City of Seattle
Open Data Risk Assessment

JANUARY 2018 – FINAL REPORT

FUTURE OF
PRIVACY
FORUM

Table of Contents

Executive Summary

The transparency goals of the open data movement serve important social, economic, and democratic functions in cities like Seattle. At the same time, some municipal datasets about the city and its citizens' activities carry inherent risks to individual privacy when shared publicly. In 2016, the City of Seattle declared in its Open Data Policy that the city's data would be "open by preference," except when doing so may affect individual privacy.[1] To ensure its Open Data Program effectively protects individuals, Seattle committed to performing an annual risk assessment and tasked the Future of Privacy Forum (FPF) with creating and deploying an initial privacy risk assessment methodology for open data.

This Report provides tools and guidance to the City of Seattle and other municipalities navigating the complex policy, operational, technical, organizational, and ethical standards that support privacy-protective open data programs. Although there is a growing body of research regarding open data privacy, open data managers and departmental data owners need to be able to employ a standardized methodology for assessing the privacy risks and benefits of particular datasets internally, without access to a bevy of expert statisticians, privacy lawyers, or philosophers. By optimizing its internal processes and procedures, developing and investing in advanced statistical disclosure control strategies, and following a flexible, risk-based assessment process, the City of Seattle – and other municipalities – can build mature open data programs that maximize the utility and openness of civic data while minimizing privacy risks to individuals and addressing community concerns about ethical challenges, fairness, and equity.

This Report first describes inherent privacy risks in an open data landscape, with an emphasis on potential harms related to re-identification, data quality, and fairness. To address these risks, the Report includes a Model Open Data Benefit-Risk Analysis ("Model Analysis"). The Model Analysis evaluates the types of data contained in a proposed open dataset, the potential benefits – and concomitant risks – of releasing the dataset publicly, and strategies for effective de-identification and risk mitigation. This holistic assessment guides city officials to determine whether to release the dataset openly, in a limited access environment, or to withhold it from publication (absent countervailing public policy considerations). The Report methodology builds on extensive work done in this field by experts at the National Institute of Standards and Technology, the University of Washington, the Berkman Klein Center for Internet & Society at Harvard University, and others,[2] and adapts existing frameworks to the unique challenges faced by cities as local governments, technological system integrators, and consumer facing service providers.[3]

[11] Exec. Order No. 2016-01 (Feb. 4, 2016), *available at* http://murray.seattle.gov/wp-content/uploads/2016/02/2.26-EO.pdf.

[2] *See infra* Appendix A for a full list of resources.

[3] *See* Kelsey Finch & Omer Tene, *The City as a Platform: Enhancing Privacy and Transparency in Smart Communities*, CAMBRIDGE HANDBOOK OF CONSUMER PRIVACY (forthcoming).

FPF published a draft report and proposed methodology for public comment in August, 2017. Following this period of public comment and input, FPF assessed the City of Seattle as a model municipality, considering the maturity of its Open Data Program across six domains:

1. Privacy leadership and management
2. Benefit-risk assessments
3. De-identification tools and strategies
4. Data quality
5. Data equity and fairness
6. Transparency and public engagement

In our analysis, we found that the Seattle Open Data Program has largely demonstrated that its procedures and processes to address privacy risks are fully documented and implemented, and cover nearly all relevant aspects of these six domains. Specifically:

- The City of Seattle is a national leader in privacy program management.
- The Seattle Open Data Program has developed and managed robust and innovative policies around data quality, public engagement, and transparency.
- The Seattle Open Data Program is working to enhance its policies and procedures for consistently assessing the benefits and risks of releasing particular datasets and for assessing and mitigating re-identification risks in open data.

Although most aspects of Seattle's programs are documented and implemented, some aspects are not as developed. This is unsurprising, given the novel challenges posed by the intersection of open government equities and privacy interests with emerging technologies and data analysis techniques.

The Report concludes by detailing concrete technical, operational, and organizational recommendations to enable the Seattle Open Data Program's approach to identify and address key privacy, ethical, and equity risks, in light of the city's current policies and practices. For example, we recommend that the City of Seattle and the Open Data Program:

- Document potential benefits and risks for each published dataset, both prospectively and retroactively for those that have not yet had a benefit-risk assessment conducted.
- Develop policies and procedures for conducting additional screening of datasets and elevating the review of risky or sensitive datasets to disclosure control experts or a disclosure review board when appropriate.
- Engage governmental decision-makers at the data collection stage with decision-makers at the data release stage (such as open data and public records staff), so that the full lifecycle of data collected by and for the city can be better understood, managed, and communicated to the public.

The City of Seattle is one of the most innovative cities in the country, with an engaged and civic-minded citizenry, active urban leadership, and a technologically sophisticated business community. By continuing to complement its growing Open Data Program with robust privacy protections and policies, the City of Seattle will be able to fulfill that program's goals, supporting civic innovation while protecting individual privacy.

Acknowledgments: We extend our thanks to the experts from the City of Seattle, Seattle Community Technical Advisory Board, University of Washington, Berkman Klein Center for Internet & Society at Harvard University, members of the FPF Smart Community Privacy Working Group, and others who provided their support and input in the development of this Report. We also thank all of those who provided public comments to this draft report for their thoughtful feedback and active participation in this process. We gratefully acknowledge the contributions of FPF Policy Interns Jeffrey Moore and Mia Little to this Report.

Special thanks to Jan Whittington, Meg Young, Ryan Calo, Mike Simon, Jesse Woo, and Peter Schmiedeskamp for their foundational scholarship on these issues at the University of Washington and to Michael Mattmiller, Jim Loter, David Doyle, Ginger Armbruster, Dena Solt, Diane Brady, and the many Open Data and Privacy Champions for their vision and dedication to making open data privacy a reality in the City of Seattle.

About FPF: Future of Privacy Forum is a nonprofit organization that serves as a catalyst for privacy leadership and scholarship, advancing principled data practices in support of emerging technologies. FPF brings together industry, academics, consumer advocates, and other thought leaders to explore the challenges posed by technological innovation and develop privacy protections, ethical norms, and workable business practices.

Background

In February 2016, City of Seattle Mayor Edward Murray issued an Executive Order calling for "all city data to be 'open by preference' – meaning city departments will make their data accessible to the public, after screening for privacy and security considerations."[4] The Executive Order "both sets the expectation that public data will be public and makes clear that [the city] has a responsibility to protect privacy."[5]

The City of Seattle Open Data Policy[6] directs the City of Seattle to perform an annual risk assessment of both the Open Data Program and the content available on its Open Data Portal. For this, the City of Seattle contracted the Future of Privacy Forum (FPF) to develop a methodology for conducting a risk assessment and to actively deploy the methodology. FPF reviewed a subset of high-risk agency datasets as well as a random sample of additional agency datasets, to evaluate privacy risks, including of re-identification, in case of release of individual datasets or multiple datasets.

From fall 2016 through summer 2017, FPF studied existing privacy and other risk assessment frameworks, created the Model Open Data Benefit-Risk Analysis, and assessed the inherent privacy risks in the municipal open data landscape for the City of Seattle as a model municipality. In doing so, FPF built on open frameworks, such as the National Institute of Standards and Technology (NIST) Special Publication 800-series. In addition to a review of available research and policy guidance related to open data privacy risk, FPF conducted interviews with privacy, open data, and disclosure control experts from around the world.

FPF also visited on-site to conduct interviews with Seattle IT and Open Data leadership, departmental Open Data and Privacy Champions, and local community advisors. These interviews included teams from the Seattle IT Department, Seattle Police Department, Seattle Department of Transportation, Planning and Development, Parks and Recreation, Civil Rights, Immigrant Affairs, and the Seattle Public Library.

FPF presented an early draft of the identified privacy risks and assessment methodology to the Seattle Community Technology Advisory Board (CTAB) for review and input in February 2017. An additional 45-day period for public comment on the report was offered from July through September 2017.

[4] Exec. Order No. 2016-01 (Feb. 4, 2016), *available at* http://murray.seattle.gov/wp-content/uploads/2016/02/2.26-EO.pdf.

[5] City of Seattle 2017 Open Data Plan, http://www.seattle.gov/Documents/Departments/SeattleIT/City%20of%20Seattle%202017%20Open%20Data%20Plan.pdf.

[6] City of Seattle, OD-1 V1.0, Open Data Policy (§ 5(k)) (2016), *available at* http://www.seattle.gov/Documents/Departments/SeattleGovPortals/CityServices/OpenDataPolicyV1.pdf.

Open Data Privacy Risks

Open and accessible public data can benefit individuals, companies, communities, and government by unleashing new social, economic, and civic innovations and improving government accountability and transparency. Tremendous benefits in healthcare, education, housing, transportation, criminal justice, and public safety are already being realized as richer and more timely datasets are made available to the public. Open data can unite the power of city and private sector abilities to improve community health and lifestyles, from bikeshare systems and commercial apps harnessing transit data to community advocates shining the light on ineffective or discriminatory practices through policing and criminal justice data.

In Seattle, for example, the Open Data Program seeks to:

- "Improve public understanding of City operations and other information concerning their communities,
- Generate economic opportunity for individuals and companies that benefit from the knowledge created by Open Data,
- Empower City employees to be more effective, better coordinated internally, and able to identify opportunities to better serve the public, and
- Encourage the development of innovative technology solutions that improve quality of life."[7]

However, open data can also pose substantial risks to the privacy of individuals whose information is collected and shared by the city. Inadequate privacy protections for open data can lead to significant financial, physical, reputational, organizational, and societal harms. For example, citizens might object to the release of their home address in connection with a crime tracking dataset, allowing nosy neighbors or prowlers to identify them or learn sensitive information about their lives. In other cases, poor quality data could lead to an individual being wrongly identified in a DUI database, causing lasting harm. And people who baulk at data brokers, advertisers, or insurance agents profiting off of or profiling their purchasing or financial habits from public datasets might cease participating in public services.

Cities must be vigilant and resourceful to deter and defend against these privacy risks, no matter how they arise. In this section, we describe the core privacy risks facing municipal open data programs: re-identification, biased or inaccurate data, and loss of public trust.

Re-identification

One of the principal and unavoidable risks of opening government datasets to the public is the possibility that the data might reveal sensitive information about a specific individual. In cases where open datasets are not adequately vetted, personally identifiable information (PII) may be published

[7] *Open Data Program*, City of Seattle, https://data.seattle.gov/stories/s/urux-ir64 (last visited July 6, 2017).

inadvertently. Even when a dataset has been scrubbed of names and other potentially identifying traits and rendered “de-identified,” there is a chance that someone (referred to in professional literature as an “adversary”) might be able to deduce that some of the data relates to a specific individual. This can be a professional skilled in re-identifying individuals from seemingly “anonymous” information; a commercial information reseller with access to millions of other data points; or an insider like a friend, coworker, or neighbor (or social media follower) who knows other personal information about an individual. If municipal employee salaries are published to an open dataset, for example, a family member who knows a particular individual’s job title may suddenly be able to easily learn how much money their relative makes.

Re-identifying a person in this way not only exposes data about the individual that would otherwise not be available to the public, but could potentially carry embarrassing, damaging, or life-threatening implications. For example, in Dallas, the names of six people who complained of sexual assault were published online by the police department. While the Dallas Police Department does not, of course, intentionally publish such sensitive information, its case classification scheme and overlapping information across datasets combined in such a way that the six injured parties could be singled out and identified when they should not have been.[8] Other re-identification attacks may reveal an individual’s home address or place of work, exposing them to increased risk of burglary, property crime, or assault.[9]

Recent advances in smart city technologies, re-identification science, data marketplaces, and big data analytics have enhanced re-identification risks, and thus increased the overall privacy risk in open datasets. As open data programs mature and shift from merely providing historic data and statistics to more granular, searchable, accessible, and comprehensive “microdata” about citizens and their activities, the risk of re-identification rises even further. Databases of calls to emergency services, civil complaints about building codes and restaurants, and even civil rights violations will potentially become available for anyone in the world to explore. The ease at which adversaries (including professional researchers, commercial organizations and data brokers, other government and law enforcement agencies, civic hackers, and individual members of the general public) could download, re-sort, and recombine these datasets carries an obvious risk for the leakage of sensitive data.

Open data programs are not only challenged by sophisticated adversaries combining multiple databases to reveal sensitive attributes about individuals. Opening administrative datasets that appear more routine or mundane (and therefore fail to raise the same privacy red flags) can also leave individuals exposed. In 2017, for example, a parent who was examining expenditure files on the Chicago Public School’s website discovered that deep within the tens of thousands of rows of vendor payment data were some 4,500 files that identified students with Individualized Educational Programs – revealing in plain text the students’ names, identification numbers, the type of special education services that were

[8] *See* Andrea Peterson, *Why the names of six people who complained of sexual assault were published online by Dallas police,* WASH. POST, Apr. 21 2016, https://www.washingtonpost.com/news/the-switch/wp/2016/04/29/why-the-names-of-six-people-who-complained-of-sexual-assault-were-published-online-by-dallas-police/.

[9] *See* SIMSON L. GARFINKEL, DE-IDENTIFYING PERSONAL INFORMATION NISTIR 8053 (NIST Oct. 2015), http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf.

being provided for them, how much those services cost, the names of therapists, and how often students met with the specialists.[10]

One of the unavoidable challenges of open data is that once information has been published publicly, it likely can never be retracted. Unfortunately, data de-identification is a moving target – data that could not be linked to an individual when it was released, could become identifiable over time. For example, if sometime in the future another dataset is published that links one record to another or if a new technique becomes available to match information across multiple datasets, the risk of re-identifying an individual in the original open dataset may increase significantly. While it is difficult to predict when such future data may become available, cutting-edge research into more dynamic de-identification techniques is underway by disclosure control experts around the world.

Re-identification also harms municipalities: when data published on an open data portal becomes re-identified and harms an individual, public trust in the city and in open data could be seriously eroded. Citizens may stop providing data, or provide false data, if they believe that it might be exposed in the future. If the data were subject to regulatory or confidentiality provisions, moreover, such disclosures could lead to new compliance costs or lawsuits. For example, in 2012, Philadelphia's Department of Licenses & Inspections published gun permit appeals as part of its open data initiative. These permits included a free text field where applicants explained why they needed the permit. Some individuals wrote they carried large sums of cash at night. As a consequence of disclosing this information, the City was ultimately charged $1.4 million as part of a class-action lawsuit. One of the lawyers behind the suit stated that the information released was a "road map for criminals."[11]

Re-identification can cause harms to individuals, organizations, government agencies, and society as a whole. Even *false* claims of re-identification can cause significant damage, leaving individuals uncertain whether their information is exposed and susceptible to lost opportunities or mistaken decisions based on data wrongly attributed to them.

Data Quality and Equity

Multiple stakeholders rely on the accuracy of information in public datasets: citizens, companies, community organizations, and other governmental entities. In some circumstances, inaccurate, incomplete, or biased open data may have little impact – for example, a list of sold city fleet vehicles may accidentally record the wrong make and model for a vehicle or two. In other circumstances, however, the consequences can be more lasting, leading to poor or inefficient decision-making, unethical or illegal data uses, or discriminatory outcomes. Publishing the wrong person's information to

[10] *See* Lauren Fitzpatrick, *CPS privacy breach bared confidential student information,* CHI. SUN-TIMES (Feb. 2, 2017), http://chicago.suntimes.com/news/cps-privacy-breach-bared-confidential-student-information/.

[11] *See* Vince Lattanzio, *Philly paying $1.4 million after posting confidential gun permit information online,* NBC PHILADELPHIA, July 22, 2014, http://www.nbcphiladelphia.com/news/local/Philly-Paying-14M-After-Posting-Confidential-Gun-Permit-Information-Online-268147322.html.

an open dataset of DUI arrests, for example, could adversely affect that person's employment, credit, and insurance prospects for years to come.

Personal data that has been made public without legal conditions may be consumed and repurposed by any number of potential actors, including identity thieves, commercial information resellers (and ultimately their clients, including potential employers, insurers, creditors, and others), companies, friends and family, nosy neighbors, stalkers, law enforcement and other government entities, and others. Some commercial "mugshot" or arrest record databases, for example, profit by gathering sensitive personal information via public records, publishing the data to private sites, and then charging individuals a fee to have them removed.[12] The lack of control over downstream uses of open data is a significant point of concern among a variety of open data stakeholders, including civic hackers, legal advocates, and industry representatives.[13]

Over the last few years, organizations increasingly rely on data to automate their decision-making in a wide variety of situations, including everything from traffic management to personalized advertising to insurance rate setting. But particularly in "smart" systems that use algorithmic decision-making and machine learning, bad data can lead to bad policies. For example, both predictive policing and criminal sentencing have repeatedly demonstrated racial bias in both the inputs (historic arrest and recidivism data) and their outputs, leading to new forms of institutional racial profiling and discrimination.[14]

In fact, even individuals who are not directly represented in an open dataset may nevertheless be impacted by inaccuracies and biases in the dataset or analysis performed on it. [15] For example, according to the City of Seattle, "residents of zip codes listed as having high rates of households below the poverty level; property owners in neighborhoods where crime rates are higher than average; [and] students at schools that are underperforming" may all be adversely effected by conclusions drawn from such datasets, especially if drawn from "low-quality data."[16] While municipal open data programs often categorize data by rough accuracy measures for the purposes of prioritization,[17] this sort of quick data sorting is not a substitute for the in-depth data quality and privacy assessments that are required prior to publication. These sorts of inferential disclosures may result in group harms that have not been traditionally viewed as privacy concerns, and may thus not be well addressed by existing municipal privacy policies and practices.

[12] Damian Ortellado, *The perils of personally identifiable pre-conviction data*, SUNLIGHT FOUNDATION (Feb. 1, 2016, 3:48 PM), https://sunlightfoundation.com/2016/02/01/the-perils-of-personally-identifiable-pre-conviction-data/.

[13] Jan Whittington et al., *Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government*, 30 BERKELEY TECH. L.J. 1899, 1913-14 (2015).

[14] *See generally* Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[15] *See* SIMSON L. GARFINKEL, *supra* note 9.

[16]*See* CITY OF SEATTLE, OPEN DATA PLAYBOOK V. 1.0, http://www.seattle.gov/Documents/Departments/SeattleGovPortals/CityServices/OpenDataPlaybook_Published_2016.08.pdf.

[17] *Id.*

Moreover, an unfair distribution of data benefits and data risks across a community may reinforce societal biases, disguise prejudiced decision-making, and block equal opportunities for marginalized or vulnerable populations. Some open data stakeholders have raised concerns that, particularly when commercialized, public municipal data may be used to "lower property values, redline insurance, et cetera, in neighborhoods with high crime rates rather than addressing those issues."[18] If data represented on the open data portal is disproportionately collected from certain populations over others; is used against certain populations over others; or if the data exposes vulnerable populations to higher privacy risks or at a higher rate than others, it may be inequitable. For example, given that minority and vulnerable populations, including immigrant communities, tend to be over-surveilled in comparison to majority populations, particularly in the context of law enforcement and social services, they may be disproportionately represented in open datasets, creating fertile grounds for inaccuracies and biases in decision-making or even just reporting of data. Governments must constantly strive to serve all their citizens fairly and equitably, however difficult it may be to strike the balance of equities.

Public Trust

Open data programs cannot succeed in their social, economic, and democratic missions without public trust. When individuals feel their privacy is violated by a particular dataset being published or that community expectations of privacy were disregarded, they may hold the open data program accountable. This can result not only in a loss of trust in the open data program, but also undermine the entire city government's ability to act as a responsible data steward. [19] Civic engagement and communication, paired with demonstrable responsible data practices, can earn the public's trust in open data. But if the public's trust in a government as a responsible data steward is damaged, individuals may become unwilling to support and participate in important civic activities and research.[20] It can also lead to the public providing false data in certain circumstances out of a fear their real information would be compromised.

Just as in the event of a data breach, individuals who believe that their personal data may have been exposed to the world can feel uncertainty and anxiety about the loss of informational control and potential long-term ramifications such as identity theft. When personally identifiable information is published to an open data portal or a re-identification attack appears successful, individuals often have little recourse. Municipal leaders must be aware that deciding what data they may *release* about individuals is inextricable from what data they *collect* about individuals. Failing to address privacy throughout the entire data lifecycle – including collection, use, sharing, retention, disposal – will impede public trust in data-driven municipal programs. In the open data context, in particular, it should be noted that once data has been made public they may be re-used and re-shared by others long after the city has disposed of them internally. For example, cities should be cautious about collecting information

[18] Whittington et al., *supra* note 13, at 1919.

[19] *See Ben Green et al*., OPEN DATA PRIVACY (2017), https://dash.harvard.edu/handle/1/30340010; Whittington et al., *supra* note 13, at 1914.

[20] SEAN A. MUNSON ET AL., ATTITUDES TOWARD ONLINE AVAILABILITY OF US PUBLIC RECORDS (2011), https://pdfs.semanticscholar.org/fa4b/e73719e5047fb97f21eef25bbe26984abbf0.pdf.

that would harm individuals if it were one day shared via the open data portal, disclosed via a public records request, or exposed via a data breach.[21]

Finally, cities must be aware that *how* data is collected and used is as important as how it is released for ensuring public trust in open data programs. Cities must communicate clearly with individuals about how and when their data can find its way to an open data portal. Vague privacy notices and a lack of an opportunity to opt in or out of data collection may shock or surprise some people, even if that information is in pseudonymized or aggregate form. And if data is used for a purpose other than the reason the collection occurred without citizens' consent to repurpose, significant privacy concerns are raised, as well as ethical and technical questions. It is possible that an individual never would have consented to the data collection if they it would ultimately be released publicly through the open data portal. Where an individual's privacy – or trust – has been violated by a government data initiative, it may be impossible to restore.

*

The transparency goals of municipal open data programs are critical to the improvement of civic life and institutions in the modern city, and rely on the release of microdata about the city and its citizens' activities. And yet people who provide personal information to their governments must be able to trust that their privacy will be protected. If individuals find their personal information exposed, or their neighborhoods singled out or discriminated against, or their data collected for one purpose and used for another, this can undermine public trust in the city as a whole and slow or even reverse the momentum of the open data program. On the other hand, where cities engage the public and communicate the benefits of the open data program while clearly addressing any shortcomings, they may build public trust. Responsible privacy practices and effective communication provide the foundation for successful, trustworthy, and innovative open data programs.

[21] *See* Liz Robbins, *New York City ID Holders Aren't a Threat, N.Y.P.D. Official Says in Court*, N.Y. TIMES (Jan. 5 2017), https://www.nytimes.com/2017/01/05/nyregion/new-york-id-program-immigrants.html?action=click&contentCollection=N.Y.%20%2F%20Region&module=RelatedCoverage&region=EndOfArticle&pgtype=article; Liz Robbins, *New York Can Destroy Documents, Judge Rules in Municipal ID Case*, N.Y. TIMES (Apr. 7, 2017), https://www.nytimes.com/2017/04/07/nyregion/new-york-can-destroy-documents-judge-rules-in-municipal-id-case.html.

Model Open Data Benefit-Risk Analysis

In the open data context, considering the risks of the dataset is merely one part of a balanced value equation; decision-makers must also take account of the project's benefits in order to make a final determination about whether to proceed with publishing the dataset openly.[22] For the purposes of this report, FPF developed this Model Analysis, which is based on risk assessment and de-identification frameworks developed by the National Institute of Standards and Technology and also builds on parallel efforts by researchers at the University of Washington, the Berkman Klein Center, and the City of San Francisco to develop robust risk-based frameworks for government data releases.[23] This Model Analysis provides a structure for vetting potential open datasets in five steps:

Step 1: Evaluate the Information the Dataset Contains. This step includes identifying whether there are direct or indirect identifiers, sensitive attributes, or information that is difficult to de-identify present in the dataset; assessing how linkable the information might be to other datasets; and considering the context in which the data was obtained.

Step 2: Evaluate the Benefits Associated with Releasing the Dataset. This step considers the potential benefits and users of the dataset, and assesses the magnitude of the potential benefits against the likelihood of their occurring.

Step 3: Evaluate the Risks Associated with Releasing the Dataset. This step considers the potential privacy risks and negative users of the dataset, and assesses the magnitude of the potential risks against the likelihood of their occurring.

Step 4: Weigh the Benefits against the Risks of Releasing the Dataset. This step combines the overall scores from steps 2 and 3 to determine an appropriate method for releasing (or not releasing) the dataset. Recommendations include releasing as open data, in a limited access environment, or not publishing at the current time. This section also overviews common methods for reducing re-identification risk in terms of their privacy-protective, utility, and operational impacts.

Step 5: Evaluate Countervailing Factors. This step provides a final opportunity to document any countervailing factors that might justify releasing a dataset openly regardless of its privacy risk, such as when there is a compelling public interest in the information.

See Appendix C for the full Model Analysis.

[22] *See infra* Appendix C.

[23] *See* Micah Altman et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 BERKELEY TECH. L.J. 1968 (2015); Jan Whittington et al., *Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government*, 30 BERKELEY TECH. L.J. 1899 (2015); Ben Green et al., *Open Data Privacy*, BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY AT HARVARD (2017); DATASF, https://datasf.org/opendata/.

The City of Seattle as a Model Municipality

Given the risks described above, FPF developed and applied the following assessment to evaluate the City of Seattle as a model municipality based on its organizational structure and data handling practices related to open data. The assessment is grounded in public documentation and interviews with privacy, open data, and disclosure control experts and with Seattle IT, Open Data, and Privacy Leadership, departmental Open Data and Privacy Champions, and local community advisors including the Community Technology Advisory Board.

Our scoring of the City of Seattle's practices in each of the following domains is based on the AICPA/CICA Privacy Maturity Model (PMM) levels:[24]

- ***Undeveloped*** – procedures or processes are absent, or are unpredictable and reactive.
- ***Ad hoc*** – procedures or processes are generally informal, incomplete, and inconsistently applied.
- ***Repeatable*** – procedures or processes exist; however, they are not fully documented and do not cover all relevant aspects.
- ***Defined*** – procedures and processes are fully documented and implemented, and cover all relevant aspects.
- ***Managed*** – reviews are conducted to assess the effectiveness of the controls in place.
- ***Optimized*** – regular review and feedback are used to ensure continuous improvement towards optimization of the given process.

A key principal of the PMM approach is the recognition that "each organization's personal information privacy practices may be at various levels, whether due to legislative requirements, corporate policies or the status of the organization's privacy initiatives. It was also recognized that based on an organization's approach to risk, not all privacy initiatives would need to reach the highest level on the maturity model."[25]

Given the relative youth of municipal open data programs in the U.S.,[26] it is to be expected that fully mature privacy practices may take years to emerge. The privacy profession itself is relatively young,[27]

[24] *See generally AICPA/CICA Privacy Maturity Model*, CHARTERED ACCOUNTANTS OF CANADA (Mar. 2011) (https://www.kscpa.org/writable/files/AICPADocuments/10-229_aicpa_cica_privacy_maturity_model_finalebook.pdf).

[25] *See id.*

[26] For example, the City of Seattle's open data program launched in 2010, and the Executive Order directing all City data to be "open by preference" was signed in 2016. CITY OF SEATTLE, OPEN DATA PROGRAM 2016 ANNUAL REPORT, https://www.seattle.gov/Documents/Departments/SeattleIT/Open%20Data%20Program%202016%20Annual%20Report.pdf.

[27] *See* KENNETH BAMBERGER & DEIRDRE MULLIGAN, PRIVACY ON THE BOOKS AND ON THE GROUND (2015) (discussing the emergence of Chief Privacy Officers in the 1990s and 2000s).

and the technical, legal, and organizational tools necessary to address the full panoply of open data privacy risks are still evolving. For example, while the science supporting the de-identification of personal data is advancing towards more mathematically grounded measures of privacy (e.g., differential privacy), for now such techniques remain difficult and costly to implement at scale.[28] Similarly, while stakeholders in both the public and private sectors recognize the possibility that new data mining and analytics techniques may lead to inequitable or discriminatory uses of personal data, the tools to prevent and remedy these unfair outcomes are still emerging.[29] We fully expect that municipal open data programs will play a role in supporting the development and implementation of these emerging tools and safeguards in the years to come.

FPF evaluated the City of Seattle's current Open Data Program by assessing PMM levels across the following six domains:

- Privacy leadership and program management
- Benefit-risk assessment
- De-identification tools and strategies
- Data quality
- Equity and fairness
- Transparency and public engagement

[28] *See* SIMSON L. GARFINKEL, NISTIR 8053: DE-IDENTIFYING PERSONAL INFORMATION (NIST Oct. 2015), http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf.

[29] *See Scholarship,* Fairness, Accountability, and Transparency in Machine Learning, http://www.fatml.org/resources/relevant-scholarship (last visited 7/17/17).

Privacy leadership and program management

- Does the municipality employ a comprehensive, strategic, agency-wide privacy program regarding its open data initiatives?
- Has the municipality designated a privacy governance leader for open data?
- Is the open data program guided by core privacy principles and policies?
- Does the open data workforce receive effective privacy training and education?
- Are the municipality's open data privacy policies and procedures updated in light of ongoing monitoring and periodic assessments?

Seattle privacy maturity score: Optimized

The City of Seattle is a national leader in municipal privacy governance. Under the guidance of the Seattle IT department, agencies citywide have demonstrated commitment to privacy-protective data practices. The Open Data and Privacy Programs, in particular, have undergone significant operational and cultural shifts to more effectively enshrine privacy protections in a short amount of time.

The Open Data Program Manager has developed and deployed a comprehensive, strategic, and citywide plan for ensuring that city departments making their data accessible to the public consistently screen for privacy, security, and quality considerations. This work is guided by the city's Privacy Principles, adopted by the City Counsel in February 2015,[30] as well as the Open Data Policy created by Executive Order 2016-01 in February 2016,[31] and supported by annual progress reports evaluating existing policies and procedures.[32] Consistent with the Open Data Policy, the city engaged external privacy experts at FPF to complete a privacy risk assessment of the Open Data Program to evaluate the effectiveness of the controls in place and to ensure continuous improvement. Even prior to the establishment of the Open Data Policy, the City of Seattle worked in partnership with University of Washington experts to analyze privacy protections for municipal data release.[33]

The city's Open Data and Privacy Programs, both situated within the Seattle IT department, work closely to ensure that data is published in compliance with the city's Privacy Principles. While the city's Chief Privacy Officer and permanent privacy staff are responsible for the privacy governance and review of open datasets, the Open Data Program manager and designated departmental "Open Data Champions" also have privacy governance responsibilities.[34] The Open Data workforce has received multiple privacy-specific trainings, including "Data Camp," a multiday workshop series designed to educate Open Data

[30] CITY OF SEATTLE, PRIVACY PRINCIPLES, https://www.seattle.gov/Documents/Departments/InformationTechnology/City-of-Seattle-Privacy-Principles-FINAL.pdf.

[31] CITY OF SEATTLE, OPEN DATA POLICY, http://www.seattle.gov/Documents/Departments/SeattleGovPortals/CityServices/OpenDataPolicyV1.pdf

[32] CITY OF SEATTLE, OPEN DATA PROGRAM 2016 ANNUAL REPORT, https://www.seattle.gov/Documents/Departments/SeattleIT/Open%20Data%20Program%202016%20Annual%20Report.pdf.

[33] Jan Whittington et al., *Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government*, 30 BERKELEY TECH. L.J. 1899 (2015), http://btlj.org/data/articles2015/vol30/30_3/1899-1966%20Whittington.pdf.

[34] Several departmental Open Data Champions are also the departmental Privacy Champion.

Champions about issues such as data quality, data privacy, data equity, and public disclosure.[35] With even non-technical employees within the City of Seattle receiving basic privacy training, data are more likely to be protected throughout their full lifecycle (collection, use, release, disposal).

Benefit-risk assessment

- Does the open data program conduct a benefit-risk assessment to manage privacy risk in each dataset considered for publication?
- Are datasets assessed based on the identifiability, sensitivity, and utility of the data prior to release?
- Are inventories of published personally identifiable information (PII) maintained?
- Are benefit-risk assessments documented and regularly reviewed?
- Does the open data program have a mechanism in place to trigger re-assessment of a published dataset in light of new facts?
- Does the open data program have an ability to elevate review of risky or sensitive datasets to disclosure control experts or a disclosure review board?

Seattle privacy maturity score: Repeatable

Seattle's processes and procedures for reviewing the benefits and privacy risks of prospective open datasets are fully documented and implemented, however these efforts are incomplete and not as robust as those in a fully mature program. Datasets do undergo documented benefit-risk assessments prior to publication, however these assessments are not regularly reviewed after publication. Nor do formal procedures appear to exist that would trigger re-assessment of previously published datasets (such as if a new dataset or re-identification technique were to be created that significantly raised the risk of re-identification for the existing data). Inventories of PII published to the Seattle Open Data portal are not centrally maintained, though they could help Open Data Champions and privacy reviewers more confidently assess whether a prospective dataset contains the same fields as the "foreign key" to another dataset (thus potentially raise the risk of re-identification).

While prospective datasets undergo a tiered privacy assessment process that leads open data submitters through progressively more intensive review processes according to the identifiability and sensitivity of the data, a full accounting of the potential benefits and risks of a particular dataset is reserved only for the most stringent review. Although this means that the datasets with the highest potential privacy impact receive the greatest review, specifically documenting the expected benefits and risks of *every* dataset at the time of their publication can serve an important accountability function. Furthermore, datasets entered into the database prior to the implementation of these new processes in early 2016 have not undergone such review.

[35] *See* CITY OF SEATTLE, OPEN DATA PROGRAM 2016 ANNUAL REPORT (https://www.seattle.gov/Documents/Departments/SeattleIT/Open%20Data%20Program%202016%20Annual%20Report.pdf).

In light of the initial draft of this report, however, the Seattle Open Data Program is already considering how to more efficiently and programmatically complete more comprehensive benefit-risk assessments like the model included in Appendix C.

De-identification tools and strategies

- Does the open data program utilize technical, legal, and administrative safeguards to reduce re-identification risk?
- Does the open data program have access to disclosure control experts to evaluate re-identification risk?
- Does the open data program have access to appropriate tools to de-identify unstructured or dynamic data types? (e.g., geographic, video, audio, free text, real time sensor data)
- Does the open data program have policies and procedures for evaluating re-identification risk across databases? (e.g., risk created by intersection of multiple municipal databases; county, state, or federal open databases; commercial databases)
- Does the open data program evaluate privacy risk in light of relevant public records laws?

Seattle privacy maturity score: Repeatable

Although the Seattle Open Data Program utilizes a variety of safeguards to reduce re-identification risk, these procedures do not currently cover all relevant aspects of a mature disclosure control program. While Seattle's current de-identification controls can address many re-identification risks (and in some cases, handle nontraditional data types),[36] the unavailability of more sophisticated statistical, technical, and administrative tools limits the Open Data Program's ability to mitigate the full range of re-identification risks.

Recognizing the potential risk of re-identification from the 'mosaic effect,'[37] the privacy and Open Data teams conduct evaluations of re-identification risk across databases as part of the privacy review process. Nevertheless, although the Open Data Program collaborates with leading academic institutions like the University of Washington and the Berkman Klein Center, it does not yet have reliable access to statistical disclosure control experts or specialized de-identification tools to evaluate and mitigate re-identification risk across multiple datasets or in a variety of formats. As noted above, the tools to adequately address these risks may not yet be commercially available or implementable at scale; however, the City of Seattle's previous partnerships with privacy research centers may help pave a path forward for future developments in municipal de-identification strategies.

Nevertheless, the Seattle Open Data Program will need to grapple with developing policies and procedures for evaluating re-identification risk that can be applied prospectively and retroactively.

[36] In response to public records requests, for example, the Seattle Police Department has worked to develop tools to de-identify video and image data from body-worn cameras.

[37] *See The Mosaic Effect*, Office of the Assistant Secretary for Planning and Evaluation, HHS (Sept. 9, 2014), https://aspe.hhs.gov/report/minimizing-disclosure-risk-hhs-open-data-initiatives/c-mosaic-effect.

Currently, datasets uploaded to the Seattle Open Data Portal prior to the development of the current policies show inconsistent applications of basic de-identification techniques. For example, while the Seattle Real Time Police 911 Calls dataset generalizes addresses to the hundreds block, the Seattle Real Time Fire 911 Calls dataset reports precise addresses. The inconsistent treatment of location data and other potentially identifiable fields in legacy data raises the potential risk of re-identification for all datasets on the Open Data portal.

While these legacy data may have already introduced re-identification risks, the city's Open Data Program does not currently have procedures to further mitigate (although not remove) those risks, such as triggering re-evaluation of particular datasets' re-identification risk in light of changing circumstances, removing datasets if or when re-identification risks rise too high, or utilizing legal or administrative controls to control access to more sensitive datasets. Ultimately, because there is currently no way to know how or by whom Seattle's existing open data have been consumed, even removing the legacy datasets may not curtail their impact on future re-identification risks for other datasets. It is for this reason that legal and administrative controls (such as data enclaves, tiered access models, contractual safeguards, or use and download restrictions) are important complements to technical de-identification tools.

Finally, given the breadth of the State of Washington's current Public Records Law, the Seattle Open Data Program is considerably constrained in its efforts reduce re-identification risks. Staff responsible for evaluating privacy risk under open data and public records requests in various city departments are often not in close communication with each other, and there is not a formal process for considering the impact of each program on the other. Because the public records law mandates the disclosure of even personally identifiable information in many circumstances of legitimate public interest, open data programs within the state must be especially cautious about releasing de-identified records that may be 'unlocked' or re-identified by information subject to public records requests.

Data quality

- Does the municipality employ policies and procedures for the open data program to ensure that personally identifiable information is accurate, complete, and current?
- Does the open data program check for, and correct as appropriate, inaccurate or outdated personally identifiable information?
- Are there procedures or mechanisms for individuals to submit correction requests for potentially incorrect personal data posted on the open data program?

Seattle privacy maturity score: Managed

For over a year, the Seattle Open Data quality review process has been fully documented and implemented. Open dataset submissions are vetted for fidelity, completeness, consistency, currency, and credibility/validity and scored consistently prior to being approved for publication. The review considers the quality of both the dataset's content and metadata, reducing the likelihood that data will be misinterpreted at a later date.

The policies and procedures include suggested testing methods and exemplars to aid Open Data Champions in making consistent determinations about quality. The data quality review process also allows for previously approved mitigation strategies to be deployed to address minor inaccuracies within a reasonable amount of time after a dataset is published.

The Open Data Program does not actively search for inaccurate or outdated PII on the open data portal currently, however the Socrata system underlying Data.Seattle.Gov could allow for updates and corrections in the future. While there are no specific mechanisms for individuals to submit correction requests, individuals with concerns may easily contact the dataset owner via the open data portal.

Equity and fairness

- Were the conditions under which the data was collected fair? (e.g., were citizens aware that the data would be published on the open data portal? Did individuals have an opportunity to opt out of data collection? If data was acquired from a third party, were terms and conditions observed in the collection, use, maintenance, and sharing of the data?)
- Does the open data program assess the representativeness of the data? (e.g. whether underserved or vulnerable populations are appropriately represented in the data, or whether underserved or vulnerable populations' interests are taken into account when determining what data to publish).
- Are any procedures and mechanisms in place for people to submit complaints about the use of data or about the publication process generally, as well as procedures for responding to those complaints?

Seattle privacy maturity score: Ad hoc

While the Seattle Open Data Program does not have specific policies or procedures for assessing the representativeness of datasets on the open data portal, the city's Race and Social Justice Initiative (RSJI) is a mature, active, funded program dedicated to eliminating racial disparities and achieving racial equality in Seattle.[38] The Open Data Program has also committed to supporting the RSJI in 2017 by "releasing open datasets that help with promoting positive RSJI outcomes."[39]

The City of Seattle Privacy Statement makes clear that some data collected by the city may be made public through public records requests or the open data portal, and the city's public engagement and transparency efforts are helping educate the general public about what open data is and how it is created. The privacy assessment process triggers further review any data collected by particular

[38] *Race and Social Justice Initiative*, SEATTLE.GOV, https://www.seattle.gov/rsji (last visited 7/17/17).
[39] CITY OF SEATTLE, CITY OF SEATTLE 2017 OPEN DATA PLAN, http://www.seattle.gov/Documents/Departments/SeattleIT/City%20of%20Seattle%202017%20Open%20Data%20Plan.pdf.

surveillance technologies (such as public cameras), data collected under regulatory regimes, or data that may lead to public backlash if published.

Beyond the Privacy Statement, however, there appear to be few coordinated efforts to provide specific notices to individuals at the time of data collection about the possibility of their data being released publicly. The hiring of a dedicated Smart Cities Coordinator in August 2017 may provide additional capacity, particularly if or when the city's deployment of smart city technologies and sensors feed back into the Open Data Portal. The Open Data Program does offer a variety of communication channels for individuals to express complaints on social media, the open data portal, the City of Seattle website, and at community meetings and events, as well as formal procedures for responding to them.

Transparency and public engagement

- Does the open data program engage and educate the public about the benefits of open data?
- Does the open data program engage and educate the public about the privacy risks of open data?
- Does the open data program provide opportunities for public input and feedback about the portal, the data available, and privacy, utility, or other concerns?
- Does the open data program engage with the public when developing of open data privacy protections?
- Does the open data program consider the public interest in determining what datasets to publish?
- Does the open data program communicate with the public about why some datasets may include PII?

Seattle privacy maturity score: Managed

The Open Data Program includes a significant amount of community outreach, including coordination with the Civic Technology program and the Seattle Community Technology Advisory Board (CTAB). In 2016, the Open Data and Civic Technology Programs supported approximately 20 public events to engage and educate the public about the benefits of open data, including hackathons, presentations to community groups, brown bag lunches, and community design workshops. Many of these events were co-hosted by local community groups, businesses, and academic institutions. Video recordings of the city's Data Camp workshop, which included training on data quality, data privacy, data equity, and public disclosure, were also made public via the Seattle Channel.

The Open Data team relies on emails from citizens for suggesting datasets, noting problems with existing datasets, or other program management issues. The Open Data team also actively engages on social media, including promoting specific discussions and presentations about privacy and open data. Communications to the public about data on the open data portal, however, are largely captured by either the dataset's metadata (which seeks to provide context to the dataset as a whole as well as its individual data fields) or through the city's Privacy Statement. Beyond the Privacy Statement, there are few if any efforts to provide specific notices to individuals at the time of data collection about the possibility of their data being released as open data, however.

In developing privacy protections for open data, Seattle engages with a wide variety of stakeholders, including local privacy academics, civic technologists, and privacy activists, community groups, and external stakeholders like FPF. Correspondingly, as part of the city's commitment to transparency and openness, this draft report was presented to CTAB for input and response during its development, and will be made available for public comment prior to publication.

The Open Data team prioritizes datasets for publication based on the public interest, taking into account a variety of public stakeholders. The Open Data Playbook specifically contemplates the impact of open data on: people and institutions represented in the data, those who might be impacted by the release of data or analysis conducted on it, people and institutions who will use the raw data, and anyone who reads or uses the information. In practice, departmental Open Data Champions often consider the frequency of public records requests as a prime indicator of public interest in the information.

Overall Seattle Open Data Program privacy maturity score:
Defined

Considered holistically, the City of Seattle's Open Data Program has largely demonstrated that its procedures and processes to address privacy risk are fully documented and implemented, and cover nearly all relevant aspects of these six domains. The City of Seattle is a national leader in privacy program management, and has robust and innovative policies around data quality and public engagement and transparency. While the city's Open Data Program appears less mature in other technical and policy domains, such as consistently applying benefit-risk analyses, deploying more sophisticated de-identification tools, and engaging in data fairness reviews, Seattle appears to be ahead of the curve in comparison to other municipal data programs today, which have also lacked the technical tools or capacity to fully address these issues.

Given the short timespan in which the Seattle Open Data Program has gained this level of privacy sophistication, the strength of its organizational foundation, and the emergence of new scholarship and tools to address de-identification and data fairness, we think it is likely that Seattle's Open Data Program will continue to mature. Below we provide specific recommendations to the city for advancing its privacy protections to the next level.

Recommendations and Conclusion

As the Seattle Open Data Program evolves and matures, it must continue developing the specialized resources and tools to address the privacy risks inherent in open data. We fully expect that years of innovation, investment, and community discussion around evolving privacy best practices will be required for fully mature municipal open data programs to emerge in the United States or elsewhere. Where municipalities are uncertain about their capacity to protect their constituents' privacy in open datasets, we urge them to err on the side of caution until sufficient protections are available to them.

The Seattle Open Data Program will be building on the strong foundation described in the section above, but there are always steps that can be taken to improve the depth and breadth of privacy protections. The following recommendations are intended to support this growth and advance the City of Seattle's leadership in open data privacy:

- To optimize privacy leadership and program management throughout the city, the City of Seattle should, as appropriate:
    - Continue to deepen workforce privacy training and education efforts throughout the city.
    - Continue to codify data handling policies and procedures to ensure continuity and consistency over time.
    - Continue to invest in the Open Data and Privacy Champions Programs to build experience and expertise internally (such as providing incentives, e.g., spot awards, increased compensation or benefits, or appointing separate staff to each role and engaging both in reviewing potential open datasets).
    - Engage governmental decision-makers at the data collection stage with decision-makers at the data release stage (such as open data and public records staff), so that the full lifecycle of data collected by and for the city can be better understood, managed, and communicated to the public.
    - Regularly review and take feedback on the Open Data Program's privacy practices to ensure continuous improvement.

- To manage and optimize its open data benefit-risk assessment process, the City of Seattle should, as appropriate:
    - Document potential benefits and risks for every published dataset, both prospectively and retroactively for those that have not yet had a benefit-risk assessment conducted.
    - Develop mechanisms to trigger re-assessment of published datasets in light of new facts.
    - Review benefit-risk assessments on a regular basis, and determine how to respond in the event of newly developed re-identification risks.
    - Develop cross-referencing inventories of direct and indirect identifiers published to the open data portal.

- To fully mature its toolbox of de-identification tools and mitigation strategies the City of Seattle should, as appropriate:
    - Develop policies and procedures for conducting additional screening of datasets and of elevating the review of risky or sensitive datasets to disclosure control experts or a disclosure review board where appropriate.
    - Develop or obtain appropriate tools to de-identify unstructured or dynamic data types.
    - Consult statistical disclosure control experts and invest in programmatic tools to evaluate re-identification risk across datasets (including King County, Washington State, federal open data, and commercial databases).
    - Consult statistical disclosure control experts about and invest in differential privacy or secure multi-party computation solutions for releasing data that poses a risk to privacy, to provide the strongest known protection against re-identification attacks today.
    - Develop policies and procedures to address legacy data on data.seattle.gov and to remove or modify existing datasets that pose an inappropriate risk of re-identification.
    - Investigate options for a limited-access or controlled-access scheme for more sensitive datasets (such as a data enclave, contractual safeguards, or tiered access model).
    - Create an internal or external disclosure review board that is accountable and transparent, with diverse representation and interdisciplinary capability to evaluate datasets requiring advanced review (such as datasets involving sensitive data, where municipal employees are data subjects, or data that could pose social justice concerns).
    - Adopt vendor contracts (such as with open data platform providers) that support the development and deployment of differentially private open data tools.

- To optimize its measures and protections for data quality, the City of Seattle should, as appropriate:
    - Develop procedures and mechanisms for individuals to submit correction requests for potentially incorrect personal data posted on the open data portal.
    - Actively check for inaccurate or outdated personal data in published datasets.
    - Develop mechanisms to trigger re-assessment of published datasets in light of new facts.
    - Monitor and maintain best practices and standards for data quality.
    - Continue to develop clear and consistent metadata standards for individual datasets, particularly communicating any data quality or privacy concerns.

- To mature its approach to equity and fairness within its Open Data Program, the City of Seattle should, as appropriate:
    - Develop policies and procedures for ensuring that individuals are provided clear notice when data they provide to the city is reasonably likely to be published publicly (particularly when data is solicited in unstructured formats, such as 311 requests).
    - Develop policies and procedures for ensuring that individuals are provided with reasonable choices about data collection (such as an ability to opt out of data collection, or to opt-out of having their data included in an open dataset).

    - Where individuals' consent to data collection will not be obtained, or where it may be too costly for some individuals to opt out of data collection, develop additional privacy controls to ensure that personal data is used fairly (such as not publishing the data publicly, or restricting its use to a narrower band of purposes).
    - Review and respect contract terms or conditions when acquiring data from third parties.
    - Provide mechanisms for individuals, community groups, and other data users to submit complaints about the open data process and develop policies and procedures for responding to those complaints.
    - Continue to invest in public engagement and communications strategies that seek to include the input of underserved or vulnerable populations into the Open Data Program.
    - Develop or obtain tools for evaluating the representativeness of the city's open data (including whether underserved or vulnerable populations are over- or under-represented in certain way).
    - Develop policies, procedures, and technical tools for evaluating the equity, fairness, and social justice impacts of releasing open datasets.
    - Consult statistical fairness experts and ethicists and invest in the development of programmatic tools for evaluating unfairness within the city's open datasets.
    - Create an internal or external ethical review board that is accountable and transparent, with diverse representation and interdisciplinary capability to evaluate datasets requiring advanced review (such as datasets involving vulnerable populations, where municipal employees are data subjects, or data that could pose social justice concerns).

- To manage and optimize efforts to engage and educate the public about open data, the City of Seattle should, as appropriate:
    - Develop additional methods for communicating with individuals at the point of data collection about how their data is reasonably likely to be used or published (particularly when data is solicited from individuals in unstructured formats, such as 311 requests).
    - Develop and share educational materials specific to privacy and open data with the public, using language and formats that are easy for diverse communities to understand.
    - Educate and engage local stakeholders in discussions about the equity, fairness, and social justice impacts of releasing open datasets.
    - Continue to directly engage local stakeholders in the development of privacy protections for open data.
    - Formalize inclusive methods for incorporating the public interest into determinations about what datasets to publish.
    - Develop and share educational materials about the intersection of open data and new sensor data from Smart City devices if and when those devices are deployed.
    - Strive to include local community stakeholders in the composition of any ethical or disclosure review board that is established.

The City of Seattle is one of the most innovative cities in the country, with an engaged and civic-minded citizenry, active city leadership, and technologically sophisticated business community. The city's appreciation for both the promises and the risks of open data is apparent in its thoughtful and thorough approach to protecting individual privacy. While there are certainly aspects of the Seattle Open Data Program that require improvement and further capacity-building, including the need to comprehensively assess the potential benefits and risks of each dataset and to evaluate re-identification risks across multiple datasets, the city's existing organizational structure and data handling practices provide a solid foundation for growth.

By continuing to complement its growing Open Data Program with robust privacy protections and policies, it will be possible for the City of Seattle to live up to the promise of its Open Data Policy, supporting civic innovation while protecting individual privacy.

Appendix A: Additional Resources

AICPA/CICA Privacy Task Force, AICPA/CICA Privacy Maturity Model, (2011), https://www.kscpa.org/writable/files/AICPADocuments/10-229_aicpa_cica_privacy_maturity_model_finalebook.pdf.

Micah Altman et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 Berkeley Tech. L.J. 1968 (2015), https://cyber.harvard.edu/publications/2016/Privacy_Aware_Government_Data_Releases.

Sean Brooks et al., An Introduction to Privacy Engineering and Risk Management in Federal Systems NISTIR 8062 (NIST Jan. 2017), http://nvlpubs.nist.gov/nistpubs/ir/2017/NIST.IR.8062.pdf.

Joseph A. Cannataci, Report of the Special Rapporteur on the right to privacy (Appendix on Privacy, Big Data, and Open Data) (Human Rights Council, Mar. 8, 2016), www.ohchr.org/Documents/Issues/Privacy/A-HRC-31-64.doc.

Lorrie Cranor*, Open Police Data Re-identification Risks*, Tech@FTC Blog (April 27, 2016, 3:31 PM), https://www.ftc.gov/news-events/blogs/techftc/2016/04/open-police-data-re-identification-risks.

David Doyle, Open Government Data: an analysis of the potential impacts of an Open Data law for Washington State (2015) (unpublished M.P.P. thesis, University of Washington Bothell), https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/34826/Doyle%20-%20Capstone.pdf?sequence=1.

Khaled El Emam, *A de-identification protocol for open data*, IAPP (May 16, 2016), https://iapp.org/news/a/a-de-identification-protocol-for-open-data/.

Khaled El Emam, Guide to the De-Identification of Personal Health Information (CRC Press, 2013).

Khaled El Emam & Waël Hassan, A PRIVACY ANALYTICS WHITE PAPER: The De-Identification Maturity Model (PrivacyAnalytics, 2013), *available at* http://www.himss.org/privacy-analytics-de-identification-maturity-model.

Federal Committee on Statistical Methodology, *Report on Statistical Disclosure Limitation Methodology* (Federal Committee on Statistical Methodology, Statistical Policy Working Paper No. 22, 2005), https://www.hhs.gov/sites/default/files/spwp22.pdf.

Kelsey Finch & Omer Tene, *Welcome to the Metropticon: Protecting Privacy in a Hyperconnected Town*, 41 Fordham Urb. L.J. 1581 (2015), *available at* http://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=2549&context=ulj.

Kelsey Finch & Omer Tene, *The City as a Platform: Enhancing Privacy and Transparency in Smart Communities*, Cambridge Handbook of Consumer Privacy (forthcoming).

Erica Finkel, DataSF: Open Data Release Toolkit (2016), https://drive.google.com/file/d/0B0jc1tmJAlTcR0RMV01PM2NyNDA/view.

Simson L. Garfinkel, SP 800-188: De-Identifying Government Datasets (NIST draft. Aug. 2016), http://csrc.nist.gov/publications/drafts/800-188/sp800_188_draft2.pdf.

Simson L. Garfinkel, NISTIR 8053: De-Identifying Personal Information (NIST Oct. 2015), http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf.

Ben Green et al., *Open Data Privacy*, Berkman Klein Center for Internet & Society at Harvard (2017), https://dash.harvard.edu/bitstream/handle/1/30340010/OpenDataPrivacy.pdf.

Emily Hamilton, The Benefits and Risks of Policymakers' Use of Smart City Technology (Oct. 2016) (unpublished paper) (on file with the Mercatus Center at George Mason University).

Information Commissioner's Office, Anonymisation: managing data protection risk Code of Conduct (2012), *available at* https://ico.org.uk/media/1061/anonymisation-code.pdf.

ISO/IEC CD 20889: Information technology – Security techniques – Privacy enhancing data de-identification techniques, https://www.iso.org/standard/69373.html.

Anna Johnston, Demystifying de-identification: An introductory guide for privacy officers, lawyers, risk managers and anyone else who feels a bit bewildered*,* (Salinger Privacy, Feb. 2017).

Joint Task Force Transformation Initiative Interagency Working Group, Guide for Conducting Risk Assessments NIST 800-30 (NIST Sep. 2012), http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf.

Jeff Jonas & Jim Harper, *Open Government: The Privacy Imperative*, *in* Open Government: Collaboration, Transparency, and Participation in Practice (O'Reilly Media, 2010).

Rob Kitchin, The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences (Sage, 1st ed. 2014).

Yves-Alexandre de Montjoye et al., Unique in the Crowd: The privacy bounds of human mobility (Scientific Reports 3, Mar. 25, 2013), https://www.nature.com/articles/srep01376.

Sean A. Munson et al., Attitudes toward Online Availability of US Public Records (2011).

Arvind Narayanan et al., *A Precautionary Approach to Big Data Privacy*, *in* 24 Data Protection on the Move: Law, Governance and Technology Series (Serge Gutwirth, Ronald Leenes, Paul de Hert eds., 2016), *available at* https://link.springer.com/chapter/10.1007/978-94-017-7376-8_13.

Opinion of the Article 29 Data Protection Working Party on Anonymisation Techniques (2014), *available at* http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

Jules Polonetsky, Omer Tene & Kelsey Finch, *Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification*, 56 Santa Clara L. Rev. 594 (2016), *available at* http://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=2827&context=lawreview.

President's Council of Advisors on Science and Technology, Exec. Office of the President, Report to the President: Technology and the Future of Cities (Feb. 2016), *available at* https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_cities_report__final_3_2016.pdf.

Ira Rubinstein & Woodrow Hartzog, *Anonymization and Risk*, 91 Wash. L Rev. 703 (2016), http://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1589/91WLR0703.pdf?sequence=1&isAllowed=y.

Sander v. State Bar of California, 58 Cal. 4th 300 (2013).

Jan Whittington et al., *Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government*, 30 Berkeley Tech. L.J. 1899 (2015), http://btlj.org/data/articles2015/vol30/30_3/1899-1966%20Whittington.pdf.

Alexandra Wood et al., *Privacy and Open Data Research Briefing*, Berkman Klein Center for Internet & Society at Harvard (2016), https://dash.harvard.edu/bitstream/handle/1/28552574/04OpenData.pdf?sequence=1.

Frederik Zuiderveen Borgesius et al., *Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework*, 30 Berkeley Tech. L.J. 2075 (2015), http://btlj.org/data/articles2015/vol30/30_3/2073-2132%20Borgesius.pdf.

Public Comments to *City of Seattle Open Data Privacy Risk Assessment: Draft Report*

Public comments to *City of Seattle Open Data Privacy Risk Assessment: Draft Report* (July-Oct. 2017), *available at* https://fpf.org/2018/01/22/public-comments-on-proposed-open-data-risk-assessment-for-the-city-of-seattle/.

Seattle Resources

City of Seattle, City of Seattle 2017 Open Data Plan, http://www.seattle.gov/Documents/Departments/SeattleIT/City%20of%20Seattle%202017%20Open%20Data%20Plan.pdf.

City of Seattle, Open Data Playbook V. 1.0, http://www.seattle.gov/Documents/Departments/SeattleGovPortals/CityServices/OpenDataPlaybook_Published_2016.08.pdf.

City of Seattle, Open Data Policy, http://www.seattle.gov/Documents/Departments/SeattleGovPortals/CityServices/OpenDataPolicyV1.pdf.

City of Seattle, Open Data Program 2016 Annual Report, https://www.seattle.gov/Documents/Departments/SeattleIT/Open%20Data%20Program%202016%20Annual%20Report.pdf.

City of Seattle, Privacy Principles, https://www.seattle.gov/Documents/Departments/InformationTechnology/City-of-Seattle-Privacy-Principles-FINAL.pdf.

Seattle Information Technology: Community Technology Advisory Board (CTAB), Seattle.gov, https://www.seattle.gov/tech/opportunities/ctab.

Seattle Information Technology: Privacy, Seattle.gov, http://www.seattle.gov/tech/initiatives/privacy.

Seattle Information Technology: Open Dataset Inventory – Privacy and PII, Seattle.gov, https://view.officeapps.live.com/op/view.aspx?src=http://www.seattle.gov/Documents/Departments/SeattleIT/OpenDatasetInventory_Privacy_PII.docx.

Appendix B: Program Maturity Assessment

Municipal open data programs create privacy risks around re-identification, data quality and equity, and public trust. FPF provides the following assessment framework in order to help municipalities around the United States better evaluate their organizational structures and data handling practices related to open data privacy.

In conducting their own assessments, we recommend municipal leaders incorporate the following into their analyses: any public statements about the municipality's open data program, privacy commitments, and use of personal data; interviews with internal and external staff who have responsibility for open data or privacy, or who contribute to or rely on the municipality's published open datasets; public discussions with local community advisors about open data and privacy values within the community; expert opinions or guidance from statistical disclosure control professionals about calculating and mitigating re-identification risks; any relevant case law or legal opinions related to the intersection of public records laws and individual privacy; and any relevant vendor contracts that might condition the sharing of personal data. These materials should support and document the municipal open data program's activities in each privacy domain, and justify its maturity measures.

Municipalities should apply a consistent scoring mechanism to their answers within this framework. Our scoring of the City of Seattle's practices in each of the following domains was based on the AICPA/CICA Privacy Maturity Model (PMM) levels, which reflect Generally Accepted Privacy Principles (GAPP):[40]

- ***Undeveloped*** – procedures or processes are absent, or are unpredictable and reactive.
- ***Ad hoc*** – procedures or processes are generally informal, incomplete, and inconsistently applied.
- ***Repeatable*** – procedures or processes exist; however, they are not fully documented and do not cover all relevant aspects.
- ***Defined*** – procedures and processes are fully documented and implemented, and cover all relevant aspects.
- ***Managed*** – reviews are conducted to assess the effectiveness of the controls in place.
- ***Optimized*** – regular review and feedback are used to ensure continuous improvement towards optimization of the given process.

A key principle of the PMM approach is the recognition that "each organization's personal information privacy practices may be at various levels, whether due to legislative requirements, corporate policies or

[40] *See* AICPA/CICA Privacy Task Force, AICPA/CICA Privacy Maturity Model, (2011), https://www.kscpa.org/writable/files/AICPADocuments/10-229_aicpa_cica_privacy_maturity_model_finalebook.pdf

the status of the organization's privacy initiatives. It was also recognized that based on an organization's approach to risk, not all privacy initiatives would need to reach the highest level on the maturity model."[41]

Privacy leadership and program management

- Does the municipality employ a comprehensive, strategic, agency-wide privacy program regarding its open data initiatives?
- Has the municipality designated a privacy governance leader for open data?
- Is the open data program guided by core privacy principles and policies?
- Does the open data workforce receive effective privacy training and education?
- Are the municipality's open data privacy policies and procedures updated in light of ongoing monitoring and periodic assessments?

Maturity score and supporting rationale:

Benefit-risk assessment

- Does the open data program conduct a benefit-risk assessment to manage privacy risk in each dataset considered for publication?
- Are datasets assessed based on the identifiability, sensitivity, and utility of the data prior to release?
- Are inventories of published personally identifiable information (PII) maintained?
- Are benefit-risk assessments documented and regularly reviewed?
- Does the open data program have a mechanism in place to trigger re-assessment of a published dataset in light of new facts?
- Does the open data program have an ability to elevate review of risky or sensitive datasets to disclosure control experts or a disclosure review board?

Maturity score and supporting rationale:

De-identification tools and strategies

- Does the open data program utilize technical, legal, and administrative safeguards to reduce re-identification risk?
- Does the open data program have access to disclosure control experts to evaluate re-identification risk?

[41] *See id.*

- Does the open data program have access to appropriate tools to de-identify unstructured or dynamic data types? (e.g., geographic, video, audio, free text, real time sensor data)
- Does the open data program have policies and procedures for evaluating re-identification risk across databases? (e.g., risk created by intersection of multiple municipal databases; county, state, or federal open databases; commercial databases)
- Does the open data program evaluate privacy risk in light of relevant public records laws?

Maturity score and supporting rationale:

Data quality

- Does the municipality employ policies and procedures for the open data program to ensure that personally identifiable information is accurate, complete, and current?
- Does the open data program check for, and correct as appropriate, inaccurate or outdated personally identifiable information?
- Are there procedures or mechanisms for individuals to submit correction requests for potentially incorrect personal data posted on the open data program?

Maturity score and supporting rationale:

Equity and fairness

- Were the conditions under which the data was collected fair? (e.g., were citizens aware that the data would be published on the open data portal? Did individuals have an opportunity to opt out of data collection? If data was acquired from a third party, were terms and conditions observed in the collection, use, maintenance, and sharing of the data?)
- Does the open data program assess the representativeness of the data? (e.g. whether underserved or vulnerable populations are appropriately represented in the data, or whether underserved or vulnerable populations' interests are taken into account when determining what data to publish).
- Are any procedures and mechanisms in place for people to submit complaints about the use of data or about the publication process generally, as well as procedures for responding to those complaints?

Maturity score and supporting rationale:

Transparency and public engagement

- Does the open data program engage and educate the public about the benefits of open data?

- Does the open data program engage and educate the public about the privacy risks of open data?
- Does the open data program provide opportunities for public input and feedback about the portal, the data available, and privacy, utility, or other concerns?
- Does the open data program engage with the public when developing of open data privacy protections?
- Does the open data program consider the public interest in determining what datasets to publish?
- Does the open data program communicate with the public about why some datasets may include PII?

Maturity score and supporting rationale:

Overall Open Data Program privacy maturity score:

Maturity score and supporting rationale:

Appendix C: Model Benefit-Risk Analysis

Step 1: Evaluate the Information the Dataset Contains

Dataset: ________________________________

Consider the following categories of information:

- *Direct Identifiers:* These are data points that identify a person without additional information or by linking to other readily available information. "Personally Identifiable Information," or PII, often falls within this category. For example, they can be names, social security numbers, or an employee ID number. (*See, e.g.,* municipal guidance like Seattle's PII/Privacy in the Open Dataset Inventory). Publishing direct identifiers creates a *very high* risk to privacy because they directly identify an individual and can be used to link other information to that individual.

- *Indirect Identifiers:* These are data points that do not directly identify a person, but that in combination can single out an individual. This could include information such as birth dates, ZIP codes, gender, race, or ethnicity. (*See, e.g.,* municipal guidance like Seattle's PII/Privacy in the Open Dataset Inventory). In general, to preserve privacy, experts recommend including no more than 6-8 indirect identifiers in a single dataset.[42] If a dataset includes 9 or more indirect identifiers there is a *high* or *very high* risk to privacy because they can indirectly identify an individual.

- *Non-Identifiable Information:* This is information that cannot reasonably identify an individual, even in combination. For example, this might include city vehicle inventory or atmospheric readings. This data creates *very low* or *low* risk to privacy.

- *Sensitive Attributes:* These data points that may be sensitive in nature. Direct and indirect identifiers can be sensitive or not, depending on context. For example, this might include financial information, health conditions, or a criminal justice records. Sensitive attributes typically create *moderate*, *high*, or *very high* risk to privacy.

- *Spatial Data and Other Information that Is Difficult to De-identify:* Certain categories or data are particularly difficult to remove identifying or identifiable information from, including: geographic locations, unstructured text or free-form fields, biometric information, and photographs or videos.[43] If data to be included in a public dataset are in one of these formats, they may create a *high* or *very high* risk to privacy.

[42] *See* Khaled El Emam, *A De-Identification Protocol for Open Data*, IAPP (MAY 16, 2016), https://iapp.org/news/a/a-de-identification-protocol-for-open-data/.

[43] *See* GARFINKEL, *supra* note 9, at 32-33.

Consider how linkable the information in this dataset is to other datasets:

- Do any of the dataset’s direct or indirect identifiers currently appear in other readily accessible open datasets (e.g., other municipal county, or state open datasets)? If this information is present in multiple open datasets, it increases the chances of identifying an individual and increases the risk to privacy.

- How often is the dataset updated? In general, the more frequently a dataset is updated—every fifteen minutes versus every quarter, for example—the easier it is to re-identify an individual and the greater the risk to privacy.

- How often is the information in this dataset requested by public records?

Consider how the information in this dataset was obtained:

- In what context was this data collected? Is this data collected under a regulatory regime? Are there any conditions, such as a privacy policy or contractual term, attached to the data? If the personal information in this dataset collected directly from the individual or from a third party?

- Would there be a reasonable expectation of privacy in the context of the data collection? For example, if the public has no notice of the data collection or data are collected from private spaces, there may be an expectation of privacy.

- Was the collection of the information in this dataset controversial? Was any of the information in this dataset collected by surveillance technologies (e.g., body-worn cameras, surveillance cameras, unmanned aerial vehicles, automatic license plate readers, etc.)?

- Has this dataset been checked for accuracy? Is there a mechanism for individuals to have information about themselves in this dataset corrected or deleted?

- Is there a concern that releasing this data may lead to public backlash or negative perceptions?

Step 2: Evaluate the Benefits Associated with Releasing the Dataset

List some of the foreseeable benefits of publishing the data fields included in this dataset and identify whether this use typically involves aggregate data or individual records. For example, measuring atmospheric data at particular locations over time may reveal useful weather patterns, and tracking building permit applications may reveal emerging demographic or commercial trends in particular neighborhoods.

Consider the likely users of this dataset. Who are the ideal users? Check all that apply.

- □ Individuals
- □ Community Groups
- □ Journalists
- □ Researchers
- □ Companies or Private Entities
- □ Other Government Agencies or Groups
- □ Other: ___________________________

Assess the scope of the foreseeable benefits of publishing the dataset:

Qualitative Value	Quantitative Value	Description
Very High	10	The dataset will likely have *multiple compelling and important* utilities for individuals, the community, other organizations, or society.
High	8	The dataset will likely have a *compelling and important* utility for individuals, the community, other organizations, or society.
Moderate	5	The dataset will likely have a *clear* utility for individuals, the community, other organizations, or society. While the utility is clear, it is not as urgent as a “high” value.
Low	2	The dataset will likely have a *limited* utility for individuals, the community, other organizations, or society.
Very Low	0	The dataset will likely have *negligible* utility for organizations, the community, other organizations, or society.

Next, assess the likelihood that the desired benefits of releasing this dataset would occur:

Qualitative Value	**Quantitative Value**	**Description**
Very High	10	The benefit is *almost certain* to occur.
High	8	The benefit is *highly likely* to occur.
Moderate	5	The benefit is *somewhat likely* to occur.
Low	2	The benefit is *unlikely* to occur.
Very Low	0	The benefit is *highly unlikely* to occur.

Combining your rating of the foreseeable benefits of the dataset with the likelihood that these benefits will occur, assess the overall benefit of this dataset:

Likelihood of Occurrence	**Impact of Foreseeable Benefits**				
	Very Low Impact	**Low Impact**	**Moderate Impact**	**High Impact**	**Very High Impact**
Very High Likelihood	Low Benefit	Moderate Benefit	High Benefit	Very High Benefit	Very High Benefit
High Likelihood	Low Benefit	Moderate Benefit	Moderate Benefit	High Benefit	Very High Benefit
Moderate Likelihood	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit	High Benefit
Low Likelihood	Very Low Benefit	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit
Very Low Likelihood	Very Low Benefit	Very Low Benefit	Low Benefit	Low Benefit	Low Benefit

Step 3: Evaluate the Risks Associated with Releasing the Dataset

Consider the foreseeable privacy risks of this dataset:[44]

- *Re-identification (and false re-identification) impacts on individuals*
  - Would a re-identification attack on this dataset expose the person to identity theft, discrimination, or abuse?
  - Would a re-identification attack on this dataset reveal location information that could lend itself to burglary, property crime, or assault?
  - Would a re-identification attack on this dataset expose the person to financial harms or loss of economic opportunity?
  - Would a re-identification attack on this dataset reveal non-public information that could lead to embarrassment or psychological harm?

- *Re-identification (and false re-identification) impacts on the organization*
  - Would a re-identification attack on this dataset lead to embarrassment or reputational damage to the City of Seattle?
  - Would a re-identification attack on this dataset harm city operations relying on maintaining data confidentiality?
  - Would a re-identification attack on this dataset expose the city to financial impact from lawsuits, or civil or criminal sanctions?
  - Would a re-identification attack on this dataset undermine public trust in the government, leading to individuals refusing to consent to data collection or providing false data in the future?

- *Data quality and equity impacts*
  - Will inaccurate or incomplete information in this dataset create or reinforce biases towards or against particular groups?
  - Does this dataset contain any incomplete or inaccurate data that, if relied upon, would foreseeably result in adverse or discriminatory impacts on individuals?
  - Will any group or community's data be disproportionately included in or excluded from this dataset?
  - If this dataset is de-identified through statistical disclosure measures, did that process introduce significant inaccuracies or biases into the dataset?

[44] Special thanks to Simson Garfinkel and Khaled El Emam whose works provide a foundation for articulating this analytic framework. *See* DE-IDENTIFICATION OF PERSONAL INFORMATION 32-33 (NIST 2015), DE-IDENTIFYING GOVERNMENT DATASETS SP 800-188; Khaled El Emam, *A De-Identification Protocol for Open Data*, IAPP (MAY 16, 2016), https://iapp.org/news/a/a-de-identification-protocol-for-open-data/; KHALED EL EMAM, GUIDE TO THE DE-IDENTIFICATION OF PERSONAL HEALTH INFORMATION (2013).

- *Public trust impacts*
  - Does this dataset have information that would lead to public backlash if made public?
  - Will local individuals or communities be shocked or surprised by the information about themselves in this dataset?
  - Is it likely that the information in this dataset will lead to a chilling effect on individual, commercial, or community activities?
  - Is there any information contained within the dataset that would, if made public, reveal nonpublic information about an agency's operations?

Consider who could use this information improperly or in an unintended manner (including to re-identify individuals in the dataset). Check all that apply.

□ General public (individuals who might combine this data with other public information)

□ Re-identification expert (a computer scientist skilled in de-identification)

□ Insiders (a municipal employee or contractor with background information about the dataset)

□ Information brokers (an organization that systematically collects and combines identified and de-identified information, often for sale or reuse internally)

□ "Nosy neighbors" (someone with personal knowledge of an individual in the dataset who can identify that individual based on the prior knowledge)

□ Other: ______________________________________

Assess the scope of the foreseeable privacy risks of publishing the dataset:

Qualitative Value	**Quantitative Value**	**Description**
Very High	10	The dataset will likely have *multiple severe or catastrophic* adverse effects on individuals, the community, other organizations, or society.
High	8	The dataset will likely have a *severe or catastrophic* adverse effect on individuals, the community, other organizations, or society.
Moderate	5	The dataset will likely have a *serious* adverse effect on individuals, the community, other organizations, or society.
Low	2	The dataset will likely have a *limited* adverse impact on individuals, the community, other organizations, or society,
Very Low	0	The dataset will likely have a *negligible* adverse impact on individuals, the community, other organizations, or society.

Next, assess the likelihood that the foreseeable privacy risks of releasing this dataset would occur:

Qualitative Value	Quantitative Value	Description
Very High	10	The risk is *almost certain* to occur.
High	8	The risk is *highly likely* to occur.
Moderate	5	The risk is *somewhat likely* to occur.
Low	2	The risk is *unlikely* to occur.
Very Low	0	The risk is *highly unlikely* to occur.

Combining your rating of the foreseeable risks of the dataset with the likelihood that these risks will occur, assess the overall risk of this dataset:

Likelihood of Occurrence	**Impact of Foreseeable Risks**				
	Very Low Impact	**Low Impact**	**Moderate Impact**	**High Impact**	**Very High Impact**
Very High Likelihood	Low Risk	Moderate Risk	High Risk	Very High Risk	Very High Risk
High Likelihood	Low Risk	Moderate Risk	Moderate Risk	High Risk	Very High Risk
Moderate Likelihood	Low Risk	Low Risk	Moderate Risk	Moderate Risk	High Risk
Low Likelihood	Very Low Risk	Low Risk	Low Risk	Moderate Risk	Moderate Risk
Very Low Likelihood	Very Low Risk	Very Low Risk	Low Risk	Low Risk	Low Risk

Step 4: Weigh the Benefits against the Risks of Releasing the Dataset

Step 4A: Combine the overall scores from the benefit and risk analyses to determine the appropriate solution for how to treat the dataset.

Benefit	**Risks**				
	Very Low Risk	**Low Risk**	**Moderate Risk**	**High Risk**	**Very High Risk**
Very High Benefit	Open	Open	Limit Access	Additional Screening	Additional Screening
High Benefit	Open	Limit Access	Limit Access	Additional Screening	Additional Screening
Moderate Benefit	Limit Access	Limit Access	Additional Screening	Additional Screening	Do Not Publish
Low Benefit	Limit Access	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish
Very Low Benefit	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish	Do Not Publish

- *Open*: Releasing this dataset to the public presents low or very low privacy risks and the potential benefits of the dataset substantially outweigh the potential privacy risks.
- *Limit Access*: Releasing this data presents moderate to very low privacy risks and the potential benefits of the dataset outweigh the potential privacy risks. In order to reduce the privacy risk, limit access to the dataset (such as by attaching contractual/Terms of Service terms to the dataset prohibiting re-identification attempts).
- *Additional Screening:* Releasing this dataset presents high privacy risks and the benefits could outweigh the potential privacy risks, or releasing this dataset presents privacy risk and the potential benefits do not outweigh the potential privacy risks. In order to reduce the privacy risk, formal application and oversight mechanisms should be considered (such as a disclosure review board, data use agreements, or a secure data enclave).
- *Do Not Publish*: Releasing this dataset presents very high to moderate privacy risks and the potential privacy risks of the dataset substantially outweigh the potential benefits. This dataset should remain closed, unless the risk can be reduced or there are countervailing public policy reasons for publishing it.

If the above table results in an “Open” categorization, then record the final benefit-risk score and continue preparing to publish the dataset. If the above table does *not* result in an “Open” categorization, then proceed to Step 4B by applying appropriate de-identification controls to mitigate the privacy risks for this dataset. The de-identification methods described below will be appropriate for some datasets, but not for others. Advances are always being made in de-identification techniques, and some tools may require disclosure control experts to properly implement. In the long-term, municipalities should strive to incorporate the expertise of disclosure control professionals and to implement mathematically provable privacy protections like differential privacy.

Consider the level of privacy risks you are willing to accept, the overall benefit of the dataset, and the operational resources available to mitigate re-identification risk. Note that the more invasive the de-identification technique, the greater the loss of utility will be in the data, but also the greater the privacy protection will be.

Technical Controls[45]

Method	**Description**	**Privacy Impact**	**Utility Impact**	**Operational Costs**
Suppression	Removing a data field or an individual record to prevent the identification of individuals in small groups or those with unique characteristics.	Removing the field removes the risk created by those fields, and lowers the likelihood of linking one dataset to another based on that information. Removing individual records can also effectively protect the privacy of those individuals. Suppression cannot guarantee absolute privacy, because there is always a	This approach removes all utility added by the suppressed field or record, and could skew the results or give false impressions about the underlying data.	This is a relatively low-cost method of de-identification. Removing entire fields of data can be both a quick and relatively low-tech process. When removing records one-by-one, particularly large datasets, there is a risk that some records may be overlooked.[46]

[45] Special thanks to the Berkman Klein Center for Internet & Society at Harvard University whose work provides a foundation for this analytic framework. BEN GREEN ET AL, OPEN DATA PRIVACY (2017), https://dash.harvard.edu/handle/1/30340010; Micah Altman et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 BERKELEY TECH. L.J. 1968 (2015), https://cyber.harvard.edu/publications/2016/Privacy_Aware_Government_Data_Releases.

[46] *See* Fitzpatrick, *supra* note 9.

Method	Description	Privacy Impact	Utility Impact	Operational Costs
		chance that the remaining data can be re-identified using an auxiliary dataset.		
Generalization/Blurring	Reducing the precision of disclosed data to minimize the certainty of individual identification, such as by replacing precise data values with ranges or sets.	The more specific a data value is, the easier it will generally be to single out an individual. However, even relatively broad categories cannot guarantee absolute privacy, because there is always a chance that the remaining data can be re-identified using an auxiliary dataset.	Generalizing data fields can render data useless for more granular analysis, and may skew results slightly or give false impressions about the underlying data.	Generalizing data fields can be a quick and straightforward process for reducing the identifiability of particular fields after the initial thresholds are set. In order to determine the appropriate level of generalization for particular data types, additional research or expert consultation may be required.
Pseudonymization	Replacing direct identifiers with a pseudonym (such as a randomly generated value, an encrypted identifier, or a statistical linkage key).	Pseudonymization removes the association between an individual and their data, and replaces it with a less easily identifiable key, lowering but not eliminating the risk of re-identification. Pseudonymization can be reversed in many circumstances, and are often considered personally identifiable information by privacy and data protection authorities.	Pseudonymization can allow for information about an individual to be linked across multiple records, increasing its utility for a wide variety of purposes.	Pseudonymization can appear relatively straightforward and cost-effective, however creating *irreversible* pseudonyms suitable for open data release can require significant effort.[47] Most successful re-identification attacks on openly released data have come from data that was

[47] *See* GARFINKEL, *supra* note 9, at 17.

Method	Description	Privacy Impact	Utility Impact	Operational Costs
				inadequately pseudonymized.[48]
Aggregation	Summarizing the data across the population and then releasing a report based on those data (such as contingency tables or summary statistics), rather than releasing individual-level data.	Aggregating data can be an effective method for protecting privacy as there is no raw data directly tied to an individual, however experts recommend minimum cell sizes of 5-10 records.[49]	Aggregation is more useful for examining the performance of a group or cohort. Because the raw data is not presented, it cannot be relied on to generate additional insights.	This method of de-identification requires slightly more expertise than simply removing fields or records. After an initial learning curve, the method can be implemented without significant costs. Expert consultants or guidance from federal statistical agencies may provide guidance in setting minimum cell sizes or addressing particular data types.[50]
Visualizations	Rather than providing users access to raw microdata, data may be presented in more privacy-protective formats, such as data visualizations or heat maps.	When data is released in non-tabular formats, individual data records are typically more obscure and harder to link to other auxiliary datasets, protecting individual privacy.	Data released in these sorts of formats may still be highly useful for a range of purposes, although not all. These formats may also limit the ways in which datasets can be combined or built on to generate new insights.	These are fairly low-cost approaches to limiting privacy risks, with numerous public resources readily available to Open Data program staff. Data that update frequently may be harder to maintain.

[48] *See* Ira Rubinstein & Woodrow Hartzog, *Anonymization and Risk,* 91 WASH. L REV. 703 (2016), http://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1589/91WLR0703.pdf?sequence=1&isAllowed=y; Jules Polonetsky, Omer Tene & Kelsey Finch, *Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification,* 56 SANTA CLARA L. REV. 594 (2016).

[49] *See* Khaled El Emam, Comment Letter on Proposed Rule to Protect the Privacy of Customers of Broadband and Other Telecommunications Services; Khaled El Emam, *Protecting Privacy Using k-Anonymity,* 15 J. AM. MED. INFORMATICS ASS'N (2008).

[50] *Id*.

Method	Description	Privacy Impact	Utility Impact	Operational Costs
			Visualizations and other alternative data formats may also be more engaging to the lay public than raw tabular data.	
Perturbation	An expert adds “noise” to the dataset (such as swapping values from one record to another, or replacing one value with an artificial value), making it difficult to distinguish between legitimate values and the “noise.”	The false data in the field makes re-identification much less likely to occur. The noise makes it difficult to determine if re-identification is associated with a specific individual.	Utility decreases as the amount of noise in the data increases. The proportionate amount of legitimate data is reduced as false data is added.	This is costly in that it requires an expert. The type of noise, as well as the amount to be added will have a drastic difference, and to ensure a retention in utility, it must be completed by an expert. However, research shows that “even relatively small perturbations to the data may make re-identification difficult or impossible.”[51]
k-Anonymity	A technique to measure and limit how many individuals in a dataset have the same combination of identifiers. K-anonymity suppresses or generalizes identifiers and perturbs outputs until a particular k-value is reached.	Privacy protection is greater as the value of “k” increases. Experts recommend that the k-value for open datasets should be at least k=11 (that is, for every combination of identifiers in a dataset, there should be at least 11 equivalent records).[52]	As with the above controls, the negative impact on utility increases as k-value increases. In order to achieve k=11, significant portions of some datasets may need to be suppressed or generalized.	This is a costly, complex, and time-consuming method. An expert in de-identification and k-anonymity is necessary to ensure that the k-value is correct and will provide the desired level of protection and utility. Subsequent research has led to additional requirements

[51] *See* GARFINKEL, *supra* note 9, at 29.
[52] El Emam, *supra* note 42.

Method	Description	Privacy Impact	Utility Impact	Operational Costs
				for the diversity of sensitive attribute within k-anonymous datasets (l-diversity) and statistical relationship to the original data (t-closeness).[53]
Differential Privacy	A formal mathematical definition of privacy, which may be satisfied by a range of techniques if the result of an analysis of a dataset is the same before and after the removal of a single data record.	Differential private solutions increase privacy for all individuals in a dataset and provide mathematical guarantees against a wider range of re-identification attacks than traditional de-identification techniques. Some differential privacy solutions rely on limiting the number of queries completed to prevent maintain a proven minimum privacy threshold (often known as the "privacy budget"). The more queries performed on a function, the more the total "leakage" increases. The leakage can never decrease, and there is an acceptable level of leakage that can occur before a privacy risk becomes likely	As with other above tools, differential private solutions decrease the accuracy of analysis performed on the dataset. The amount of noise is calibrated to the amount of privacy protection offered, and in larger datasets may be negligible.[56] In other deployments, the level of utility in a differentially private dataset may be dependent upon the number of queries to be made in the dataset. Once the leakage threshold is hit, the dataset can no longer be used. However, if the desired task can be accomplished under the leakage threshold, the dataset retains great	Differential privacy requires an expert to calculate the leakage threshold, the amount of noise to add, and other statistical nuances. It may also require an interactive query system to be established, or trained users who can create data summaries for release and use. Therefore, it carries a higher operational cost than other methods of de-identification. Differential privacy is an active research area, and while to date it has only been applied to a few operational system,[59] differential privacy tools for use by non-experts in privacy, computer science, and statistics are also

[53] *See* GARFINKEL, *supra* note 9, at 12.

[56] *Comment by Alexandra Wood, Micah Altman, Suso Baleato, and Salil Vadhan* to Future of Privacy Forum (Oct. 3, 2017), *available at* https://fpf.org/wp-content/uploads/2018/01/Wood-Altman-Baleato-Vadhan_Comments-on-FPF-Seattle-Open-Data-Draft-Report.pdf.

[59] *See* GARFINKEL, *supra* note 9, at 7-9.

Method	Description	Privacy Impact	Utility Impact	Operational Costs
		and the dataset must be abandoned. Non-interactive differential privacy solutions such as synthetic data also provide strong privacy protection when sharing statistics,54 as “the privacy loss budget can be spent in creating the synthetic dataset, rather than in responding to interactive queries.”[55]	utility with little risk to privacy. In other cases, such as synthetic data (see below), differentially private tools may be non-interactive and so not limited by query amounts, such as by enabling data or data summaries to be released and used.[57] Datasets that may otherwise be too sensitive to share in individual-level formats could still be safely analyzed in differentially private formats, as well.[58]	currently in development.[60]
Synthetic Data	A process in which seed data from an original dataset is used to create artificial data that has some of the statistical characteristics as the seed	Synthetic datasets can make it very difficult and costly to map artificial records to actual people, and supports mathematical privacy guarantees with differential privacy that can remain in	Synthetic data “can be confusing to the lay public,” as they may contain artificial individuals who “appear quite similar to actual individuals in the population.”[64] The utility of synthetic data also	Synthetic databases may be confusing to both researchers and lay people, requiring additional efforts to educate data users about the dataset’s contents and limitations.

[54] *See* Wood et al., supra note 56 (citing Census, Google, Apple, Uber).

[55] GARFINKEL, *supra* note 9, at 52.

[57] *See* Wood et al., supra note 56.

[58] *See* Wood et al., supra note 56.

[60] *See* Wood et al., supra note 56. (citing e.g., Marco Gaboardi et al., PSI (Ψ): A Private Data Sharing Interface, Working Paper (2016), *available at* https://arxiv.org/abs/1609.04340).

[64] *Id*.

Method	Description	Privacy Impact	Utility Impact	Operational Costs
	data.[61] Datasets may be partially synthetic (in which some of the data is inconsistent with the original dataset) or fully synthetic (in which there is no one-to-one mapping between any record in the original dataset and the synthetic dataset).[62]	force "even if there are future data releases."[63]	depends on the model used to create it. Synthetic databases, unlike some differential privacy deployments, do not need to be released via interactive query systems, as "the privacy loss budget can be spent in creating the synthetic dataset, rather than in responding to interactive queries."[65]	

Administrative and Legal Controls

Method	Description	Privacy Impact	Utility Impact	Operational Costs
Contractual provisions	Data is made available to qualified users under legally binding contractual terms (such as commitments not to attempt to re-identify individuals or link datasets, to update the information	Contractual controls alone do not necessarily reduce the risk of re-identification, but when complementing the technical controls above can provide more flexible and contextual privacy protections. Contractual	Contractual provisions do not impede utility for acceptable data uses, although the compliance costs may deter some potential data users. Contractual terms prohibiting commercial uses may deter certain categories of users	Consistent contractual provisions must be developed and deployed, but this is a less extensive process than many of the technical measures above. Contractual provisions can also be tailored to the specific risk profiles of each

[61] GARFINKEL, *supra* note 9, at 48-49.
[62] *Id. at* 49-54.
[63] *Id*. at 51.
[65] *Id*. at 52.

	periodically, or to use data in noncommercial and nondiscriminatory ways).	terms are more robust when backed up by audit requirements and penalties for noncompliance.	(such as businesses or data brokers).[66]	dataset. There may be legal limits on how governments can restrict the use of data as well.[67]
Access fees	Charging users for access to data increases accountability and may discourage improper use of data.	Because fees are likely to deter many casual browsers of a particular datasets, the likelihood of accidental re-identification of an individual by a curious friend, neighbor, or acquaintance generally decreases. Tiered fee structures (e.g., that charge more for commercial access or remote versus in-person data access) may also lower the risk of re-identification by other actors. Charging fees may also introduce registration and audit capabilities, allowing Open Data program staff to identify which data users accessed which datasets.	The deterrent effect of access fees on the general public will impede the potential utility of the dataset and could limit access by some marginalized or vulnerable communities (e.g., those without credit cards, technological sophistication, or new market entrants).	Introducing access fees comes with initial and ongoing administrative overhead, and requires thoughtful determination of when particular datasets or classes of users warrant the use of fees.

[66] *See* Jan Whittington et al., supra note 13, at 1962.
[67] *Id.* at 1963.

Data enclaves	Physical or virtual environments are created that enable "authorized users to access confidential data and analyze the data using provided statistical software."[68]	Risks of re-identification are almost entirely removed by restricting external access to even de-identified data and introducing accountability and oversight measures. Technical controls may not need to be as strict, when complemented by administrative and legal safeguards (such as requiring researchers to apply for access, describe the proposed research, agree to confidentiality laws and penalties, audit logs, and authentication measures).	Data utility can be maximized for qualified researchers, as privacy protections are no longer purely technical. Researchers may be limited in what research questions can be asked and in the format of their results. But data utility is completely removed for any individual or organization that is not approved to access the dataset.	There are significant operational costs to maintaining a secure data enclave, including establishing policies and procedures for granting qualified researcher queries, for processing queries on de-identified data, for establishing the enclave, and for monitoring the program over time.
Tiered access controls	Systems in which data are made available to different categories of users through different mechanisms.[69]	Tiered access controls permit municipalities to craft more granular and contextual privacy protections depending on the sensitivity and identifiability of the data, and may support more accountability mechanisms (e.g., providing more	Limiting access to some datasets to particular types of users may increase the utility of data to those who qualify for greater access but decrease it for those who do not or cannot satisfy the access requirements. This may deter some members of	Establishing and monitoring an access-control system may require meaningful operational overhead. Consistent access terms and conditions will need to be defined, and deployed, and enforced. Access models that intend to do individualized

[68] *See* Micah Altman et al., *supra* note 23, at 40; GARFINKEL, *supra note* 9 at ix.
[69] *See* Wood et. al., supra note 56.

		sensitive or identifiable data only to potential data users who sign enforceable data use agreements or have their research questions vetted in advance).	the public from engaging with certain open datasets, but it may also provide municipal data leaders more oversight and insight into which data are most valuable to users.	vetting of some subsets of data users will likely require additional staffing.
Ethical and/or disclosure review board	Particularly risky or ambiguous policy decisions about a dataset are escalated to an advisory group with broad expertise and community engagement for further review.[70]	Review boards with diverse backgrounds and subject matter expertise can more robustly debate the benefits and risks of releasing a dataset and can address any additional dimensions not captured by the privacy risk assessment.	A review board may determine that a dataset's utility ultimately outweighs its impact on individual privacy; it may also determine that the benefits do *not* outweigh the risks.	Establishing and maintaining an accountable and transparent body of experts can be a challenging operational endeavor, although guidance and models from academic data research are available.[71]

[70] *See generally* Conference Proceedings: Beyond IRBs: Ethical Guidelines for Big Data Research, Future of Privacy Forum (Dec. 10, 2015), https://fpf.org/wp-content/uploads/2017/01/Beyond-IRBs-Conference-Proceedings_12-20-16.pdf.

[71] *See* 45 C.F.R. 46.102; Omer Tene & Jules Polonetsky, Beyond IRBs: Ethical Guidelines for Big Data Research 1 (Dec. 2015), https://bigdata.fpf.org/wp-content/uploads/2015/12/Tene-Polonetsky-Beyond-IRBs-Ethical-Guidelines-for-Data-Research1.pdf.

Step 4B: After determining and applying appropriate privacy controls and mitigations for the dataset, re-assess the overall risks and benefits of the dataset (Steps 1-3). Note any mitigation steps taken, and record the final benefit-risk score:

Benefit	**Risks**				
	Very Low Risk	**Low Risk**	**Moderate Risk**	**High Risk**	**Very High Risk**
Very High Benefit	Open	Open	Limit Access	Additional Screening	Additional Screening
High Benefit	Open	Limit Access	Limit Access	Additional Screening	Additional Screening
Moderate Benefit	Limit Access	Limit Access	Additional Screening	Additional Screening	Do Not Publish
Low Benefit	Limit Access	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish
Very Low Benefit	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish	Do Not Publish

If the score is still not "Open," consider using another mitigation method. If this is not possible, then determine whether to publish the dataset. If there may be countervailing public policy factors that should be considered, move on to Step 5.

- *Open*: Releasing this dataset to the public presents low or very low privacy risks and the potential benefits of the dataset substantially outweigh the potential privacy risks.
- *Limit Access*: Releasing this data presents moderate to very low privacy risks and the potential benefits of the dataset outweigh the potential privacy risks. In order to reduce the privacy risk, limit access to the dataset (such as by attaching contractual/Terms of Service terms to the dataset prohibiting re-identification attempts).
- *Additional Screening:* Releasing this dataset presents high privacy risks and the benefits could outweigh the potential privacy risks, or releasing this dataset presents privacy risk and the potential benefits do not outweigh the potential privacy risks. In order to reduce the privacy risk, formal application and oversight mechanisms should be considered (such as a disclosure review board, data use agreements, or a secure data enclave).
- *Do Not Publish*: Releasing this dataset presents high or very high privacy risks and the potential privacy risks of the dataset substantially outweigh the potential benefits. This dataset should remain closed, unless the risk can be reduced or there are countervailing public policy reasons for publishing it.

Step 5: Evaluate Countervailing Factors

Sometimes, a dataset with a very high privacy risk is still worth releasing into the open data portal in light of public policy considerations. For example, a dataset containing the names and salaries of elected officials would likely be considered high-risk due to the inclusion of a direct identifier. However, there is a compelling public interest in making this information available to citizens that outweighs the risk to individual privacy.

Additionally, there are always risks associated with maintaining and releasing any kind of data relating to individuals. Two key considerations when deciding whether to release the data irrespective of a potentially high or very high risk to individual privacy are:

1. If you are on the edge between two categories, analyze the dataset holistically but err on the side of caution. A dataset that is not released immediately can still be released at another date, as additional risk mitigation techniques become available. A dataset that has been released publicly, however, cannot ever be fully pulled back, even if it is later discovered to pose a greater risk to individual privacy. Be particularly cautious about moving data from an original recommendation of *Do Not Publish* to *Open*, and ensure that the potential benefits of releasing the data are truly so likely and compelling that they outweigh the existing privacy risks.

2. Any time you deviate from the original analysis, document your reasoning for doing so. This will not only help you decide whether the deviation is, in fact, the correct decision, but also provides accountability. Should the need arise, you will have a record of your reasoning, including analysis of the expected benefits and the recognized risks at the time. Where personally identifiable information is published notwithstanding the privacy risk, accountability mechanisms help maintain trust in the Open Data program that may otherwise be lost.

Appendix D: Model Analysis Applied to Current Seattle Open Data Content

The following sample datasets are included for illustrative purposes only to demonstrate some of the factors that could be considered in this type of benefit-risk analysis. As such, FPF has only provided an initial analysis (Steps 1-4A) of the current datasets, and has not prescribed specific mitigation interventions or potentially countervailing public policy rationales for publishing data that may pose a risk to individual privacy (Steps 4B and 5). We are outsiders to the City of Seattle and cannot substitute our judgment for those of the civic leaders and community members who must determine when privacy concerns outweigh the potential utility of data to the public.

- **Real Time Fire 911 Calls** – Moderate Benefit/Very High Risk.
  - Assessment: Do not publish (unless mitigated or countervailing public policy values identified).
- **Building Permits (Current)** – High Benefit/High Risk.
  - Assessment: Additional Screening (unless mitigated or countervailing public policy values identified).
- **Sold Fleet Equipment** – Moderate Benefit/Low Risk.
  - Assessment: Limit Access (unless mitigated or countervailing public policy values identified).
- **Seattle Communities Online Inventory** – Very High Benefit/Low Risk.
  - Assessment: Open.
- ***Road Weather Information Systems** – Very High Benefit/Very Low Risk.
  - Assessment: Open
  - **Note that as the Road Weather Information Systems dataset does not contain personally identifiable information, it typically would not undergo the full Benefit-Risk Assessment process. It is included here for illustrative purposes.*

Dataset: **Real Time Fire 911 Calls**

Step 1: Evaluate the Information the Dataset Contains

Consider the following categories of information:

- *Direct Identifiers:* There are no data that directly identify individuals in the Real Time Fire 911 Calls dataset.
- *Indirect Identifiers:* Indirect identifiers in this dataset include the address, latitude, and longitude of each call; the date and time of the call, and the type of response. If incident numbers in this dataset correlate to other city datasets, they may also help indirectly identify individuals.
- *Non-Identifiable Information:* If incident numbers are unique to this dataset, they may not be directly or indirectly identifying of individuals. All other data in this dataset is potentially identifiable.
- *Sensitive Attributes:* The type of response to a particular address could reveal sensitive information about individuals' reported health, safety and criminal justice conditions (e.g., "Assault with weapons," "Activated CO detector," "Aid Resp[iratory] Infections," "Fire in Single Family Res," "Illegal Burn," "Multiple casualty incident").
- *Spatial Data and Other Information that Is Difficult to De-identify:* Precise latitude and longitude and addresses are spatial data that are difficult to de-identify without impeding the data's utility to the public.

Consider how linkable the information in this dataset is to other datasets:

- *Do any of the dataset's direct or indirect identifiers currently appear in other readily accessible open datasets?* 911 incident data appear on the Seattle "911 Incidents & Police Reports" map for the first 24 hours after officers are dispatched and the incident is considered safe to close out, prior to being made available on the open data portal. 911 calls are also broadcast live on a variety of websites and mobile apps, which report on incidents in real time. Broadcastify.com, for example, allows listeners to hear live Seattle Fire response calls, which typically includes a report of the incident, its location, and the units sent on a live feed. Additional sensitive information may also be transmitted, depending on the nature of the call. Fire 911 call data may involve any number of locations, and addresses would be more or may sensitive depending on the type (e.g., single home residences vs. apartment complexes vs. commercial real estate vs. public lands or roadways, etc.). Depending on the kind of location, some addresses may also be available on building permit, park and recreation, food bank, or other Seattle open datasets. The King County and Washington State open data portals also contain numerous addresses related to public buildings and spaces, businesses, and community activities. County, state, and federal entities will also have access to private databases of address data linked to individuals who reside, own, or manage the property. Data brokers, commercial marketers, or any number of businesses that deliver to/provide services at such locations may also have information tying addresses to individuals. And of course the individuals who reside, own, or manage these properties – as well as their friends, family, coworkers, neighbors, and other community members – will also be able to connect those addresses to particular individuals. Social media posts, newspaper reports, and other public documents may also be used to easily tie individuals to particular addresses or incidents.

- *How often is the dataset updated?* Every five minutes.
- *How often is the information in this dataset requested by public records?* Unknown to FPF, but as of this report there were over 60,400 views and 459,000 downloads of the Sold Fleet Equipment dataset.

Consider how the information in this dataset was obtained:

- *In what context was this data collected? Is this data collected under a regulatory regime? Are there any conditions, such as a privacy policy or contractual term, attached to the data? If the personal information in this dataset collected directly from the individual or from a third party?* This data was collected by the city from its own departments to document 911 calls to the dispatcher. Any contractual terms or regulatory requirements are unknown to FPF. The most potentially identifying information in this dataset – the address of the incident – could be provided in a number of ways, including by the individual who owns the property, a bystander, a commercial alarm system, or by city employees or sensors.
- *Would there be a reasonable expectation of privacy in the context of the data collection?* The Fire 911 Calls dataset covers primarily situations where there is a health or safety emergency, and in which expectations of privacy may be outweighed by the need for an immediate response and the public's right to know about the actual deployment of emergency services. Nevertheless, individuals may be surprised to learn that information about the incident, including the address and type of response, would be made available forever on the open data portal. While there is unsettled case law surrounding the video broadcast of emergency crews at work on patients in some U.S. states, the sharing of more limited location and incident type information would seem to pose a much lower level of intrusion on individual privacy.
- *Was the collection of the information in this dataset controversial? Was any of the information in this dataset collected by surveillance technologies?* The collection of this data by the city is not controversial, and the data was not collected by surveillance technology.
- *Has this dataset been checked for accuracy? Is there a mechanism for individuals to have information about themselves in this dataset corrected or deleted?* It is unclear whether the dataset has been checked for accuracy, and there does not appear to be a mechanism for individuals to request information to be corrected or deleted, other than contacting the dataset owner (Seattle Fire Department).
- *Is there a concern that releasing this data may lead to public backlash or negative perceptions?* If the information were or could be used, in part, to identify individual(s) who were involved in sensitive health or criminal incidents, that could potentially create negative public impacts. The

ready availability of this information on commercial sites, social media, and news reports in real time, however, make the open data portal unlikely to be the primary source of re-identification or the primary recipient of negative sentiment.

Step 2: Evaluate the Benefits Associated with Releasing the Dataset

List some of the foreseeable benefits of publishing the data fields included in this dataset and identify whether this use typically involves aggregate data or individual records:

> The Real Time Fire 911 Calls dataset provides accountability for how the city responds to fire and emergency situations, which may be of interest to taxpayers, impacted individuals or companies, and local media. Journalists may be interested in tracking response and incident trends throughout the community, such as clusters of incidents around particular locations or building types or upticks in calls around holiday or major event dates. Individuals or communities could use this information to gain more insight into the type and relative frequency of responses in their neighborhoods, and to help residents better prepare or prevent such situations. Insurance, construction, or other kinds of companies could also potentially use this data to inform their business operations (such as times of year or locations most likely to need remediation or recovery services after a fire). Researchers or other government agencies, however, likely have access to richer databases for their analyses than the open dataset. These uses would typically involve aggregate data.

Consider the likely users of this dataset. Who are the ideal users? Check all that apply.

✓ Individuals
✓ Community Groups
✓ Journalists
□ Researchers
✓ Companies or Private Entities
□ Other Government Agencies or Groups
□ Other: ___________________________

Assess the scope of the foreseeable benefits of publishing the dataset:

Qualitative Value	Quantitative Value	Description
Moderate	5	The dataset will likely have a *clear* utility for individuals, the community, other organizations, or society. While the utility is clear, it is not as urgent as a "high" value.

Next, assess the likelihood that the desired benefits of releasing this dataset would occur:

Qualitative Value	Quantitative Value	Description
High	8	The benefit is *highly likely* to occur.

Combining your rating of the foreseeable benefits of the dataset with the likelihood that these benefits will occur, assess the overall benefit of this dataset:

Likelihood of Occurrence	**Impact of Foreseeable Benefits**				
	Very Low Impact	**Low Impact**	**Moderate Impact**	**High Impact**	**Very High Impact**
Very High Likelihood	Low Benefit	Moderate Benefit	High Benefit	Very High Benefit	Very High Benefit
High Likelihood	Low Benefit	Moderate Benefit	[✓ Moderate Benefit]	High Benefit	Very High Benefit
Moderate Likelihood	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit	High Benefit
Low Likelihood	Very Low Benefit	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit
Very Low Likelihood	Very Low Benefit	Very Low Benefit	Low Benefit	Low Benefit	Low Benefit

Step 3: Evaluate the Risks Associated with Releasing the Dataset

Consider the foreseeable privacy risks of this dataset:

- *Re-identification (and false re-identification) impacts on individuals*: The location of the incident, response type, and date/time of the 911 call are all available on this dataset. If the address or incident type identify a single family home or an area with a low population density, it is highly likely that individuals in the community would be able to identify the individuals involved with little or no effort. There are many ways that others could learn whatever additional information is needed to link an individual(s) to an address in this dataset, including news reports on the incident; social media references to the owner or resident of the address; other public records or open datasets; live broadcasts of emergency operations during the incident on 911 scanners; and commercial databases or transactions that may have involved the owner or resident of the address. Once the connection between an individual and an address is made, the incident type alone could reveal sensitive information. On the contrary, someone who knows the individual, the type of response, and the approximate time/date of the incident could potentially also use this dataset to learn the individual's address. Being identified as the potential source or subject of a 911 call (even falsely) – including incidents like a house fire, a car accident, a boating rescue, an encounter with an armed individual, proximity to an illegal burn, etc. – could open an individual up to significant harms. Depending on the nature of the incident, individuals could be targeted for identity theft, burglary, assault; they could be targeted by scammers or for insurance fraud; and they could have long-term reputational or emotional damage. At the same time, however, many Fire 911 calls are about commercial or public properties, and are less likely to lead to re-identification or harms to individuals. Small businesses may be more at risk than larger organizations, as well.
- *Re-identification (and false re-identification) impacts on the organization:* If information in the Real Time Fire 911 Calls dataset were used to re-identify an individual, there could be serious reputational damage to the city. While this data may be readily available on commercial sites, social media, and news reports in real time, the permanence and consolidation of incident data on the open data portal heightens the risk of re-identification in the long term. Even if the open data is not the only source for a re-identification attack, it could attract negative sentiments.
- *Data quality and equity impacts*: This dataset contains information about that, if inaccurate or incomplete, could potentially cast a negative light on individuals or communities, such as portraying a particular neighborhood as having erroneously high crime or incident rates. Inaccurate or incomplete data about Fire 911 Calls could also significantly affect individuals, for example if it conflicted with other reports and created confusion or was the basis for insurance rate changes across an area. Vulnerable or minority populations may be less inclined to instigate a call to the 911 system in general, and analyses relying on these data should keep such factors in mind. Perturbing or slightly modifying the existing data fields for the purposes of limiting the overall risk of re-identification – such as by providing less-precise location data for incidents – could mitigate biased impacts or re-identification risks towards individuals and groups. The Seattle Police 911 Calls dataset, for example, provides location only at the 100s block without any apparent impact on the dataset's popularity and usability. Statistical disclosure experts in other fields

might recommend generalizing data to differing geographic zone sizes (like neighborhood, census track, or ZIP codes) depending on factors like population density, property zoning and use, and re-identification risks. The Health Insurance Portability and Accountability Act (HIPAA) provides specific guidance for the de-identification of health data, including geographic limits (https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html).

- *Public trust impacts*: If an individual were re-identified, discriminated against, or adversely impacted by their inclusion in this dataset, there could certainly be public mistrust of the open data and 911 systems. It is unlikely that individuals would cease to use city emergency services, but it could possibly lead to individuals providing incomplete or misleading information out of fear or confusion about what information could be made public. The re-identification of crime victims from municipal data has attracted media attention before in cities like Dallas (See Report, p. 8). It is unknown to FPF if the Fire 911 Calls dataset would reveal any nonpublic information about the Fire Department's operations, but one anticipated use of this dataset by the public/academics/nonprofits/policy analysts would be investigations of the agency responsiveness to particular incident types or geographic areas.

Consider who could use this information improperly or in an unintended manner (including to re-identify individuals in the dataset). Check all that apply:

✓ General public
✓ Re-identification expert
✓ Insiders
✓ Information brokers
□ "Nosy neighbors"
□ Other: ______________________________

Assess the scope of the foreseeable privacy risks of publishing the dataset:

Qualitative Value	**Quantitative Value**	**Description**
Very High	10	The dataset will likely have a *multiple severe or catastrophic* adverse impact on individuals, the community, other organizations, or society.

Next, assess the likelihood that the foreseeable privacy risks of releasing this dataset would occur:

Qualitative Value	**Quantitative Value**	**Description**
High	8	The risk is *highly likely* to occur.

Combining your rating of the foreseeable risks of the dataset with the likelihood that these risks will occur, assess the overall risk of this dataset:

Likelihood of Occurrence	**Impact of Foreseeable Risks**				
	Very Low Impact	**Low Impact**	**Moderate Impact**	**High Impact**	**Very High Impact**
Very High Likelihood	Low Risk	Moderate Risk	High Risk	Very High Risk	Very High Risk
High Likelihood	Low Risk	Moderate Risk	Moderate Risk	High Risk	[✓ Very High Risk]
Moderate Likelihood	Low Risk	Low Risk	Moderate Risk	Moderate Risk	High Risk
Low Likelihood	Very Low Risk	Low Risk	Low Risk	Moderate Risk	Moderate Risk
Very Low Likelihood	Very Low Risk	Very Low Risk	Low Risk	Low Risk	Low Risk

Step 4: Weigh the Benefits against the Risks of Releasing the Dataset

Step 4A: Combine the overall scores from the benefit and risk analyses to determine the appropriate solution for how to treat the dataset.

Benefits	**Risks**				
	Very Low Risk	**Low Risk**	**Moderate Risk**	**High Risk**	**Very High Risk**
Very High Benefit	Open	Open	Limit Access	Additional Screening	Additional Screening
High Benefit	Open	Limit Access	Limit Access	Additional Screening	Additional Screening
Moderate Benefit	Limit Access	Limit Access	Additional Screening	Additional Screening	[✓ Do Not Publish]
Low Benefit	Limit Access	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish
Very Low Benefit	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish	Do Not Publish

***Do Not Publish*:** Releasing this dataset presents high or very high privacy risks and the potential privacy risks of the dataset substantially outweigh the potential benefits. This dataset should remain closed, unless the risk can be reduced or there are countervailing public policy reasons for publishing it. The City would assess appropriate risk mitigation measures and such countervailing interests in Steps 4B & 5.

Steps 4B & 5: At this point, the City would determine whether appropriate technical, legal, or administrative controls could lower the privacy risk of the Real Time Fire 911 Calls dataset further; reevaluate the dataset's risks and benefits; and identify any countervailing factors in favor of publication. The City would take into account the risk to privacy, the overall utility of the dataset, and the operational costs of further mitigations among other factors in determining what steps to take to complete the assessment.

Dataset: **Building Permits (Current)**

Step 1: Evaluate the Information the Dataset Contains

Consider the following categories of information:

- *Direct Identifiers:* Applicants' full names are available, as well as the address, latitude, and longitude for which permits are issued.
- *Indirect Identifiers:* Indirect identifiers in this dataset include the description of the work to be completed; the category of use or occupancy of the building where work is proposed; the value of the work being proposed; permit and complaint URLs; master use permit; and dates related to application and permit issuance/final inspections/expirations. Each of these could, in combination with other information, help identify or single out an individual property owner or occupant even if they are not already named as an applicant.
- *Non-Identifiable Information:* Permit type, status, action type, and work type could be considered non-identifiable, as they relate to permit administration generally and do not typically enable the look-up of a particular permit or individual.
- *Sensitive Attributes:* There value of the work being proposed may reveal information about individuals' financial status and the description of the work may reveal information about individuals' homes (e.g., moving a fireplace or bathroom, or repairing fire damage inside a single family home).
- *Spatial Data and Other Information that Is Difficult to De-identify:* Precise latitude/longitude are present related to permit worksites, and work description may be open text fields.

Consider how linkable the information in this dataset is to other datasets:

- *Do any of the dataset's direct or indirect identifiers currently appear in other readily accessible open datasets?* Information on Seattle's building permits are also available on the Seattle Department of Construction and Inspections website, which provides additional detail about inspections, reviews, land use, fees and receipts, occupancy and uses, and contacts related to permits and complaints. Seattle building permit data is also available in other formats, such as a data story and visualization at the Evergreen Data Library (https://evergreen.data.socrata.com/stories/s/5ru4-56sa) and combined with other Washington and Oregon State entities at the Daily Journal of Commerce (https://www.djc.com/const/bp.html). Building permit data is also widely used by data brokers and information resellers. Address data may also appear on Seattle Fire 911 and other datasets, depending on the type of building (residential, industrial, institutional, commercial, etc.) and any events or incidents at that location (such as a fire, an emergency, or a community event).

- *How often is the dataset updated?* The dataset was last updated Dec. 19, 2017. This dataset is refreshed daily.
- *How often is the information in this dataset requested by public records?* Unknown to FPF, but as of this report there were over 58,700 views and 61,300 downloads of the Building Permits (Current) dataset.

Consider how the information in this dataset was obtained:

- *In what context was this data collected? Is this data collected under a regulatory regime? Are there any conditions, such as a privacy policy or contractual term, attached to the data? If the personal information in this dataset collected directly from the individual or from a third party?* This data was collected by the city directly from prospective builders to record the issuance of building permits, who may be individuals, business owners, licensed contractors, or other direct stakeholders. Individuals who apply for permits online at the Seattle Department of Construction and Inspections can see a link to the City of Seattle Privacy and Security Policy at the lower right corner of the webpage throughout the application process.
- *Would there be a reasonable expectation of privacy in the context of the data collection?* The information in the dataset was collected from a government site with a link to the city's privacy policy and public links to other building permits, so there is likely no reasonable expectation of privacy for the permit applicant. Building permit data is also widely published on other municipal open datasets and used by local news and researchers. However, building owners and other occupants still may not be aware that detailed information about their building would be made permanently and publicly available once the permit application was filed.
- *Was the collection of the information in this dataset controversial? Was any of the information in this dataset collected by surveillance technologies?* The collection of this data by the city is not controversial, and the temperature sensor data was not collected by surveillance technology.
- *Has this dataset been checked for accuracy? Is there a mechanism for individuals to have information about themselves in this dataset corrected or deleted?* It is unclear whether this dataset has been checked for accuracy. There does not appear to be a mechanism for individuals to request information be corrected or deleted, other than contacting the dataset owner (Seattle Department of Planning and Development).
- *Is there a concern that releasing this data may lead to public backlash or negative perceptions?* If individuals are targeted for, e.g., insurance or marketing purposes based on information about their building's design or the value of work being done to it or are harassed based on publicly linkable complaints data that is related to the open dataset, then there may be negative perception about the public nature of this data.

Step 2: Evaluate the Benefits Associated with Releasing the Dataset

List some of the foreseeable benefits of publishing the data fields included in this dataset and identify whether this use typically involves aggregate data or individual records:

> The Building Permits dataset provides accountability for how the city manages its building permits, particularly in dense areas where many individuals and businesses may be impacted by construction. City departments may be interested in tracking how much and what type of development is occurring, and forecasting the potential impacts of the construction activities (such as impeding traffic or pedestrian flows, noise or odor concerns, inspections needed, etc.). Commercial entities and data brokers may use the data to gain business intelligence, such as a competitor's new building activity, housing renovation trends, or contracting and labor values. Individuals may use the information to inform their own building decisions. This data is typically used in both aggregate form (for trend analysis) and individual records (for accountability into specific building activities).

Consider the likely users of this dataset. Who are the ideal users? Check all that apply.

- ✓ Individuals
- □ Community Groups
- □ Journalists
- ✓ Researchers
- ✓ Companies or Private Entities
- ✓ Other Government Agencies or Groups
- □ Other: ___________________________

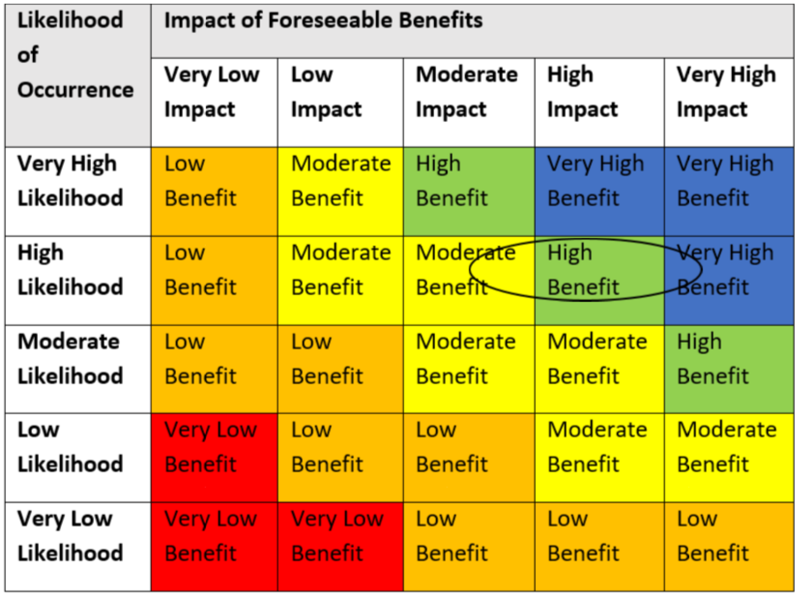
Assess the scope of the foreseeable benefits of publishing the dataset:

Qualitative Value	Quantitative Value	Description
High	8	The dataset will likely have a *compelling and important* utility for individuals, the community, other organizations, or society.

Next, assess the likelihood that the desired benefits of releasing this dataset would occur:

Qualitative Value	Quantitative Value	Description
High	8	The benefit is *highly likely* to occur.

Combining your rating of the foreseeable benefits of the dataset with the likelihood that these benefits will occur, assess the overall benefit of this dataset:



Likelihood of Occurrence	Impact of Foreseeable Benefits				
	Very Low Impact	Low Impact	Moderate Impact	High Impact	Very High Impact
Very High Likelihood	Low Benefit	Moderate Benefit	High Benefit	Very High Benefit	Very High Benefit
High Likelihood	Low Benefit	Moderate Benefit	Moderate Benefit	[✓ High Benefit]	Very High Benefit
Moderate Likelihood	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit	High Benefit
Low Likelihood	Very Low Benefit	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit
Very Low Likelihood	Very Low Benefit	Very Low Benefit	Low Benefit	Low Benefit	Low Benefit

Step 3: Evaluate the Risks Associated with Releasing the Dataset

Consider the foreseeable privacy risks of this dataset:

- *Re-identification (and false re-identification) impacts on individuals*: Building owners and contractors are explicitly identified in this dataset, but the property location and details could be combined with other data to identify or reveal information about other occupants. The precise location and nature of the work being done, along with its estimated value, can expose individuals and their properties to harassment, squatters, or appliance and materials theft during and after the building activity. Data on the value, complexity, and timeline of the work being done to the property may lead to owners or contractors being targeted for scams or unwanted ads and solicitations. Descriptions of the interiors of personal homes may also lead to psychological concern or anxieties for some individuals, particularly if they reveal sensitive information (the location of home entrances, safes, or security systems; the presence of nurseries; assisted living or disability accommodations; etc.).
- *Re-identification (and false re-identification) impacts on the organization:* As Building Permits include explicitly personal data, re-identification attacks are unnecessary. There may be some concern that the city's dataset is consolidating and highlighting properties and individuals that would make particularly good targets for financial or property crimes, which could make it a target for lawsuits.
- *Data quality and equity impacts*: This dataset's connection to the Construction and Inspections department's trackers and other workflows suggests that inaccuracies regarding particular properties are likely to be caught and corrected – how long such inaccuracies persist in the open dataset are unclear. Permitting trends by neighborhood and links to inspection reports/fines/related data may reveal information about development and gentrification trends, the price of services in different parts of the city, or particularly divisive building projects. If inaccurate data about a particular location were used to target a person or property for financial or property crimes, then that would be a significant adverse impact (just as it would be if someone were targeted on the basis of accurate data).
- *Public trust impacts:* If any financial or property crimes, unfair competition activities, fraud, insurance rate hikes, or other generally negative impacts on individuals or communities could be tied back to the availability of the open dataset, then this could significantly impair public trust. The precision of the location data, explicit identification of individual owners and contractors, and granular detail about the nature of the work, inspections, and fines heightens the risk – while individuals might expect some of this information to be made public, they may be surprised that all of it is linked and searchable. Other occupants of relevant properties could also be surprised to find the details and value of their home or work environments publicized on a city website. Concern about such data being made public may chill building activities by vulnerable populations.

Consider who could use this information improperly or in an unintended manner (including to re-identify individuals in the dataset). Check all that apply.

✓ General public (individuals who might combine this data with other public information)
✓ Re-identification expert (a computer scientist skilled in de-identification)
✓ Insiders (a municipal employee or contractor with background information about the dataset)
✓ Information brokers (an organization that systematically collects and combines identified and de-identified information, often for sale or reuse internally)
✓ "Nosy neighbors" (someone with personal knowledge of an individual in the dataset who can identify that individual based on the prior knowledge)
□ Other: N/A______________________________________

Assess the scope of the foreseeable privacy risks of publishing the dataset:

Qualitative Value	Quantitative Value	Description
High	8	The dataset will likely have a *severe or catastrophic* adverse impact on individuals, the community, other organizations, or society.

Next, assess the likelihood that the foreseeable privacy risks of releasing this dataset would occur:

Qualitative Value	Quantitative Value	Description
High	8	The risk is *highly likely* to occur.

Combining your rating of the foreseeable risks of the dataset with the likelihood that these risks will occur, assess the overall risk of this dataset:

Likelihood of Occurrence	Impact of Foreseeable Risks				
	Very Low Impact	Low Impact	Moderate Impact	High Impact	Very High Impact
Very High Likelihood	Low Risk	Moderate Risk	High Risk	Very High Risk	Very High Risk
High Likelihood	Low Risk	Moderate Risk	Moderate Risk	[✓ High Risk]	Very High Risk
Moderate Likelihood	Low Risk	Low Risk	Moderate Risk	Moderate Risk	High Risk
Low Likelihood	Very Low Risk	Low Risk	Low Risk	Moderate Risk	Moderate Risk
Very Low Likelihood	Very Low Risk	Very Low Risk	Low Risk	Low Risk	Low Risk

Step 4: Weigh the Benefits against the Risks of Releasing the Dataset

Step 4A: Combine the overall scores from the benefit and risk analyses to determine the appropriate solution for how to treat the dataset.

Benefits	Risks				
	Very Low Risk	Low Risk	Moderate Risk	High Risk	Very High Risk
Very High Benefit	Open	Open	Limit Access	Additional Screening	Additional Screening
High Benefit	Open	Limit Access	Limit Access	Additional Screening	Additional Screening
Moderate Benefit	Limit Access	Limit Access	Additional Screening	Additional Screening	Do Not Publish
Low Benefit	Limit Access	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish
Very Low Benefit	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish	Do Not Publish

Additional Screening: Releasing this dataset presents high privacy risks and the benefits could outweigh the potential privacy risks, or releasing this dataset presents privacy risk and the potential benefits do not outweigh the potential privacy risks. In order to reduce the privacy risk, formal application and oversight mechanisms should be considered (such as a disclosure review board, data use agreements, or a secure data enclave). The City would assess appropriate risk mitigation measures and countervailing interests in Steps 4B & 5.

Steps 4B & 5: At this point, the City would determine whether appropriate technical or legal/administrative controls could lower the privacy risk further; reevaluate the dataset; and identify any countervailing factors in favor of publication. The City would take into account operational budgets, desired outcomes of the dataset, and the overall utility as a few of the factors when deciding the appropriate steps to take.

Dataset: **Sold Fleet Equipment**

Step 1: Evaluate the Information the Dataset Contains

Consider the following categories of information:

- *Direct Identifiers:* There are no data that directly identify individuals in the Sold Fleet Equipment dataset.
- *Indirect Identifiers:* Indirect identifiers in this dataset might include the sale price, sale date, auctioneer ("Sold_by"), and the year/make/model of the vehicle sold. While this information does not directly relate to an individual, it could possibly be combined with other information (such as user names published on an online auction website, or a friend/family member/coworker's personal knowledge) to identify the purchaser of a sold fleet vehicle.
- *Non-Identifiable Information:* The equipment ID number, general vehicle description, and the department that previously owned the department are information about the vehicle when it was owned by the city, and would not reasonably identify an individual buyer.
- *Sensitive Attributes:* If an individual could be linked to a particular sold fleet vehicle, the sale price may reveal information about their financial condition, which some may consider sensitive. At the same time, the underlying vehicle sales would have taken place at public or online auctions, a context where the sales price of an item and who is bidding on it may not be as sensitive.
- *Spatial Data and Other Information that Is Difficult to De-identify:* The Sold Fleet Equipment data is structured, with no spatial information or freeform entries.

Consider how linkable the information in this dataset is to other datasets:

- *Do any of the dataset's direct or indirect identifiers currently appear in other readily accessible open datasets?* Information on Seattle's fleet equipment may be captured in the Seattle open Active Fleet Complement or Current Fleet Surplus/Auction List datasets before they appear in the Sold Fleet Equipment dataset, but this information is not otherwise generally available in other Seattle open datasets. The same or similar information is available on the listed auctioneers' websites and sites like eBay.
- *How often is the dataset updated?* Monthly. At the time of this report, it had been last updated Nov. 17, 2017.
- *How often is the information in this dataset requested by public records?* Unknown to FPF, but as of this report there were over 99,400 views and 1,831 downloads of the Sold Fleet Equipment dataset.

Consider how the information in this dataset was obtained:

- *In what context was this data collected? Is this data collected under a regulatory regime? Are there any conditions, such as a privacy policy or contractual term, attached to the data? If the personal information in this dataset collected directly from the individual or from a third party?* This data was collected by the city from its own departments and vendors to record the sales of city-owned vehicles (excluding Seattle City Light). The contractual terms are unknown to FPF, but are unlikely to restrict the publication of this data. Nor does the dataset publish personally identifiable information, so privacy-related regulatory restrictions are also unlikely.
- *Would there be a reasonable expectation of privacy in the context of the data collection?* The information in the Sold Fleet Equipment dataset was collected from public or online auctions and describing formerly-public vehicles, so there are generally no expectations of privacy attached. Individuals may have somewhat higher expectations of privacy if the vehicles were purchased at purely online auctions, where the use of pseudonyms or other privacy-protective measures may be more available than in-person auctions.
- *Was the collection of the information in this dataset controversial? Was any of the information in this dataset collected by surveillance technologies?* The collection of this data is not controversial, and was not collected by surveillance technology.
- *Has this dataset been checked for accuracy? Is there a mechanism for individuals to have information about themselves in this dataset corrected or deleted?* It is unclear whether the dataset has been checked for accuracy, and there does not appear to be a mechanism for individuals to request information to be corrected or deleted, other than contacting the dataset owner (Seattle Finance and Administrative Services – Fleet Management).
- *Is there a concern that releasing this data may lead to public backlash or negative perceptions?* If the information were or could be used, in part, to identify an individual purchaser of a fleet vehicle and the amount paid, that could potentially create negative public impacts.

Step 2: Evaluate the Benefits Associated with Releasing the Dataset

List some of the foreseeable benefits of publishing the data fields included in this dataset and identify whether this use typically involves aggregate data or individual records:

The Sold Fleet Equipment dataset provides accountability for how the city manages its fleet inventory and recoups value from sold equipment, which may be of interest to taxpayers, open government groups, and local media. City departments may be interested in tracking how their surplus equipment is disposed of, or the data could help detect or deter fraudulent sales or activity around surplus fleet inventory. The dataset's historic sales data could also help individuals and local businesses inform their purchasing habits. These uses typically rely on individual records.

Consider the likely users of this dataset. Who are the ideal users? Check all that apply.

✓ Individuals
□ Community Groups
✓ Journalists
□ Researchers
✓ Companies or Private Entities
✓ Other Government Agencies or Groups
□ Other: ___________________________

Assess the scope of the foreseeable benefits of publishing the dataset:

Qualitative Value	Quantitative Value	Description
Moderate	5	The dataset will likely have a *clear* utility for individuals, the community, other organizations, or society. While the utility is clear, it is not as urgent as a "high" value.

Next, assess the likelihood that the desired benefits of releasing this dataset would occur:

Qualitative Value	Quantitative Value	Description
High	8	The benefit is *highly likely* to occur.

Combining your rating of the foreseeable benefits of the dataset with the likelihood that these benefits will occur, assess the overall benefit of this dataset:

Likelihood of Occurrence	Impact of Foreseeable Benefits				
	Very Low Impact	Low Impact	Moderate Impact	High Impact	Very High Impact
Very High Likelihood	Low Benefit	Moderate Benefit	High Benefit	Very High Benefit	Very High Benefit
High Likelihood	Low Benefit	Moderate Benefit	[✓ Moderate Benefit]	High Benefit	Very High Benefit
Moderate Likelihood	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit	High Benefit
Low Likelihood	Very Low Benefit	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit
Very Low Likelihood	Very Low Benefit	Very Low Benefit	Low Benefit	Low Benefit	Low Benefit

Step 3: Evaluate the Risks Associated with Releasing the Dataset

Consider the foreseeable privacy risks of this dataset:

- *Re-identification (and false re-identification) impacts on individuals*: If an individual could be identified as the purchaser of a sold fleet vehicle, additional information may allow them to be located or raise the risk of vehicle theft or harassment. However, this would require significant additional information and effort, and is an extreme possibility that is not very likely to occur. Given that the vehicle would have been purchased at auction, the revelation of more sales information is also less likely to be damaging than in other contexts.
- *Re-identification (and false re-identification) impacts on the organization:* If information in the Sold Fleet Dataset were used to re-identify an individual, there could be a chance of reputational damage to the city. Data about the city’s sold fleet equipment does not typically carry confidentiality concerns or legal liability, or depend on individuals volunteering information to the city.
- *Data quality and equity impacts*: This dataset does not contain information about that would typically cast a negative light on individuals or groups, and inaccurate or incomplete data about the sold fleet vehicles is unlikely to significantly affect individuals. This is historic data updated only once per month, so it is easily corrected and it is unlike that individuals, businesses, or city departments are highly dependent on this dataset being perfectly accurate to accomplish their goals. Perturbing the existing data fields for the purpose of limiting the overall risk of re-identification could significantly impact the utility of this dataset, however. For example, masking the auctioneer data could lead to incorrect evaluations of the city’s relationships with particular vendors; perturbing sales price or vehicle information could obscure whether the city received adequate value for the sold vehicles).
- *Public trust impacts*: If an individual were re-identified in part through this data, that individual would likely be surprised to find that their purchase information was made public. It is possible that some individuals could cease purchasing city-owned equipment at auction for that reason; however, this seems highly unlikely.

Consider who could use this information improperly or in an unintended manner (including to re-identify individuals in the dataset). Check all that apply:

✓ General public
□ Re-identification expert
✓ Insiders
✓ Information brokers
✓ “Nosy neighbors”
□ Other: ____________________________________

Assess the scope of the foreseeable privacy risks of publishing the dataset:

Qualitative Value	Quantitative Value	Description
Low	2	The dataset will likely have a *limited* adverse impact on individuals, the community, other organizations, or society.

Next, assess the likelihood that the foreseeable privacy risks of releasing this dataset would occur:

Qualitative Value	Quantitative Value	Description
Low	2	The risk is *unlikely* to occur.

Combining your rating of the foreseeable risks of the dataset with the likelihood that these risks will occur, assess the overall risk of this dataset:

Likelihood of Occurrence	Impact of Foreseeable Risks				
	Very Low Impact	Low Impact	Moderate Impact	High Impact	Very High Impact
Very High Likelihood	Low Risk	Moderate Risk	High Risk	Very High Risk	Very High Risk
High Likelihood	Low Risk	Moderate Risk	Moderate Risk	High Risk	Very High Risk
Moderate Likelihood	Low Risk	Low Risk	Moderate Risk	Moderate Risk	High Risk
Low Likelihood	Very Low Risk	[✓ Low Risk]	Low Risk	Moderate Risk	Moderate Risk
Very Low Likelihood	Very Low Risk	Very Low Risk	Low Risk	Low Risk	Low Risk

Step 4: Weigh the Benefits against the Risks of Releasing the Dataset

Step 4A: Combine the overall scores from the benefit and risk analyses to determine the appropriate solution for how to treat the dataset.

Benefit	Risks				
	Very Low Risk	Low Risk	Moderate Risk	High Risk	Very High Risk
Very High Benefit	Open	Open	Limit Access	Additional Screening	Additional Screening
High Benefit	Open	Limit Access	Limit Access	Additional Screening	Additional Screening
Moderate Benefit	Limit Access	Limit Access	Additional Screening	Additional Screening	Do Not Publish
Low Benefit	Limit Access	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish
Very Low Benefit	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish	Do Not Publish

***Limit Access*:** Releasing this data presents moderate to very low privacy risks and the potential benefits of the dataset outweigh the potential privacy risks. In order to reduce the privacy risk, limit access to the dataset (such as by attaching contractual/Terms of Service terms to the dataset prohibiting re-identification attempts). The City would assess appropriate risk mitigation measures and countervailing interests in Steps 4B & 5.

Steps 4B & 5: At this point, the City would determine whether appropriate technical or legal/administrative controls could lower the privacy risk further; reevaluate the dataset; and identify any countervailing factors in favor of publication. The City would take into account operational budgets, desired outcomes of the dataset, and the overall utility as a few of the factors when deciding the appropriate steps to take.

Dataset: **Seattle Communities Online Inventory**

Step 1: Evaluate the Information the Dataset Contains

Consider the following categories of information:

- ○ *Direct Identifiers:* The Email Contact column sometimes includes individual names (e.g., Thomas.Whittemore@Seattle.gov).
- ○ *Indirect Identifiers:* The URLs of the community organizations often point to websites that include photos and names of individuals who are affiliated with the group or perhaps live in a represented neighborhood. This data could be combined or used separately to single out or identify individual Seattleites. The names and category descriptions of the groups could also shine a light on an individual's activities.
- ○ *Non-Identifiable Information:* Whether a group is community owned and operated or commercial; the type of online tool being used; and the neighborhood, region, and district information are all non-personally identifiable information in this context.
- ○ *Sensitive Attributes:* Some of the names and category descriptions could be considered sensitive (e.g., parenting, education, affordable groups).
- ○ *Spatial Data and Other Information that Is Difficult to De-identify:* Neighborhood, district, and region are spatial data, although they represent wide geographic zones already. Given nature of this data, it may be collected in unstructured formats.

Consider how linkable the information in this dataset is to other datasets:

- ○ *Do any of the dataset's direct or indirect identifiers currently appear in other readily accessible open datasets?* Information on Seattle's community groups may also appear on King County's open data portal, sometimes with more precise locations listed than in the Seattle dataset. The community organizations' own URLs also clearly point to more detailed information on the groups' location, activities, and membership.
- ○ *How often is the dataset updated?* The dataset was last updated Oct. 6, 2016. It is not clear how often it is regularly updated.
- ○ *How often is the information in this dataset requested by public records?* Unknown to FPF, but as of this report there were over 29,500 views and 5,585 downloads of the Seattle Communities Online Inventory dataset.

Consider how the information in this dataset was obtained:

- ○ *In what context was this data collected? Is this data collected under a regulatory regime? Are there any conditions, such as a privacy policy or contractual term, attached to the data? If the personal information in this dataset collected directly from the individual or from a third party?* This data was provided by individuals or organizations who publicized their presence online with the intent of attracting and engaging with Seattle community members. Individuals are also encouraged to add new sites to the list via http://www.seattle.gov/communitiesonline/addform.htm/jmzd-2qjz. This data was not collected under a regulatory regime, and there are unlikely any confidentiality conditions attached.
- ○ *Would there be a reasonable expectation of privacy in the context of the data collection?* There is no reasonable expectation of privacy in the data presented, which is voluntarily provided and about public-facing organizations rather than individuals.
- ○ *Was the collection of the information in this dataset controversial? Was any of the information in this dataset collected by surveillance technologies?* The collection of this data by the city is not controversial, and the data was not collected by surveillance technology.
- ○ *Has this dataset been checked for accuracy? Is there a mechanism for individuals to have information about themselves in this dataset corrected or deleted?* It does not appear that the dataset has been checked for accuracy recently – for example, one listing is for the community organization "facebook," in the listed neighborhood of "twitter." There does not appear to be a mechanism for individuals to request information to be corrected or deleted, other than contacting the dataset owner (Seattle IT).
- ○ *Is there a concern that releasing this data may lead to public backlash or negative perceptions?* If information on the linked community organization sites includes identifying information from unwitting individuals, then the primary backlash would likely be against the site that actually published the information, although there could be some concern that the open data portal amplified the information.

Step 2: Evaluate the Benefits Associated with Releasing the Dataset

List some of the foreseeable benefits of publishing the data fields included in this dataset and identify whether this use typically involves aggregate data or individual records:

The Seattle Communities Online Inventory provides insight and connections for Seattleites interested in engaging more with their local communities. New residents may use this dataset to learn more about their neighborhoods, existing residents may use them to identify community councils, watch groups, or meetings they would like to participate in. Individuals with affinities of all kinds can use this inventory to find others who share their interests – merchants' associations, parenting groups, bike enthusiasts, arts enthusiasts, Filipino heritage, park usage, etc. City departments can use this information to better engage and communicate with local communities. This data is primarily useful in individual record format.

Consider the likely users of this dataset. Who are the ideal users? Check all that apply.

✓ Individuals
✓ Community Groups
□ Journalists
□ Researchers
✓ Companies or Private Entities
✓ Other Government Agencies or Groups
□ Other: ___________________________

Assess the scope of the foreseeable benefits of publishing the dataset:

Qualitative Value	Quantitative Value	Description
Very High	10	The dataset will likely have a *multiple compelling and important* utilities for individuals, the community, other organizations, or society.

Next, assess the likelihood that the desired benefits of releasing this dataset would occur:

Qualitative Value	Quantitative Value	Description
High	8	The benefit is *highly likely* to occur.

Combining your rating of the foreseeable benefits of the dataset with the likelihood that these benefits will occur, assess the overall benefit of this dataset:

Likelihood of Occurrence	Impact of Foreseeable Benefits				
	Very Low Impact	Low Impact	Moderate Impact	High Impact	Very High Impact
Very High Likelihood	Low Benefit	Moderate Benefit	High Benefit	Very High Benefit	Very High Benefit
High Likelihood	Low Benefit	Moderate Benefit	Moderate Benefit	High Benefit	[✓ Very High Benefit]
Moderate Likelihood	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit	High Benefit
Low Likelihood	Very Low Benefit	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit
Very Low Likelihood	Very Low Benefit	Very Low Benefit	Low Benefit	Low Benefit	Low Benefit

Step 3: Evaluate the Risks Associated with Releasing the Dataset

Consider the foreseeable privacy risks of this dataset:

- ○ *Re-identification (and false re-identification) impacts on individuals*: The dataset includes links that reference various community activities and events. It is possible that someone could visit the sites and meetings happening in that neighborhood and learn new information about the groups' members – or learn some information that would make the locals more susceptible to crime, surveillance, or abuse. It largely depends on how much information is revealed on each linked website. Contact emails for particular organizations that reflect individuals' names may identify an individual, but as that information was intentionally made public it is not a *re*-identification risk.
- ○ *Re-identification (and false re-identification) impacts on the organization:* As the City's interaction with this data is fairly limited – posting the inventory without much obvious curation or solicitation of information – and this data is from organizations whose purpose is to engage the Seattle public, re-identification of individuals arising from this dataset is unlikely to harm city operations or create liability.
- ○ *Data quality and equity impacts*: The dataset contains URLs and email addresses, but the groups that are pointed to sometimes contain information about events and conditions in particular neighborhoods that could be offensive to some. The city does not appear to endorse any of the groups listed, however. Community groups with less robust digital literacy or resources may not be represented in this dataset; to the extent that internal or external stakeholders rely on this inventory for public engagement strategies or input, they may be inadvertently excluding such organizations.
- ○ *Public trust impacts*: This dataset was created largely by and for the Seattle community, and individuals would likely not be surprised to learn that public-facing organizations with digital presences (whether websites, blogs, social media, etc.) could be accessed online. This dataset does not appear to include any non-public information.

Consider who could use this information improperly or in an unintended manner (including to re-identify individuals in the dataset). Check all that apply.

□ General public
□ Re-identification expert
□ Insiders
✓ Information brokers
✓ "Nosy neighbors"
□ Other: ______________________________

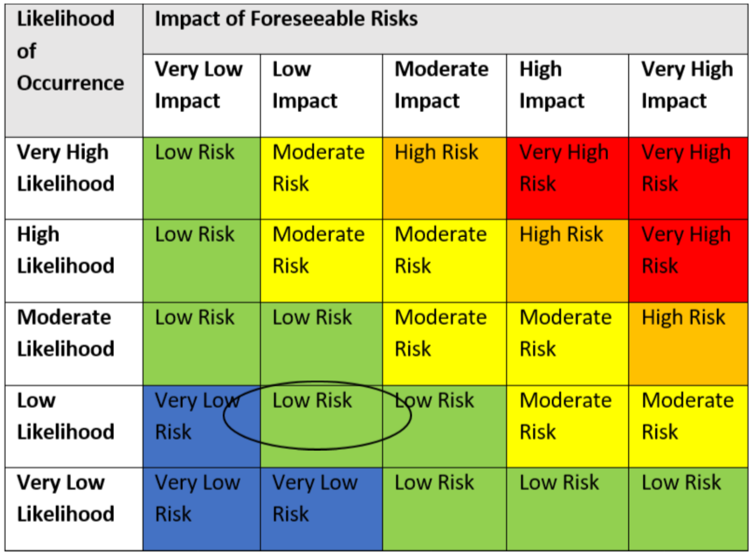
Assess the scope of the foreseeable privacy risks of publishing the dataset:

Qualitative Value	Quantitative Value	Description
Low	2	The dataset will likely have a *limited* adverse impact on individuals, the community, other organizations, or society.

Next, assess the likelihood that the foreseeable privacy risks of releasing this dataset would occur:

Qualitative Value	Quantitative Value	Description
Low	2	The risk is *unlikely* to occur.

Combining your rating of the foreseeable risks of the dataset with the likelihood that these risks will occur, assess the overall risk of this dataset:



Likelihood of Occurrence	Impact of Foreseeable Risks				
	Very Low Impact	Low Impact	Moderate Impact	High Impact	Very High Impact
Very High Likelihood	Low Risk	Moderate Risk	High Risk	Very High Risk	Very High Risk
High Likelihood	Low Risk	Moderate Risk	Moderate Risk	High Risk	Very High Risk
Moderate Likelihood	Low Risk	Low Risk	Moderate Risk	Moderate Risk	High Risk
Low Likelihood	Very Low Risk	[✓ Low Risk]	Low Risk	Moderate Risk	Moderate Risk
Very Low Likelihood	Very Low Risk	Very Low Risk	Low Risk	Low Risk	Low Risk

Step 4: Weigh the Benefits against the Risks of Releasing the Dataset

Step 4A: Combine the overall scores from the benefit and risk analyses to determine the appropriate solution for how to treat the dataset.

Benefits	**Risks**				
	Very Low Risk	**Low Risk**	**Moderate Risk**	**High Risk**	**Very High Risk**
Very High Benefit	Open	Open	Limit Access	Additional Screening	Additional Screening
High Benefit	Open	Limit Access	Limit Access	Additional Screening	Additional Screening
Moderate Benefit	Limit Access	Limit Access	Additional Screening	Additional Screening	Do Not Publish
Low Benefit	Limit Access	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish
Very Low Benefit	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish	Do Not Publish

***Open*: Releasing this dataset to the public presents low or very low privacy risks and the potential benefits of the dataset substantially outweigh the potential privacy risks. The City would continue with appropriate review processes and advance towards publishing this dataset openly.

Dataset: **Road Weather Information Stations**

Step 1: Evaluate the Information the Dataset Contains

Consider the following categories of information:

- *Direct Identifiers:* There are no direct identifiers in this dataset.
- *Indirect Identifiers:* There are no indirect identifiers that in combination could single out an individual.
- *Non-Identifiable Information:* Ambient air temperature, road surface temperature, date/time of collection, and the geolocation of the roads are all non-personally identifiable information.
- *Sensitive Attributes:* There are no sensitive attributes in this dataset.
- *Spatial Data and Other Information that Is Difficult to De-identify:* This data is structured, and spatial data is present related to Seattle road weather stations.

Consider how linkable the information in this dataset is to other datasets:

- *Do any of the dataset's direct or indirect identifiers currently appear in other readily accessible open datasets?* Information on Seattle's road weather conditions may be captured indirectly via Seattle's local weather reports; however, this granular information is not generally available in other local open datasets.
- *How often is the dataset updated?* The dataset was last updated Dec. 19, 2017. The data collected by the sensors are averaged into temperature readings that are recorded by the station every minute. The dataset is updated every 15 minutes with new data.
- *How often is the information in this dataset requested by public records?* Unknown to FPF, but as of this report there were over 191,000 views and 16,500 downloads of the Road Weather Information Systems dataset.

Consider how the information in this dataset was obtained:

- *In what context was this data collected? Is this data collected under a regulatory regime? Are there any conditions, such as a privacy policy or contractual term, attached to the data? If the personal information in this dataset collected directly from the individual or from a third party?* This data was collected by the city from its own departments and vendors to record road conditions within the Seattle city limits. Contractual terms are unknown to FPF, but are unlikely to restrict the publication of this data. Furthermore, the dataset does not publish personally identifiable information, so regulatory restrictions are also unlikely.

- *Would there be a reasonable expectation of privacy in the context of the data collection?* There is no reasonable expectation of privacy in the data presented, which is about atmospheric conditions and public roadways.
- *Was the collection of the information in this dataset controversial? Was any of the information in this dataset collected by surveillance technologies?* The collection of this data by the city is not controversial, and the temperature sensor data was not collected by surveillance technology.
- *Has this dataset been checked for accuracy? Is there a mechanism for individuals to have information about themselves in this dataset corrected or deleted?* It is unclear whether this dataset has been checked for accuracy. There is not information about individuals to be corrected or deleted.
- *Is there a concern that releasing this data may lead to public backlash or negative perceptions?* There is not a concern that data may lead to public backlash relating to the sharing of this data. If road conditions or sensor data are in conflict or their deployment concerns citizens in other ways, the amplification of this program via the open data portal may have some impact.

Step 2: Evaluate the Benefits Associated with Releasing the Dataset

List some of the foreseeable benefits of publishing the data fields included in this dataset and identify whether this use typically involves aggregate data or individual records:

The Road Weather Information Stations dataset provides accountability for how the city tracks and monitors road conditions. This data could be used by city departments to advise their staff and the public about hazards, or to better route city services (such as construction, snow plows, transit, etc.) depending on weather and road conditions. Businesses may also rely on this data for similar reasons (such as taxi or ridesharing drivers, mapping companies, or others). Historic data from this program could also improve how public and private entities route traffic during inclement weather. This data is useful typically in aggregate forms.

Consider the likely users of this dataset. Who are the ideal users? Check all that apply.

□ Individuals
□ Community Groups
□ Journalists
□ Researchers
✓ Companies or Private Entities
✓ Other Government Agencies or Groups
□ Other: ___________________________

Assess the scope of the foreseeable benefits of publishing the dataset:

Qualitative Value	Quantitative Value	Description
High	8	The dataset will likely have a *compelling and important* utility for individuals, the community, other organizations, or society.

Next, assess the likelihood that the desired benefits of releasing this dataset would occur:

Qualitative Value	Quantitative Value	Description
Very High	10	The benefit is *almost certain* to occur.

Combining your rating of the foreseeable benefits of the dataset with the likelihood that these benefits will occur, assess the overall benefit of this dataset:

Likelihood of Occurrence	Impact of Foreseeable Benefits				
	Very Low Impact	Low Impact	Moderate Impact	High Impact	Very High Impact
Very High Likelihood	Low Benefit	Moderate Benefit	High Benefit	[✓ Very High Benefit]	Very High Benefit
High Likelihood	Low Benefit	Moderate Benefit	Moderate Benefit	High Benefit	Very High Benefit
Moderate Likelihood	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit	High Benefit
Low Likelihood	Very Low Benefit	Low Benefit	Low Benefit	Moderate Benefit	Moderate Benefit
Very Low Likelihood	Very Low Benefit	Very Low Benefit	Low Benefit	Low Benefit	Low Benefit

Step 3: Evaluate the Risks Associated with Releasing the Dataset

Consider the foreseeable privacy risks of this dataset:

- ○ *Re-identification (and false re-identification) impacts on individuals*: This data does not relate to individuals, and poses no risk of re-identification.
- ○ *Re-identification (and false re-identification) impacts on the organization:* This data does not relate to individuals, and poses no risk of re-identification.
- ○ *Data quality and equity impacts*: This dataset does not contain information about individuals or groups. If the sensors are spaced inequitably throughout the city, such that some populations are receiving disproportionate benefits from this data, then there may be some fairness concerns. If sensors are producing inaccurate information that individuals or organizations are relying on, then there may be negative impacts on productivity or safety as well.
- ○ *Public trust impacts:* This dataset does not include any non-public information that requires public trust to be collected. Other factors around the deployment of these sensors, however, could be amplified by the data’s presence in the open data portal (such as if the sensors were deployed inequitably throughout the city or if the sensors were inaccurate).

Consider who could use this information improperly or in an unintended manner (including to re-identify individuals in the dataset). Check all that apply.

□ General public
□ Re-identification expert
□ Insiders
□ Information brokers
□ “Nosy neighbors”
✓ Other: N/A

Assess the scope of the foreseeable privacy risks of publishing the dataset:

Qualitative Value	Quantitative Value	Description
Low	2	The dataset will likely have a *limited* adverse impact on individuals, the community, other organizations, or society.

Next, assess the likelihood that the foreseeable privacy risks of releasing this dataset would occur:

Qualitative Value	Quantitative Value	Description
Very Low	1	The risk is *highly unlikely* to occur.

Combining your rating of the foreseeable risks of the dataset with the likelihood that these risks will occur, assess the overall risk of this dataset:

Likelihood of Occurrence	Impact of Foreseeable Risks				
	Very Low Impact	Low Impact	Moderate Impact	High Impact	Very High Impact
Very High Likelihood	Low Risk	Moderate Risk	High Risk	Very High Risk	Very High Risk
High Likelihood	Low Risk	Moderate Risk	Moderate Risk	High Risk	Very High Risk
Moderate Likelihood	Low Risk	Low Risk	Moderate Risk	Moderate Risk	High Risk
Low Likelihood	Very Low Risk	Low Risk	Low Risk	Moderate Risk	Moderate Risk
Very Low Likelihood	Very Low Risk	[✓ Very Low Risk]	Low Risk	Low Risk	Low Risk

Step 4: Weigh the Benefits against the Risks of Releasing the Dataset

Step 4A: Combine the overall scores from the benefit and risk analyses to determine the appropriate solution for how to treat the dataset.

Benefits	**Risks**				
	Very Low Risk	**Low Risk**	**Moderate Risk**	**High Risk**	**Very High Risk**
Very High Benefit	[✓ Open]	Open	Limit Access	Additional Screening	Additional Screening
High Benefit	Open	Limit Access	Limit Access	Additional Screening	Additional Screening
Moderate Benefit	Limit Access	Limit Access	Additional Screening	Additional Screening	Do Not Publish
Low Benefit	Limit Access	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish
Very Low Benefit	Additional Screening	Additional Screening	Do Not Publish	Do Not Publish	Do Not Publish

***Open*:** Releasing this dataset to the public presents low or very low privacy risks and the potential benefits of the dataset substantially outweigh the potential privacy risks. The City would continue with appropriate review processes and advance towards publishing this dataset openly.

FUTURE OF PRIVACY FORUM

1400 EYE STREET NW | SUITE 450 | WASHINGTON, DC 20005 · **FPF.ORG**