Networked Policy Series

September 2016

Translating Research for Action:
Ideas and Examples for
Informing Digital Policy

Privacy and Open Data

Research Briefing

Alexandra Wood, David R. O’Brien, and Urs Gasser

for more from this series visit
cyber.harvard.edu

BERKMAN
KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

Privacy and Open Data

Research Briefing

Alexandra Wood, David R. O’Brien, and Urs Gasser

Suggested citation: Wood, Alexandra, O'Brien, David and Gasser, Urs, Privacy and Open Data Research Briefing (September 26, 2016). Networked Policy Series, Berkman Klein Center Research Publication No. 2016-16. Available at SSRN: http://ssrn.com/abstract=2842816

23 Everett Street | Second floor | Cambridge, Massachusetts 02138
+1 617.495.7547 | +1 617.495.7641 (fax) | http://cyber.harvard.edu
press@cyber.harvard.edu

Overview

Political leaders and civic advocates are increasingly recommending that open access be the "default state" for much of the information held by government agencies.[1] Over the past several years, they have driven the launch of open data initiatives by hundreds of national, state, and local governments.[2] These initiatives are founded on a presumption of openness for government data and have led to the public release of large quantities data through a variety of channels.[3] At the same time, much of the data that have been released, or are being considered for release, pertain to the behavior and characteristics of individual citizens, highlighting tensions between open data and privacy.[4]

The Berkman Klein Center for Internet & Society at Harvard University ("BKCIS") has prepared this research briefing on open data and privacy as a guide for decision-makers who are shaping the development of open government data releases. In this briefing document, enabled by generous support by the Ford Foundation and building on deep institutional knowledge on privacy approaches as well as consultations with open data stakeholders from the public and private sectors, the BKCIS team seeks to summarize selected research findings into practical considerations and takeaways for non-academic actors.

Part I of this briefing is an **ecosystem map**, providing an overview of key developments related to privacy and open government data at the U.S. federal, state, and local levels, including landscape shifts, actors, drivers (focusing on the technological, legal, regulatory, policy-based, and behavioral), current and emerging issues related to privacy and openness, and values underlying decisions related to privacy and open data.

Part II is an **action map**, surveying key ongoing and emerging issues in open data and privacy, categorized by governance approaches and values stakeholders are considering as they respond to related concerns.

Part III is a **navigation aid** to serve as a guide for decision-makers who seek to identify and pursue values-based goals – independently or collaboratively – in the complex, pressing, and ongoing dialogues and debates regarding privacy and open data.

[1] See, e.g., Exec. Order No. 13,642, 3 C.F.R. 244 (2014) (Making Open and Machine Readable the New Default for Government Information).

[2] See Data.gov, Open Data in the United States, https://www.data.gov/open-gov/ (last visited Sept. 20, 2016) (reporting that 40 U.S. states, 48 U.S. cities and counties, 52 foreign countries, and 164 international regions have established online repositories of open data).

[3] See Paul M. Schwartz, Privacy and Participation: Personal Information and Public Sector Regulation in the United States, 80 Iowa L. Rev. 553 (1995); Harlan Yu & David G. Robinson, The New Ambiguity of "Open Government," 59 UCLA L. Rev. Discourse 178 (2012).

[4] See the selection of articles from the 19th Annual BCLT/BTLJ Symposium: Open Data: Addressing Privacy, Security, and Civil Rights Challenges, published in Volume 30, Issue 3, of the Berkeley Technology Law Journal (2015), http://www.btlj.org/2016/05/volume-30-issue-3.

I. Ecosystem Map

The ecosystem map that follows offers:

- A brief description of the overarching **tectonic shifts** in our increasingly data-driven world that underlie many of the recent developments related to open data and data privacy; and
- A **snapshot** of today's open data and privacy landscape, including key actors, drivers, and tensions.

1. Tectonic Shifts

In light of recent developments in open data and privacy, we observe that foundational changes at the intersection of technology, society, law, behavior, and related spheres are rapidly disrupting and energizing familiar institutions and activities:

* **Falling costs of data collection, processing, storage, analysis, and release** enabled by advances in technology are lowering barriers for institutions, large and small, across public and private sectors to collect, store, and analyze large quantities of data.[5] Examples of new data-intensive activities in a diverse range of areas include:
    - Consumer use of social networking platforms to connect with friends and family, and commercial use of such services for personalized service delivery and targeting of advertisements.
    - Development of smartphone apps providing real-time, customized information to users, while enabling companies to collect streams of data from individuals.
    - Use of data modeling techniques across a wide range of new services, such as services offering credit risk scoring, lending, and personal finance tools.

* **Demand for data is growing across sectors**, as businesses, governments, researchers, and other individuals increasingly engage in data-driven analysis, decision-making, and service delivery, and otherwise derive tremendous value from data. Examples include:
    - Businesses of all types, from app developers to brick and mortar retailers, are continually seeking new sources of rich, fine-grained data to guide their investment and commercial decisions, offer personalized and up-to-the-minute products and services, and tailor their advertising and marketing campaigns, all of which in turn drive innovation and competition in the marketplace.
    - Government agencies at all levels are moving to collect and analyze data to inform decision-making and service delivery in areas such as public safety, health and human services, infrastructure, and education.[6]

[5] See President's Council of Advisors on Science and Technology, Big Data and Privacy: A Technological Perspective (May 2014), https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.

[6] See Stephen Goldsmith & Susan Crawford, The Responsive City: Engaging Communities Through Data-Smart Governance (2014).

- Researchers, journalists, educators, and activists are regularly pursuing new data sources to support their research, education, and civic activities.

* **Data privacy risks are growing due to advances in analytical capabilities** and calling into question traditional approaches to privacy. A number of high-profile attacks have demonstrated that it is possible to re-identify or learn details about individuals described in releases of data, even when commonly used techniques for de-identifying data or generating aggregate statistics are applied:[7]
    - De-identified hospital records have been re-identified using information found in public records, such as an individual's date of birth, gender, and ZIP code.[8]
    - De-identified records of video ratings by Netflix customers were re-identified using information from other web sites such as the Internet Movie Database.[9]
    - Researchers have shown it is possible to use Amazon's product recommendation system to infer information about an individual user's transactions.[10]
    - Researchers demonstrated they could confirm whether a particular person is included in a database of aggregate mixtures of genomic DNA collected from hundreds of individuals, and thereby determine that an individual participated in a research study investigating a particular disease.[11]

* **New tools for privacy protection are being developed** to address the weaknesses of common approaches to privacy. A wide range of procedural, economic, educational, legal, and technical controls for data privacy and security are now available to institutions that collect, store, analyze, and publish data.[12]
    - Advanced data sharing models such as contingency tables, synthetic data, data visualizations, interactive mechanisms, and multiparty computations are used across government and industry and can provide stronger privacy protection than releases relying solely on traditional de-identification techniques.[13]
    - These data sharing models are also compatible with strong mathematical definitions of privacy from the computer science literature, such as differential privacy, which are provably resilient to a large class of potential misuses.[14]

[7] More generally, these examples illustrate the fundamental law of information recovery, which "states, informally, that 'overly accurate' estimates of 'too many' statistics completely destroy privacy." Cynthia Dwork & Guy N. Rothblum, Concentrated Differential Privacy, Working Paper (2016) (citing Irit Dinur & Kobbi Nissim, Revealing information while preserving privacy, Proceedings of PODS 202–210 (2003)).

[8] See Latanya Sweeney, Weaving Technology and Policy Together to Maintain Confidentiality, Journal of Law, Medicine and Ethics (1997).

[9] See Arvind Narayanan & Vitaly Shmatikov, Robust De-anonymization of Large Sparse Datasets, Proceed ings of the 2008 IEEE Symposium on Security and Privacy 111 (2008).

[10] See Joseph A. Calandrino et al., "You Might Also Like:" Privacy Risks of Collaborative Filtering, IEEE Symposium on Security and Privacy (2011).

[11] See Nils Homer et al., Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-density SNP Genotyping Microarrays, 4 PLoS Genetics 8 (2008).

[12] For a catalog illustrating various privacy and security controls that are available for data releases, see Micah Altman, Alexandra Wood, David R. O'Brien, Salil Vadhan & Urs Gasser, Towards a Modern Approach to Privacy-Aware Government Data Releases, 30 Berkeley Tech. L.J. 1967, 2015-31 (2015).

[13] See id.

[14] See id.

- Tiered access mechanisms can be constructed for closely calibrating appropriate combinations of privacy and security controls to different risks and intended uses at each stage of the information lifecycle.[15]

2. Snapshot of Today's Open Data & Privacy Landscape

The tectonic shifts outlined above are specifically reflected in recent developments observed within the current and emerging open data and privacy landscape.

* **Demand for access to government data is high.** Journalists, civic groups, researchers, and other members of the public seek to reuse government data in ways that advance transparency and accountability, improve the effectiveness and responsiveness of government agencies, promote innovation, and create economic benefits.
    - Scientists seek to combine data from government records and businesses such as telephone and utility providers with traditional sources.[16] For example, researchers apply big data methods to analyze data from sensors and administrative records to guide urban policymaking and operations.[17]
    - Data analysts and technology companies seek to use data collected by government agencies to develop services such as apps offering real-time public transit tracking,[18] and searchable, interactive maps for reviewing 311 complaints and inspection violations filed for apartment buildings across a city.[19]

* **Technological advances in data collection, processing, and sharing are leading government agencies at all levels to adopt open data policies.** Such policies call for open access to be the "default state" for government information.[20]
    - Open data policies encourage government agencies to adopt a presumption of openness, to the extent the law allows, and publish information online in open formats that can be accessed and analyzed through a variety of applications.[21]
    - To date, 40 states and 48 cities and counties have launched open data web sites,[22] offering tools for analyzing and downloading large volumes of data.
    - This represents a fundamental shift in the way governments release data, as agen-

[15] See id.

[16] See Daniel Tumminelli O'Brien, Robert J. Sampson & Christopher Winship, Econometrics in the Age of Big Data: Measuring and Assessing "Broken Windows" Using Administrative Records (Bos. Area Research Initiative, Working Paper No. 3, 2013).

[17] See Steven E. Koonin, Ctr. for Urban Sci. & Progress, The Promise of Urban Informatics (2013), http://cusp.nyu.edu/wp-content/uploads/2013/07/ CUSP-overview-May-30-2013.pdf.

[18] See Stephen Goldsmith & Susan Crawford, The Responsive City: Engaging Communities Through Da ta-Smart Governance 78-79 (2014).

[19] See Karen Eng, Check before you rent: How a TED Fellow is holding New York City landlords accountable, TEDBLOG (Apr. 10, 2015), http://blog.ted.com/how-ted-fellowyalefoxis-holding-new-york-city-landlords-accountable.

[20] See, e.g., Exec. Order No. 13,642, 3 C.F.R. 244 (2014) (Making Open and Machine Readable the New Default for Government Information), https://www.gpo.gov/fdsys/pkg/CFR-2014-title3-vol1/pdf/CFR-2014-title3-vol1-eo13642.pdf.

[21] E.g., Peter R. Orszag, Office of Mgmt. & Budget, Exec. Office of the President, M-10-06, Memorandum on Open Government Directive (Dec. 8, 2009), http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf.

[22] See Open Government, Data.gov, https://www.data.gov/open-gov/ (last visited Sept. 20, 2016).

cies are directed to "proactively use modern technology to disseminate useful information, rather than waiting for specific requests under FOIA."[23]

* **Governments rely on a narrow subset of tools to analyze and mitigate privacy risks,** and release decisions are often based on an ad-hoc balancing of interests that does not systematically address the privacy risks identified in the scientific literature.[24]

- A wide range of privacy and security measures can be implemented and calibrated to the risks and intended uses in a given data release. However, most agencies rely exclusively on a single intervention: withholding or redacting records that contain information deemed to be identifying.[25]
- Despite attempts by agencies to redact personal information prior to release, it is in some cases evident that released datasets could reveal sensitive details about individuals.[26] It is now understood that privacy is not simply associated with the presence of specific types of information in a released set of data, as harm can also stem from what one can infer about individuals from the data release as a whole or when the data are linked with other data sources.[27]
- Agencies' data release decisions also likely result in the withholding of useful information that could be safely shared using alternative data sharing models.

* **Guidance on interpreting and applying regulatory standards for privacy protection is limited,** contributing to variations in handling of data across government agencies.[28]

- While high-level reports on techniques for protecting privacy are widely available, there is limited practical guidance on evaluating privacy risks and selecting and implementing privacy interventions in specific settings. In particular, agencies lack sufficient guidance for determining when it is appropriate to use newly emerging privacy tools for their data publications.[29]
- Agency- or sector-specific regulatory requirements and an agency's initial choice of release mechanism often dictate its approach to privacy. As a consequence, similar privacy risks – and, in some cases, even identical datasets – are managed differently by different government actors.
- Taken together, the laws, policies, and practices compelling and constraining government releases of information can create uncertainty for government data managers, discourage some data sharing, and fall short of providing strong privacy protection for individuals.

[23] Id.

[24] See Micah Altman, Alexandra Wood, David R. O'Brien, Salil Vadhan & Urs Gasser, Towards a Modern Approach to Privacy-Aware Government Data Releases, 30 Berkeley Tech. L.J. 1967, 2006-07 (2015).

[25] See id. at 2015-31.

[26] See id. at 2048-70.

[27] See, e.g., Arvind Narayanan & Vitaly Shmatikov, Robust De-anonymization of Large Sparse Datasets, Proceedings of the 2008 IEEE Symposium on Research in Security and Privacy 111 (2008); Latanya Sweeney, k-anonymity: A Model for Protecting Privacy, 10 Int'l J. of Uncertainty Fuzziness & Knowledge-based Systems 557 (2002).

[28] See Micah Altman, Alexandra Wood, David R. O'Brien, Salil Vadhan & Urs Gasser, Towards a Modern Approach to Privacy-Aware Government Data Releases, 30 Berkeley Tech. L.J. 1967, 2007-09 (2015).

[29] See, e.g., U.S. General Accounting Office, GAO-01-126SP, Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information 105 (2001), http://www.gao.gov/new.items/d01126sp.pdf.

II. Action Map

The action map that follows offers:

1. A broad-brush taxonomy of different **governance approaches** to addressing digital information questions in the privacy and open data space, including examples of representative uses of each approach by key actors; and
2. A brief identification of the **values** that seem to be embedded in each of these approaches, which inform key actors' privacy commitments, with related questions for stakeholders interspersed where appropriate.

Governance approaches: As we consider the new ways in which "information is created, shared, accessed, and used in the globalized digital world" of privacy and open data, we see five broad categories of governance approaches: **technology-based, market-based, human-centered, law-based, and blended governance.**[30]

[30] Urs Gasser, Perspectives on the Future of Digital Privacy, 134 Zeitschrift für Schweizerisches Recht [ZSR] 335, 341-42, 444 (2015).

Action Map

	Tech-based	Market-based	Human-centered	Law & policy	Blended
Description	Technologies for enhancing privacy in open data	Market incentives for protecting privacy in open data	Mechanisms that rely on one or more forms of inter-personal engagement	Addition of new or reform of existing laws, regulations, policies, and agreements at all levels to address privacy challenges in open data	Interrelated use of more than one of the tech, market, human and law-based mechanisms to address multi-dimensional privacy problems in open data
Approaches[31]	SDL techniques Aggregate statistics Encryption Data visualizations Synthetic data Interactive query systems Secure multiparty computation Differential privacy Immutable audit logs	Collection fees Markets for personal data Access/use fees Fines	Transparency Notice Educational materials Public forums Blog posts Metadata Data asset registers Privacy dashboards Advisory committees Personal data stores	Privacy and information security management laws Data policies Privacy impact assessments Checklists Data use agreements	Tiered access Best practices
Examples	The U.S. Census Bureau applies various technical approaches to privacy when publishing data. For example, the **Synthetic Longitudinal Business Database (SynLBD)**[32] allows researchers to study economic data at the establishment level using synthetic data that do not reveal confidential information about businesses, and **OnTheMap**[33] provides an online interface for exploring the commuting patterns of U.S. workers, based on computations on synthetic data that satisfy a variant of differential privacy.	Commentators have suggested that governments could impose small **access/use fees**[34] that would make it more costly for criminals and others to misuse the data. While fees are not appropriate for all data releases, in some cases where privacy risks are high, the economic value of the data is great, and the release of the data is not compelled by statute, agencies may consider fees as one of many tools available.	The City of Seattle's open data initiative incorporates several human-based approaches to privacy. Examples include launching a **Privacy Initiative** to develop a citywide set of policies and practices for addressing privacy,[35] forming a **Privacy Advisory Committee** comprised of "privacy researchers, practitioners, and community representatives,"[36] and announcing plans to create a **Privacy Dashboard** to help open data managers analyze privacy risks associated with individual datasets and make decisions for handling them.[37]	Existing laws require agencies to have in place processes for screening the data they hold for privacy risks.[38] Examples include **privacy impact assessments**, with which federal executive agencies examine their information systems and specify the practices that will be put in place to mitigate privacy risks,[39] and **checklists**,[40] which agencies use to ensure appropriate disclosure limitation practices are followed when publishing datasets online.	Governments can implement **tiered access models**, which make data available to different categories of users through different combinations of legal, technical, and educational tools.[41] For example, public access to some data could be enabled for differentially private statistics computed from the data, and access to full datasets could be granted after application to and review by an oversight body and the signing of a data use agreement prescribing storage, use, and redisclosure of the data.

[31] For an extended discussion of the range of approaches to privacy that are available to governments in open data releases, see Micah Altman, Alexandra Wood, David R. O'Brien, Salil Vadhan & Urs Gasser, Towards a Modern Approach to Privacy-Aware Government Data Releases, 30 Berkeley Tech. L.J. 1967, 2015-31 (2015).

[32] See Ron S. Jarmin, Thomas A. Louis & Javier Miranda, Expanding the Role of Synthetic Data at the U.S. Census Bureau, CES 14-10 (2014), https://www2.census.gov/ces/wp/2014/CES-WP-14-10.pdf.

[33] See Ashwin Machanavajjhala et al., Privacy: Theory Meets Practice on the Map, IEEE 24th Int'l Conf. Data Engineering 277 (2008), http://lehd.ces.census.gov/doc/help/ICDE08_conference_0768.pdf.

[34] See, e.g., Kieron O'Hara, Transparent Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office, UK Cabinet Office (2011), https://www.gov.uk/government/publications/independent-transparency-and-privacy-review,

[35] See City of Seattle, Privacy Initiative, http://www.seattle.gov/tech/initiatives/privacy (last visited Sept. 20, 2016).

[36] City of Seattle, Privacy Advisory Committee, http://www.seattle.gov/tech/initiatives/privacy/privacy-advisory-committee (last visited Sept. 20, 2016).

Values	Emerging privacy technologies, formal privacy models in particular, can provide robust privacy protection for individuals and businesses. If finely matched to the specific privacy risks and intended uses in a given release, such tools can in some cases also bring gains in data utility over alternative approaches.	Market-based solutions such as access fees can be used to reduce some risk of informational harm, while also enabling uses of high economic value.	Human-centered approaches to privacy can be implemented to empower decision-makers to make data release decisions that are based on a more informed understanding of privacy risks and the community's consensus on balancing values such as transparency, utility, and privacy.	Law-based approaches can be used to direct actors to implement strong privacy protections, as well as provide mechanisms for transparency, accountability, and redress that can be paired with approaches from other categories.	Blended approaches can enable the use of controls that are finely tailored to the intended uses and privacy risks associated with a specific data release, and thereby bring gains in data privacy and utility.
Questions (Sample)	How can third-party open data developers be incentivized to incorporate new privacy technologies into their platforms?	How can agencies determine where access fees might be an appropriate solution?	How can the design of open data programs be informed by the expectations and preferences of members of the community?	How can laws and policies be updated to reflect recent advances in the understanding of data privacy risks and effective measures for privacy protection?	How can detailed guidance materials be developed to help open data managers and community representatives make data release decisions and choose appropriate privacy and security controls?

Flash Case Study

When the City of New York released more than 173 million records of taxi trips taken during 2013, a number of privacy-related vulnerabilities were discovered, illustrating some of the challenges of protecting privacy when releasing open data. For example, the NYC Taxi and Limousine Commission attempted to protect the privacy of taxi drivers by applying a cryptographic hash function to transform the driver medallion and license numbers contained in the dataset, but, given the uniform representation of these numbers, members of the public were able to uncover the medallion and license numbers in "less than two minutes."[42] Researchers illustrated the privacy risks associated with releasing "anonymized" taxi trip data. For example, they showed that trips originating at strip clubs and ending at residences reveal the home addresses of employees and patrons, and that locations and times of taxi trips could be matched with paparazzi photographs found online and used to identify the taxi trips taken by various celebrities.[43]

[37] See Sharon Griggins, Measuring Privacy in Municipal Open Data Sets, Knight News Challenge Entry from the University of Washington/City of Seattle Open Data/Privacy Team, https://www.newschallenge.org/challenge/data/entries/measuring-privacy-in-municipal-open-data-sets (Sept. 28, 2015).

[38] See, e.g., E-Government Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899.

[39] See Office of Mgmt. & Budget, Memorandum M-03-22, OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002 (2003).

[40] See, e.g., Data.gov, National/Homeland Security and Privacy/Confidentiality Checklist and Guidance, http://www.data.gov/sites/default/files/attachments/Privacy and Security Checklist.pdf (last accessed May 31, 2016).

[41] For a discussion of an approach to designing a tiered access mechanism, see Micah Altman, Alexandra Wood, David R. O'Brien, Salil Vadhan & Urs Gasser, Towards a Modern Approach to Privacy-Aware Government Data Releases, 30 Berkeley Tech. L.J. 1967 (2015).

[42] See Alex Hern, New York taxi details can be extracted from anonymised data, researchers say, The Guardian, June 27, 2014, https://www.theguardian.com/technology/2014/jun/27/new-york-taxi-details-anonymised-data-researchers-warn.

[43] See Anthony Tockar, Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset, Neustar Research, Sept. 15, 2014, https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset.

III. Navigation Aid (Identification of Key Opportunities)

As decision-makers move to respond to tectonic shifts in the open data and privacy landscape — both in terms of dealing with challenges related to privacy and utility and embracing tremendous opportunities afforded by new sources of open data — the following domains related to the “new societal operating system”[44] should be considered as important action areas:

1. **Data** – the creation of better data (and data about data) for making informed decisions;
2. **Values** – engagement in multistakeholder processes to study difficult normative problems;
3. **Design/instruments** – implementation of novel tools and approaches;
4. **Evaluation** – of outputs in the above categories.

Some key examples of potential action — but by no means the only forms of action — in each domain include:

* **Data:** Open data managers should consider adopting standardized processes for analyzing the data scheduled for release through their open data systems.

An information lifecycle model, combined with information security approaches to characterizing data uses and privacy risks, can provide a systematic framework for analyzing factors relevant to the management of specific sets of data.[45] Example factors to examine include the analytic value of the data, the intended uses of the data, characteristics of the data related to the sensitivity of the information and the potential for learning about individuals in the data, and the expected benefits from releasing the data.[46] These factors should inform the choice of privacy and security controls implemented in the information systems and open data release platforms used.

* **Values:** Open data managers should consider engaging in multistakeholder dialogues to surface the values of communities of data users and data subjects.

Open data managers should consult with privacy experts, potential consumers of open data, community leaders, and members of the public to develop an understanding of the relevant stakeholders’ values with respect to open data and privacy. In designing programs for data collection, storage, use, and release, governments necessarily make decisions weighing tradeoffs between protecting individual privacy interests and enabling a wide range of uses of information about individuals. Governments should consider hosting public forums, holding

[44] Urs Gasser, On Handling the Chances and Risks of a Digital Society 1 (2013) (unpublished manuscript; on file with authors).

[45] See Micah Altman, Alexandra Wood, David R. O’Brien, Salil Vadhan & Urs Gasser, Towards a Modern Approach to Privacy-Aware Government Data Releases, 30 Berkeley Tech. L.J. 1967 (2015).

[46] See id.

regular multistakeholder meetings, releasing lists of the data they hold and descriptions of the privacy and security safeguards put in place, and publishing details about data release decisions, in order to motivate public dialogues on the benefits of open data, the privacy risks associated with collecting and releasing data about individuals, and the safeguards that should be adopted.

* **Design/instruments:** Open data managers should design tailored data management and release programs and consider incorporating new procedural and technological solutions where appropriate.

Recent advances in science and technology offer the opportunity for conducting a systematic analysis of privacy risks and harms, as well as implementing a range of legal, technical, economic, procedural and educational interventions enabling different uses and types of protection.[47] Open data managers should consider calibrating their data programs to the intended uses of the information collected; the benefits of uses of the data; and the threats, vulnerabilities, and harms associated with activities involving personal data.[48] They should also explore the appropriateness of various instruments for privacy and security from the wide range of tools that are available, rather than rely on a small subset of controls such as de-identification techniques and binary access control. Examples of new and emerging interventions to consider adopting include tiered access systems for providing making data available to different categories of data users through different mechanisms; data sharing models such as privacy-aware contingency tables, data visualizations, and synthetic data, including such tools that provide formal privacy guarantees; and tools for transparency, notice, and consent, such as data asset registers and dynamic consent procedures.

* **Evaluation:** Open data managers should continually review and evaluate their data release decisions and data sharing mechanisms.

Open data managers should consider implementing procedures for reviewing and updating their data management and release processes over time. This will ensure flexibility and adaptability in response to advances in technology, evolving understanding of data privacy risks, new research questions and data analysis techniques, regulatory shifts, and changes in societal expectations of privacy. They should also consider adopting post-release review, accountability, and redress mechanisms to monitor and detect misuses of data and enable enforcement in response to privacy breaches. In addition, documenting and publishing assessments of the expected data uses, potential risks, and the privacy and security interventions implemented at each stage of the information lifecycle, as described above, could improve transparency and accountability by enabling members of the public to review agencies' data management and release decisions.

[47] See id.

[48] See id.

About the Authors

Alexandra Wood is a Fellow at BKCIS, a lawyer, and a member of the law and policy team contributing to the Privacy Tools for Sharing Research Data project at Harvard University.

David R. O’Brien is a Senior Researcher at BKCIS, and a lawyer. He leads research efforts at BKCIS related to privacy and cybersecurity, including the Berklett Cybersecurity project and the Privacy Tools for Sharing Research Data project.

Urs Gasser is the Executive Director of BKCIS and a Professor of Practice at Harvard Law School. He is a visiting professor at the University of St. Gallen (Switzerland) and at KEIO University (Japan), and he teaches at Fudan University School of Management (China).

For the past four years, BKCIS has been focused intensively on privacy and data sharing issues through its involvement with the Privacy Tools for Sharing Research Data project and related privacy initiatives. This research briefing summarizes, in large part, joint work with collaborators through the Privacy Tools project.[49] This paper and the other briefings included as part of the Networked Policy Series are generously supported by the Ford Foundation and the John D. and Catherine T. MacArthur Foundation. In addition, this material is based upon work supported by the National Science Foundation under Grant No. 1237235, the Berkeley Center for Law & Technology, and Microsoft Corporation.

[49] See Micah Altman, Alexandra Wood, David R. O’Brien, Salil Vadhan & Urs Gasser, Towards a Modern Approach to Privacy-Aware Government Data Releases, 30 Berkeley Tech. L.J. 1967 (2015); Salil Vadhan et al., Comments to the Department of Health and Human Services Re: Advance Notice of Proposed Rulemaking: Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, Docket No. HHS-OPHS-2011-0005 (Oct. 26, 2011).