).

292 Narayanan et al., *supra* note 219, at 12.

293 *Id. See also id.* at 13, 15.

294 *See* Scassa, *supra* note 73, at 405.

295 *See* for a similar approach, balancing interests in access to court records against other considerations Conley, Datta, Nissenbaum & Sharma, *supra* note 236, at 797-798.

296 A similar conclusion is reached by Narayanan et al., *supra* note 219, at 15 ("[I]t almost never will be the case that an unlimited release of a dataset to the entire public will be the optimal choice.").

rather than as anonymous data. On the other hand, non-personal data can generally be released as open data. For purportedly anonymized data, it is more complicated. As stated previously, irreversible anonymization is difficult, and perhaps impossible to achieve.[297] Therefore, in some cases anonymized data should not be released as fully open data.

A. WEIGHT OF THE GOALS PURSUED

The goals pursued by disclosing data are relevant. The consideration is not only what the (theoretical) aim of the public body is. An assessment might also be made of the most likely uses of the data by other public bodies, the private sector, and citizens. True, this runs counter to the idea behind open data that serendipitous re-use is positive, and that it is impossible for the government to predict potential uses.[298] But it is naïve to assume that uses will all be benevolent.

What is the primary goal pursued with releasing data and how important is releasing this type of information, in this form, to achieving that goal? Could the objective be adequately addressed by disclosing information in a less privacy-sensitive form? Is it likely that the data will be used primarily by the press or similar public watchdogs, or are the data primarily interesting for commercial purposes?[299] The more relevant data are to key aspects of democratic participation, the stronger the case for release as open data. As Solove notes, when deciding whether to release personal data, political transparency has more weight than pure commercial interests of re-users:

> Access should be granted for uses furthering traditional functions of transparency such as the watchdog function; access should be denied for commercial solicitation uses because such uses do not adequately serve the functions of transparency. Rather, such uses make public records a cheap marketing tool, resulting in the further spread of personal information, which is often resold among marketers.[300]

Furthermore, not all uses of public sector information are equal before the law. Additionally, the national legal system makes a difference. For instance, the strength of rights to access to information and the discretionary space for public authorities differs from country to country. For example, in the United States, the First Amendment influences decisions regarding data disclosure.[301] In Europe, access to information to foster political transparency also has backing in human rights treaties.[302] But in Europe, legal privacy and data protection rights have more relative weight than in the United

---

297 *See supra* Section V.E ("Fuzzy Boundaries"). See also Arvind Narayanan & Vitaly Shmatikov, supra note 268, 26.

298 The G8 OPEN DATA CHARTER *supra* note 3, for example contains the pledge of governments to ensure "…that the data are available to the widest range of users for the widest range of purposes*"* (pledge 22) Assessing the market for public sector information based products and services in the U.K, Deloitte concludes that "it is hard to foresee specifically where innovation might take place in the UK. Often innovation takes place in areas which are hard to predict..." DELOITTE, MARKET ASSESSMENT OF PUBLIC SECTOR INFORMATION, *supra* note 40, at 41.

299 The U.S. Supreme Court noted that different purposes have different weights in the context of inspecting and copying judicial records (Nixon v. Warner Commc'ns, Inc., 435 U.S. 589 (1978)). In the FOIA context, the Supreme Court arrived at a similar conclusion (Dep't of Justice v. Reporters Comm. for Freedom of the Press, 489 U.S. 749, 773 (1989).

300 Solove, *supra* note 9, at 1192.

301 *See id.* at 1200-6. *See also* Sorrell v. IMS Health Inc., 131 S. Ct. 2653 (2011).

302 Mireille van Eechoud & Katleen Janssen, *supra* note 119, at 483-488.

States.[303] An important factor in this respect is the role of people whose data are considered for release. Do the data concern somebody who holds a public function or a powerful position? What is the level of responsibility of the person? To what extent is the information needed in open, machine-readable form in order to facilitate democratic accountability? The higher the level of responsibility of the person, the more likely it is that transparency trumps privacy interests.

While access to information to foster democratic transparency has backing in constitutions and human rights documents, the legal backing of releasing information for business opportunities or for improving public sector efficiency is less evident.[304] If there is a good case for sharing data within the public sector because this contributes to efficient government, governments should regulate such sharing with specific laws that contain appropriate safeguards. Cost and efficiency savings in and of themselves may not outweigh the protection of individual privacy unless there are other overriding concerns about, for example, public accountability, corruption, or the exercise of democratic oversight.

Apart from the difference in national legal systems, the weight of the goal also depends on the national situation. For example, in a country where there are many problems with corruption by state officials, disclosing detailed wealth records of public functionaries makes more sense than in a country with virtually no corruption. And in some countries there may be more widespread acceptance of the public disclosure of salaries.[305]

B. WEIGHT OF THE PRIVACY INTERESTS

Arguments against releasing data, or against releasing data without restrictions, include the following: there are considerable risks associated with releasing the data; the potential harm is serious, rather than a minor inconvenience; the privacy of many people (not a few) is at risk; and the privacy threat is immediate rather than remote. For instance, a theoretical privacy infringement has less weight than would a clear danger. A clear privacy danger might occur, for example, with a dataset containing names of people with HIV. People could be discriminated against if it becomes publicly known that they have HIV.

The nature of the harm also matters: for example, if the data relate to people fulfilling public functions and concern professional conduct, a risk of reputational harm is unlikely to be of concern (unless there is doubt about the accuracy of the data). That would be the case with disclosing expenses claims. If disclosure leads to a security risk, e.g., disclosing an itinerary or detailed information about a politician's movements, the case is different.

Expectations of privacy can also be a factor. How were the data collected? If there was a promise or understanding of confidentiality, the case is different than if people have volunteered data after they were warned of future possible disclosures. Because of asymmetry in information

303 *See generally* Kranenborg, *supra* note 141.

304 The Charter of Fundamental Rights of the European Union does recognize the right to do business (Charter of Fundamental Rights of the European Union, *supra* note 70, Art. 16).

305 For instance, in Finland, the tax authorities disclose the income of people whose income exceeds certain thresholds (*see* Tietosuojavaltuutettu v Satakunnan Markkinapörssi Oy, Satamedia Oy, C-73/07 [CJEU] Dec. 18, 2008).

relationships between public authorities and citizens, it cannot be readily assumed that data was truly volunteered.

In conclusion, a case-by-case analysis is required when deciding whether to release data, and whether the data can be disclosed as fully open data, or whether access or use should be restricted. We proposed a starting point for a circumstance catalogue that would help to assist in decisions about data disclosure.

VIII. CONCLUSION

Open data are held to contribute to a wide variety of social and political goals – including strengthening transparency, public participation and democratic accountability, promoting economic growth and innovation, and enabling greater public sector efficiency and cost savings. But releasing datasets as open data may threaten privacy, for instance if they contain personal or re-identifiable data. Potential privacy problems include chilling effects on people communicating with the public sector, a lack of individual control over personal information, and discriminatory practices enabled by the released data.

Can privacy and related interests be respected, while not unduly hampering open data benefits? The Fair Information Principles (FIPs), as expressed in the OECD Privacy Guidelines, provide a framework to balance privacy and other interests. From a FIPs perspective, the main problem with open data is that open data can be used by anyone for any purpose. A complete lack of re-use restrictions would clash with the purpose specification principle of the FIPs. It follows from the purpose specification principle that personal data should only be collected for a purpose that is specified in advance, and that those data should not be used for incompatible purposes.

Compromises are possible to balance privacy and open data interests. We distinguish between four data categories with different risk levels: raw personal data, pseudonymous data, anonymized data, and non-personal data. With raw personal, no attempt has been made to make identification harder. Pseudonymous data are data for which the individual's name is changed to another unique identifier. Anonymized data are ex-personal data; people cannot be re-identified in the dataset. Non-personal data, such as data about weather conditions or public transport times, never contain personal data.

Non-personal data can generally be released without restrictions as fully open data. As a rule of thumb, raw personal data should not be released as fully open data. Pseudonymous data must generally be treated as a type of personal data – not as anonymous data. Anonymized data is more complicated. Anonymized data can, in theory, be disclosed as open data, without re-use restrictions. However, irreversible anonymization is exceedingly difficult, and perhaps impossible. And even in aggregated and purportedly anonymized data, individuals can sometimes be re-identified. Therefore, some purportedly anonymized datasets should only be disclosed with access and re-use restrictions.

Sometimes, a compromise can be found by releasing anonymized data with access and re-use restrictions. Restricting openness can be done in various ways. For instance, the public sector body could attach a license to the data, requiring the re-user to only use certain data for a certain purpose (say medical research) and to promise not to re-identify the data. Other limitations on openness can

also be envisaged. For instance, if a research interest is important, but the personal data are sensitive, researchers could be required to visit the lab where the data are held.

Hence, a case-by-case analysis is required when deciding whether to release data, and whether the data can be disclosed as fully open data, or whether access or use should be restricted. To assist in decisions about data disclosure, a circumstance catalogue may be of help: a list of circumstances to consider when deciding about releasing data. For instance: what is the goal pursued by releasing the data? Is there another way to pursue that goal? What are the risks involved with releasing the data? Are the privacy-related risks negligible or probable? If the risk materializes, what is the harm that results? Is the privacy of a few or of millions of people at stake?

In conclusion, in many instances public sector datasets that contain, or are based on, personal data should not be released as fully open data. When arguments for disclosure do outweigh privacy interests, open data should not be considered the only route. Other options might include disclosing information with access or re-use restrictions.

* * *