



Open.Data.Policy.Lab

# The Emergence of a Third Wave of Open Data

How To Accelerate the Re-Use of Data for Public Interest Purposes While Ensuring Data Rights and Community Flourishing

Stefaan G. Verhulst, Andrew Young, Andrew J. Zahuranec,  
Susan Ariel Aaronson, Ania Calderon, and Matt Gee

October 2020



GOVLAB



# Table of Contents

<b>Table of Contents</b>	<b>2</b>
<b>The Past, Present, and Future of Open Data</b>	<b>3</b>
<b>The Emergence of a Third Wave of Open Data</b>	<b>7</b>
I.    Publishing with Purpose	9
II.   Fostering Partnerships and Data Collaboration	11
III.  Advancing Open Data at the Subnational Level	14
IV.  Prioritizing Data Responsibility and Data Rights	15
<b>Riding the Third Wave: The Way Forward</b>	<b>18</b>
I.    Fostering and Distributing Institutional Data Capacity	19
II.   Articulating Value and Building an Impact Evidence Base	20
III.  Creating New Data Intermediaries	20
IV.  Establishing Governance Frameworks and Seeking Regulatory Clarity	21
V.    Creating the Technical Infrastructure for Reuse	23
VI.   Fostering Public Data Competence	24
VII.  Track, Monitor, and Clarify Decision and Data Provenance	25
VIII. Creating and Empowering (Chief) Data Stewards:	25
<b>Accelerating the Re-Use of Data for Public Interest Purposes while Ensuring Data Rights and Community Flourishing</b>	<b>27</b>
<b>Summer of Open Data: The Complete List of Panels and Conversations</b>	<b>28</b>

Cover Photo by [Uriel SC](#) on [Unsplash](#)

*Introduction:*

# **The Past, Present, and Future of Open Data**



**W**e live, as we are so often told, in an era of data abundance. Yet we also live during a period marked by tremendous inequities and asymmetries when it comes to data access. These contradictory phenomena represent one of the great paradoxes of our times, reflective of more general socio-economic and political trends. Even as ever greater amounts of data are generated and stored, the capability to actually access and re-use this data to spur positive social change remains stunted. This discrepancy occurs despite the success of the open data movement, which built on earlier Freedom of Information (FOI) legislation in pushing at the boundaries of transparency and accessibility. Around the world, open data has played an [important role](#) in improving government accountability and service delivery, empowering citizens to make better decisions, creating economic opportunity, and solving big public problems. But recent developments—including the struggles to use data effectively to address COVID-19—show these successes are not enough. In April 2020, almost 500 data practitioners and organizations signed The GovLab’s [Call for Action](#) urging the development of data infrastructure and an ecosystem capable of tackling the pandemic and other dynamic threats.

There are many drivers of data asymmetries, both existing and emerging. First, one key challenge is that much—possibly the majority—of generated data today resides in the private sector, hidden away in silos collected, controlled, and often monetized by companies and other entities. New models for [collaborating](#) and accessing private sector data are emerging to break these silos, but it is imperative that we further build on them.

In addition, despite its undeniable accomplishments, the open data movement has thus far been limited by an overwhelming focus on the national and supranational levels. Much data resides at the subnational, local level, and attention there could considerably open access.

Finally, recent controversies and early failures to consider citizen's concerns and their data rights have done little to boost the public's trust in open data or the re-use of data. Though there is now greater awareness of the need for [data responsibility](#) and respecting data rights to address this deficit, it needs to be central and incorporated into projects in a systematic fashion to answer public questions about accountability and security.

In this piece, we argue these (and many other) limitations necessitate additional improvements to accelerate the re-use of data to unleash the public good potential of the digital era. We reflect on the potential of what we call the “Third Wave<sup>1</sup> of Open Data.” We discuss both what we see as the emerging elements of this wave and the actions several players have taken that may contribute to its realization. In doing so, we propose actions for how policy makers and data practitioners might address the limitations of previous waves and outline a set of actions that could accelerate data re-use and collaboration.

**“[T]he emergent Third Wave adopts a more purpose-driven approach; it seeks not simply to open data for the sake of opening, but to focus on impactful re-use.”**

As mentioned above, the First Wave relied on FOI laws and used regulation and legislation to unlock data upon specific requests. The Second Wave, marked by the arrival of the open source and Web 2.0 era, relied primarily on open government data (though some private sector data was also included) and sought to make

---

<sup>1</sup> This piece uses the metaphor of “waves” to capture how open data could build on its previous successes and failures. While this concept is fundamental to the piece, the authors note that, in the time since they began developing this paper, the term “wave” has become synonymous with the increasing surge of COVID-19 cases. Though this paper refers to the pandemic and the need to use data to address it, this similarity in language is unintentional. We extend our sympathies to all those affected by the pandemic and are committed to advance the [responsible re-use of data](#) to counteract its worst effects.

data [open by default](#). It often opened datasets without a clear understanding or mapping of how they could best be used. In contrast to its predecessors, the emergent Third Wave adopts a more purpose-driven approach; it seeks not simply to open data for the sake of opening, but to focus on impactful re-use, especially through inter-sectoral collaborations and partnerships. The Third Wave pays at least as much attention to the demand as to the supply side of the data equation; and the way its use impacts the public-at-large. It is concerned not simply with data itself but with the broader technical, social, political, and economic context within which data is produced and consumed.

This paper elaborates on these and other features of the Third Wave. Specifically, we synthesize lessons learned from the [Summer of Open Data](#), a three-month series of conversations and consultations co-organized by the [Open Data Policy Lab](#) (at [The GovLab](#)), [Brighthive](#), [Digital Trade and Governance Hub](#), [Open Data Charter](#), and the [Open Data Institute](#) that tapped into insights from 29 global open data leaders and experts (see full list of panelists and conversations [here](#)). Over the course of ten panels, we sought to jointly discover how to build a collaborative, purposeful, and responsible future of open data. In part A, we highlight four key attributes that appear emergent in the Third Wave of Open Data. In part B, we outline eight steps that could accelerate and amplify these elements emergent in the Third Wave for the public's benefit. Part C concludes with a brief reflection on future priorities.

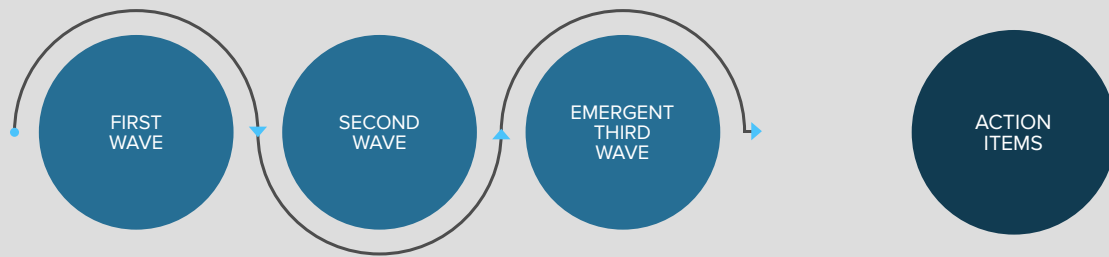


# SUMMER OF OPEN DATA

ACCELERATING DATA COLLABORATION

*Part A*

# The Emergence of a Third Wave of Open Data



Concept:	Freedom of Information	Open Government Data	Re-use of Public and Private Data	<ul style="list-style-type: none"> <li>▶ Foster institutional data capacity</li> <li>▶ Increase Data Liquidity</li> </ul>
Value Proposition:	Transparency	Transparency and Problem Solving	<ul style="list-style-type: none"> <li>▶ Evidence based policy making</li> <li>▶ Innovation and Entrepreneurship</li> </ul>	<ul style="list-style-type: none"> <li>▶ Articulate value and build an evidence base of the impact of data re-use for public interest purposes</li> </ul>
Approach:	Data on Request (Right to Know)	Open by Default (Right to Share)	Publish with Purpose	<ul style="list-style-type: none"> <li>▶ Create new data intermediaries</li> <li>▶ Invest in a Data Demand Identification and Assessment Methodology</li> <li>▶ Increase accessibility and findability of high-value datasets</li> </ul>
Focus:	Pull Focused	Push Focused	Partnerships (Data Collaboratives)	<ul style="list-style-type: none"> <li>▶ Establish governance frameworks and seek regulatory clarity</li> <li>▶ Document and experiment with the number of operational models that are "fit for purpose"</li> <li>▶ Develop common data sharing agreements and new licensing regimes</li> </ul>
Geographical Emphasis	National	International and National	<ul style="list-style-type: none"> <li>▶ Sub-national and Local</li> <li>▶ Purpose-driven Cross-border Data Flow</li> </ul>	<ul style="list-style-type: none"> <li>▶ Create the technical infrastructure for re-use at the local level</li> <li>▶ Invest in Subnational Capacity, Guidance, Legal Frameworks and Best Practice</li> </ul>
Audience/ Demand:	<ul style="list-style-type: none"> <li>▶ Journalists</li> <li>▶ Lawyers and Activists</li> </ul>	<ul style="list-style-type: none"> <li>▶ Civic technologists and "Data Geeks"</li> <li>▶ Government agencies</li> <li>▶ Corporations, technology start-ups</li> </ul>	<ul style="list-style-type: none"> <li>▶ Community-based organizations</li> <li>▶ NGOs, Human Rights Organizations, and Social Justice Advocates</li> <li>▶ Academia and Research Institutions</li> <li>▶ Small businesses and startups</li> <li>▶ Government</li> </ul>	<ul style="list-style-type: none"> <li>▶ Bolster public competence</li> <li>▶ Develop a data demand assessment and segmentation methodologies</li> <li>▶ Strengthen accountability mechanisms to support rights-based data re-use.</li> </ul>
Risks and Policies:	Secrecy and obfuscation	Privacy — Mosaic effect, demographically identifiable information (DII)	Data Responsibility and Data Rights Framework	<ul style="list-style-type: none"> <li>▶ Track, monitor, and clarify decision and data provenance</li> <li>▶ Invest in Data Responsibility Frameworks and Technologies</li> <li>▶ Engage the public, especially vulnerable communities, to document public perceptions and opinions on data re-use</li> </ul>
Institutional Responses:	Information Auditors	<ul style="list-style-type: none"> <li>▶ Data Officer</li> <li>▶ Open Data Portals</li> </ul>	<ul style="list-style-type: none"> <li>▶ Chief Data Steward</li> <li>▶ Intermediaries</li> </ul>	Create and empower (chief) data stewards





**W**hat are the attributes of the emergent Third Wave of Open Data, and how are they distinguished from those that characterized earlier waves? In this section, we describe four key elements that appear to be arising in the Third Wave.

As suggested in the summary provided in Table 1, it is important to recognize that each wave builds on previous ones rather than replacing them. If matched properly, the net effect will be cumulative—an overall trend toward greater openness and transparency, and more purpose-driven (re-)use of data to support social transformation.

## **I. Publishing with Purpose**

We expect the Third Wave to pursue a much more purpose-directed approach to data provision than prior waves; it will seek not simply to open data but to do so in a way that focuses on impactful re-use. The Third Wave will pay at least as much attention to the demand as to the supply side of the data equation, and will be concerned not simply with data itself but with the broader technical, social, political and economic context within which data is produced and consumed.

Each wave has redefined the nature of openness and built on earlier assumptions about how to release previously siloed datasets for maximum social impact. In the First Wave, FOI laws sought to open government data based on a doctrine of *Need to Know*. The government had a responsibility to respond to citizen requests for data related to particular projects or topic areas. In the Second Wave, this conception of openness

expanded to include some private-sector data along with government data. Openness in this wave was premised on a doctrine of *Duty to Share*. In contrast to the pull nature of FOI requests, organizations shared data preemptively with a goal of creating public value from previously siloed assets.



The Third Wave will seek to extend these earlier efforts. It will include a wider spectrum of openness and data types. In particular, it will increase the scope of data work to

include private-sector data. Among the main distinguishing features of this wave will be a greater effort to match supply and demand, and to target which datasets are released so as to achieve maximum social impact.

Built around the emerging doctrine of [Publish with Purpose](#), the participants in our Summer of Open Data argued Third Wave begins with an understanding that resources (whether they be financial, technical, or human) are limited for both data suppliers and recipients. It acknowledges, as Chief Digital Officer for British Columbia Jaimie Boyd [argues](#), that organizations cannot just “pump out data for the sake of data and hope that maybe someone will find [it].” Rather, as Jean-Noé Landry of Open North [describes](#), data advocates can “document and reach out to different user communities to understand their specific data needs.”

**“You have to anchor data in the golden use case that aligns everyone’s interests.”**

– Matt Gee  
Co-Founder & CEO of BrightHive

This demand-driven approach calls on practitioners to think critically about what they intend to achieve, to identify what Matt Gee of BrightHive [calls](#) the “golden use case that aligns everyone’s interests.” Theo Blackwell, London’s chief digital officer, for instance, has highlighted the importance of well-defined and formulated questions to enable effective, and purposeful, data re-use. “The framing of a data question is quite the art,” he argues. It can

make the difference between successfully addressing a major challenge, or failing to do so.

In short, maximizing the value of open data, practitioners need to identify problems or opportunities that lack requisite data, engage actors and communities representing the demand for that data, and formulate clear and effective questions to guide data sharing and re-use.

## II. Fostering Partnerships and Data Collaboration

In the First Wave, the establishment of FOI laws, regulations and institutions made national and local government records available on request to an audience (largely)



Photo by [Jeremy Bishop](#) on [Unsplash](#)

composed of journalists, lawyers, and activists. The Second Wave, enabled by the advent of open source and the web 2.0 era, called upon governments to make their data open by default to civic technologists, government agencies, and corporations.

Because of the very nature of open data initiatives, they often have attracted mostly specialized professionals—lawyers, good government watchdogs, journalists, civic technologists, and data scientists within large corporations. Such specialization has the undeniable benefit of channeling efforts toward individuals and groups with a baseline understanding of the related issues. However, it means open data projects have had limited impact and often failed to engage the diversity of actors who could create public value from data made available.

**“[T]he Third Wave will seek to expand the circle of those involved in open data projects and enable more direct collaboration between data holders and data users.”**

In our conception, the Third Wave will seek to expand the circle of those involved in open data projects and enable more direct collaboration between data holders and data users. These partnerships can be forged with community-based organizations, NGOs, small businesses, local government, and other actors who have a broad-based awareness of conditions on the ground and can help translate data and data insights into meaningful change. It can involve adopting what Ania Calderon, Executive Director of the Open Data Charter, [describes](#) as a [feminist approach](#) to data production and use that centers communities that have traditionally not held power.

Reconceptualizing the notion of what constitutes a “data scientist” is at the heart of expanding the circle. As the World Resources Institute’s Amen Ra [has argued](#), data scientists too often are seen (and see themselves) as apart from the social and governance structures of a local community. Emerging evidence suggests the Third Wave will seek to embed data scientists within broader processes and forms of knowledge. For example, it could connect data scientists with local decision-makers and subject-matter experts who can contextualize data-derived insights in ways reminiscent of [medical anthropologists’](#) community-based contextualization work. As part of this process, data scientists can connect to citizens and citizen groups, who play a similarly vital role in embedding data knowledge. Denise Riedl, Chief Innovation Officer for South Bend, Indiana and a Fellow with the Benton Institute [argues](#):

“[T]echnology is reshaping cities, their services, and their public spaces. It's key for residents to feel that they both understand and have a voice in that change. We have to raise the bar on civic engagement when it comes to data and technology innovations in our communities.”

[Data collaboratives](#) that bring both public and private-sector to bear for public problem solving will also play a key role in the Third Wave. Stephen Chacha of the Tanzania Data Lab [notes](#), for instance, the capacity of data collaboratives to improve lives, support the data demand that exists on the ground, and inform policy for various community actors that work with the lab. It can inform efforts to improve basic processes as well as achieve larger sustainable development goals or inform COVID-19 response.



These arrangements provide for several new operational and governance models for inter-sectoral engagement, models in which actors work together to identify and use privately held data that can contribute social value. For example, the Swedish nonprofit Flowminder used data from telecommunications operators in [Nepal](#) to facilitate aid deliveries after the 2015 earthquake there. In the United States, clinical data holders can share their data with the [Yale University Open Data Access](#) program which, in turn, shares the information with researchers seeking to develop new drugs and treatments. BrightHive [similarly has led](#) a project with Goodwill and Google.org. The relationship deploys Google staff and resources to increase Goodwill's ability to use data to better assess the success of Goodwill organizations. The GovLab's [Data Collaboratives Explorer](#) contains a list of over 200 other examples of data collaboratives, and provides real-world [examples of different types](#) of collaboratives.

Private-sector data assets can be integral in filling information gaps. London's Chief Digital Officer Theo Blackwell [describes](#) the capacity well when he notes "open [government] data alone does not cover the entire universe of what we do," and that merging open government data and data held by the private sector has become increasingly essential for helping the city thrive.

### **III. Advancing Open Data at the Subnational Level**

Previous waves demonstrated how the potential of open data could be harnessed at the national and international levels. The emergent conception of the Third Wave places a greater emphasis on building open data capacity and meeting open data demand at the subnational level. Data held by the public sector and other institutions in cities, municipalities, states, and provinces are, by definition, more targeted and narrower in scope than data made available at the national or supranational level. Subnational open data is more likely to align with the direct and immediate needs of citizens, and the actors representing the demand for that data are likely to be more proximate to the people they intend to benefit and more familiar with their needs.

Cities, municipalities, states, and provinces are increasingly [addressing local needs](#) through open data and data collaboration. Yet, despite the indisputable importance (and potential) of subnational open data efforts, local actors also encounter some major

obstacles. The Beeck Center’s Tyler Kleykamp, [highlights](#) the fundamental issue of subnational data capacity, arguing, “Chief data officers are often just one person in city government trying to open data, share data [...] they just lack the capacity to do [open data] in a very thoughtful manner.” Rudi Borrmann, the Deputy Director of OGP Local, expresses similar sentiments. “Capacity building inside the government, that’s a conversation we are seeing today accelerated by COVID. The future is here but the skills and resources aren’t evenly distributed,” he [states](#). This is especially true in smaller towns or rural areas, which often lack the technical and financial capabilities to support even a lean open data program or team.

In addition, all of these factors—and others—combine to create an information paucity at the local level. Often, there simply isn’t enough locally relevant data available (especially of sufficient quality) to justify and initiate open data projects that could result in meaningful change. Moreover, the particularity and regional emphasis of subnational efforts means that they are often seen as limited in scope and impact, and consequently have trouble attracting finance and other forms of support necessary to scale up pilot projects. As Denise Riedl Linn of South Bend, Indiana [describes](#), the small town data ecosystem is still evolving. They look to peer cities and philanthropies such as [Bloomberg’s What Works Cities](#) and the [National Neighborhood Indicators Partnership](#) for support in developing data capacity.

#### IV. Prioritizing Data Responsibility and Data Rights

Lastly, our conversations and research suggest the Third Wave of open data will take a responsibility-by-design approach to open data activities. Throughout our work, data practitioners emphasized the need to better promote fairness, accountability, and transparency across all stages of the data lifecycle to manage risks and maximize value. Importantly, this work involves not just preserving data rights and needs but also measuring benefits against risks. Practitioners acknowledged the societal costs associated with both acting and *not* acting. While open data champions in earlier waves acknowledged risks to personal privacy, much of the focus was narrow and did not fully anticipate risks across the data value chain. Practitioners generally did not take into consideration risks related to, for instance, potential bias in the analysis and use of certain open datasets or how open data initiatives might negatively impact the rights of

people and communities. Privacy is, of course, key in any open data project, but privacy does not exist in a vacuum. There are other risks from and to the data ecosystem that need to be considered and protected alongside (not to the exclusion of) privacy. In addition, there exist power asymmetries in how data is made available that may reinforce existing inequities.

Organizations and countries can sometimes cite privacy or security as a reason not to share data (data hoarding), in the process exacerbating other harms. For instance, a company that refuses to provide functional access to data in the name of privacy may simply be perpetuating its own monopolistic ownership and, in the process, exacerbating existing inequalities—both in terms of access to data but also broader socio-economic disparities that often map onto data inequalities. Lawrence Kay of the Open Data

Institute [writes](#) in *apolitical* how closed systems make collaboration harder, make people poorer, and make innovation more difficult.

**“Privacy is, of course, key in any open data project, but privacy does not exist in a vacuum. There are other risks from and to the data ecosystem that need to be considered and protected alongside (not to the exclusion of) privacy.”**

There can be other serious costs to this data hoarding, especially in crises or urgent situations. A refusal to share data in the name of privacy can threaten lives and livelihoods. Arturo Franco of Mastercard [has described](#) how his organization’s [Principles for Data Responsibility](#) emphasize the need to “put [data] in the public interest and achieve social impact” and give consumers control over how others use their data. Ania Calderon of the Open Data Charter and Swee Leng Harris of the Luminate Group similarly echoes this sentiment. For Ania, it is important to engage people in how and when organizations use their data [because](#) “people might feel comfortable with certain types of data being shared for certain purposes and not others.” Swee Leng, meanwhile, [speaks](#) of the need for organizations to articulate the context in which data was produced so we could have “systems worthy of trust.”

Data-holding institutions are better served by integrating privacy-protective behavior within a more multifaceted data responsibility framework that seeks to identify and act upon both opportunities and risks across the data lifecycle—from data collection through processing, analysis, sharing, and (re-)use. Importantly, the emerging conception of the Third Wave suggests data suppliers can be proactive in assessing the ethical implications of data re-use, and take steps to ensure that external actors do not use data in a way that could harm data subjects, as outlined in the British Government’s upcoming [Data Ethics Framework](#).

**“Data-holding institutions are better served by [...] a more multifaceted data responsibility framework that seeks to identify and act upon both opportunities and risks across the data lifecycle.”**

As discussed in more detail below, the field largely lacks the types of governance frameworks that can ensure this more sophisticated and end-to-end approach to data responsibility. Establishing such guidelines will be an essential step in responsibly unlocking the value of the Third Wave of Open Data.

*Part B*

# **Riding the Third Wave: The Way Forward**



Emerging research and practice, as well as lessons learned from the [Summer of Open Data](#) series make clear a number of concrete steps will be needed to enable the Third Wave. Our discussions with 29 global open data experts working in government, the private sector, and civil society organizations highlighted eight key actions that can be taken to foster a Third Wave built around equitable, impactful public benefits. These are:

## I. Fostering and Distributing Institutional Data Capacity

In previous waves, open data and innovation capacity within institutions were largely relegated to particular teams or units. This consolidation and siloing of data skills and resources limited the impact of open data and minimized its ability to filter into daily institutional operations. Arturo Muent Kunigami of the Inter-American Development Bank referred to this point in [his panel](#), noting the existence within public institutions of both vertical silos (creating barriers between domains, such as transportation and education) and horizontal silos (creating barriers between skills or capacities, such as open data, information technology, and citizen engagement). While Chief Data Stewards and data stewardship teams will be important actors going forward, institutions should take care that they are not disconnected from normal business operations.

Breaking down these horizontal and vertical silos is important to Rudi Borrmann of the Open Government Partnership, who [speaks](#) to the need to “evenly distribute” skills and resources across local governments to enable them to meet public needs. The OECD’s

Barbara Ubaldi similarly [highlights](#) the importance of institutions acting as a “learning organization that has people who are aware they are sitting on data” that could be used to forge new relationships and create public value.

## II. Articulating Value and Building an Impact Evidence Base

Earlier waves of open data were often built on the normative value and importance of providing public access to public information as well as arguments that opening data can create concrete, real-world impacts, such as supporting economic growth. However, the normative arguments for open data risk losing their influence as institutional budgets tighten and political attitudes shift in many parts of the world. As Sage Bionetworks’ John Wilbanks [explains](#), the Second Wave of Open Data too often championed the existence or availability of datasets, rather than “celebrating the number of users of data” and the real-world impacts of their work.

Demonstrating the concrete, tangible value of increased access to and re-use of data, therefore, is increasingly essential. Arturo Muent Kunigami [describes](#) the need to show that open data “is not a ‘nice-to-have,’ but rather, it “actually improves the lives of people, actually improves equity, actually produces economic growth. Natalia Domagala of the UK Cabinet Office similarly [notes](#) that “simple explanations of the long-term value of open data” can help make the case to the “policymakers and decision-makers” that hold the keys to long-term sustainability of open data programs. Demonstrating this value can also, as New York City Open Data Program Manager Zachary Feder [says](#), communicate to the public what types of impact can be created through open data, increasing their recognition and support of such efforts.

## III. Creating New Data Intermediaries

Our emerging conception suggests data intermediaries will be important actors in the Third Wave, helping to lower transaction costs between data suppliers and users. Some intermediaries can facilitate data collaboration by matching supply and demand actors, ensuring that both public and institutional objectives can be achieved in a responsible manner. Jean-Noé Landry of Open North [advocates](#) for the establishment of more and more targeted data intermediaries to foster collaboration between large public

institutions and local community stakeholders. “Data intermediaries have opened up a space for topics that in the first wave of the open data movement, weren’t at the forefront [...] like risk analysis around data release as a trust building measure.” He says, arguing that these intermediaries can “bring more openness to the decision making process” as it relates to open data.

Other intermediaries can provide additional technical expertise to a collaboration by analyzing data from the supply side and passing on actionable insights to users representing the demand. This approach can advance data collaboration without requiring resources to be expended by data providers or data users beyond their capacity. Prominent examples of intermediaries include New Zealand’s [Data Ventures](#), the commercial arm of the country’s statistical agency created to pull and redistribute data; [Open North](#), which facilitates multi-stakeholder partnerships around data usage in Canada; and [BrightHive](#), a co-sponsor of the Summer of Open Data and an organization that leads efforts to link organizations and data. These data intermediaries are continuing to emerge, but their number will need to expand significantly for the Third Wave to achieve its potential.

#### IV. Establishing Governance Frameworks and Seeking Regulatory Clarity

Many data suppliers in the public and private sectors lack fit-for-purpose governance frameworks and face significant legal and regulatory uncertainty. Ania Calderon, Executive Director of the Open Data Charter, for instance, [notes](#) “[o]ur legislation and policies are insufficient to respond to the nuances we are seeing in terms of how data is being used, abused” and has advocated for designing new governance frameworks to promote public trust. Senior Data Scientist for the World Bank Malar Veerappan similarly [argues](#) that governance frameworks are essential both for incentivizing data use and re-use, and for “creat[ing] safeguards that mitigate risks from harmful outcomes.”

**“[O]ur legislation and policies are insufficient to respond to the nuances we are seeing in terms of how data is being used, abused”**

**– Ania Calderon, Executive Director of the Open Data Charter**



Photo by Josh Calabrese on Unsplash

In the public sector, this lack of clarity manifests at the international and national level, but can be especially pronounced at the subnational level. While organizations such as the European Union [have sought](#) to develop strategies for maximizing the effectiveness of public and private data stores, many US states host open data portals but do not have an open data policy driving their use. Christian Troncoso of BSA | The Software Alliance [argues](#) that, in the absence of such policy direction, “those portals tend to be pretty malnourished, not very usable, and not very responsive to the community of users who would otherwise engage with the data being made available.”

Some private sector organizations face similar uncertainty regarding what types of data sharing and collaboration are appropriate and legally sound. The Tanzania Data Lab’s Stephen Chacha [indicates](#) that, for instance, “Telecom operators are waiting to share their data but they aren’t sure what datasets are okay to share and which are not okay to share. And if they end up sharing the ones that are not okay to share [...] what will happen to them?” Meanwhile, in the United States, a recent [MIT survey](#) found that 64% of business executives are reluctant to fully embrace open data as a result of regulatory uncertainty. As it

stands, the legal basis for data collaboratives involving private sector data holders tends to be more ad hoc and bilateral in nature. For these reasons, organizations such as the United States-based [Data Coalition](#) have sprung up to ensure the private sector can engage with and benefit from open data policies, but more work remains to be done.

Despite the existence of some open data licensing regimes, bespoke and unrepeatable [contracts](#), memoranda of understanding, and similar legal instruments drive activity in the space. BSA's Christian Troncoso [suggests](#) that competition and privacy regulators could be empowered to establish an expedited review process to approve proposed data sharing arrangements to help accelerate activity in the space. Stakeholders could also look to establish new data licensing regimes and contract templates that are explicitly aimed at unlocking purposeful cross-sector data collaboration and re-use.

## V. Creating the Technical Infrastructure for Reuse

In many countries, Freedom of Information laws were enshrined long before the proliferation of digital technologies. More recently, and with a few [notable exceptions](#), parties have used relatively common technologies like web forms and email to request and receive desired first-wave public information. The open data portal remains the central technical component enabling the second wave of open data. These portals commingle various institutional datasets and allow users to browse, filter, search, and download data to their machines. Despite clear calls for common data standards and machine-readability, the format, structure and quality of data provided through both first and second wave technological infrastructures can be inconsistent and unpredictable.

The open data portal will likely remain a common piece of technical infrastructure in the Third Wave, but the focus on data collaboration and data responsibility will necessitate new and more sophisticated technological development. As described in The GovLab's [Call for Action](#) to create the infrastructure and ecosystem necessary to leverage data to address emerging threats like COVID-19, creating this technical infrastructure will likely require an intersectoral, multidisciplinary research and development effort. This effort could focus initially on core needs such as privacy-preserving technologies, security technologies, access-control technologies, and, as [highlighted](#) by Patrick McGarry of



data.world, the technical means for increasing the interoperability of disparate data systems.

For the Third Wave to reach its full potential, John Wilbanks also [argues](#) for improvements to technical infrastructure on the demand and use side. This could involve new strategies for institutions to subsidize computing capacity among target users and demographics, especially in fields where the datasets in question are prohibitively large and complex.

## VI. Fostering Public Data Competence

The general public is an important stakeholder in data re-use efforts, whether they are users of open data made publicly accessible, the intended beneficiaries of data collaboratives, or somehow put at risk as data subjects. Minister Audrey Tang of Taiwan [emphasizes](#) the need to encourage more than just data literacy among the public, a

“[I]nstitutions need to foster data competence so anyone can participate fully in data efforts.”

common call for general familiarity with key data science concepts. Rather, she argues, institutions need to foster data competence so anyone can participate fully in data efforts. This approach can help the public contribute directly to a data ecosystem from which they might currently feel disconnected. “We do not want our children to feel that they are merely media literate, that they are only consumers of media, only consumers of data, consumers of digital creative products,” says Tang. “I want them to think that they are producers.”

This data competency is not just a matter of enabling creativity or new productive capacity but to provide the public with a means for disrupting the power asymmetries that persist in the current data and digital era. “We want to think about the contributions they are making and the trade-offs they are making. Once they see themselves as data producers, they are in a position to negotiate.”

## VII. Track, Monitor, and Clarify Decision and Data Provenance

The lack of visibility into both decision and data provenance can limit the ability of actors to identify the optimal intervention points for mitigating data risks and to avoid missed use of potentially impactful data.

[Decision provenance](#) involves identifying key decision points impacting data's collection, processing, sharing, analysis, and (re-)use; and determining which internal and external parties influence those decision points. This distillation can help stakeholders on the supply and demand side of open data and data collaboration pinpoint any gaps and develop strategies for improving decision-making processes.

Relatedly, Luminate principal Swee Leng Harris [advocates](#) for a greater focus in the Third Wave on the social implications of data provenance. She argues, “We need to understand the context in which data was produced and the potential harms that might result from unhelpful or inadvertent use.”

## VIII. Creating and Empowering (Chief) Data Stewards:

Finally, we argue Third Wave data projects will likely emphasize the importance of creating and nurturing the right institutional structures to support impactful re-use of data. Some leading governments have taken steps to establish open data champions and cross-agency open data teams. Data collaboratives initiated in the private sector, on the other hand, tend to be one-off and ad hoc.

**“While the data steward role is relatively new, there is a growing body of evidence demonstrating the impact and value of their work.”**

Our research suggests the Third Wave will likely seek to systematize and build on prior experiences and lessons to reimagine institutional arrangements that can support more data-driven ways of working. One key piece of this institutional shift lies in the creation of new roles and responsibilities. [Data stewards](#), also known as Chief Data Stewards, are a

particularly important emerging actor. These individuals are responsible data leaders within organizations empowered to identify opportunities for data sharing and seek new



ways of creating public value through cross-sector data collaboration. They may take the form of individuals or groups of individuals, but either way, consist of dedicated teams or employees that help initiate and sustain collaboration. Data stewards serve five key roles in their work: 1) partnership and community engagement; 2) internal coordination and staff engagement; 3) data audit, ethics, and assessment of value and risk; 4) dissemination and communication of findings; and 5) nurture data collaboratives to sustainability. While the data steward role is relatively new, there is a [growing body of evidence](#) demonstrating the impact and value of their work.



## *Conclusion*

# Accelerating the Re-Use of Data for Public Interest Purposes while Ensuring Data Rights and Community Flourishing

Over the coming months and weeks, the Open Data Policy Lab and its network of partners will seek to spur these changes needed toward the realization of the Third Wave of Open Data. This effort will involve engagement with public and private partners on open data and data re-use issues. It will also involve developing the expertise and tools that data practitioners need to be collaborative. The Open Data Policy Lab, for instance, is currently developing a course on data stewardship to help leaders in the public and private sector develop data re-use strategies to solve public problems.

As we go about this work, we welcome comments and reactions on this piece. We encourage readers to annotate this post, outlining their thoughts on the findings and recommendations. All feedback will factor into our subsequent research. With your support, we can ride a new wave of open data that is more collaborative, innovative, and demand-driven.

*Appendix:*

# Summer of Open Data: The Complete List of Panels and Conversations



The [Summer of Open Data](#) is a three-month project spearheaded by the [Open Data Policy Lab](#) (an initiative of [The GovLab](#) with support from Microsoft) in partnership with the [Digital Trade & Data Governance Hub](#), [Open Data Institute](#), the [Open Data Charter](#), and [BrightHive](#). From July through September, we spoke with data experts in local and regional governments, national statistical agencies, international bodies, and private companies to advance our understanding of how to establish a vision of open data focused on collaboration, responsibility, and purpose.

Below, we provide all those conversations and their participants. We welcome visitors to the Open Data Policy Lab to use the links provided to read our blogs summarizing key points and to watch videos of the full sessions. We're confident these discussions have major value, not just in improving how we use data but also in ensuring we can achieve a third wave of open data that addresses data gaps and is fueled by enhanced data collaboration.

### **Summer of Open Data:**

**Tuesday, 14 July 2020: [Announcement of the Summer of Open Data](#)** on [Medium](#)

**Wednesday, 22 July 2020: [Our Summer of Open Data: What Are The Contours of The Third Wave of Open Data?](#)** (Kick-Off Panel [Blog](#) and [Video](#))

- ▶ Co-Founder and CEO of BrightHive [Matt Gee](#)
- ▶ Executive Director of the Open Data Charter [Ania Calderon](#)
- ▶ Vice President and Chief Strategy Adviser of the Open Data Institute [Jeni Tennison](#)

The panel focused on the key contours, opportunities and challenges of the **Third Wave of Open Data**. Distinguished by its interest in opening up data silos, the third wave emphasizes the importance of responsibly reusing public and private data held by local governments and businesses through the use of partnerships. It emphasizes the importance of making data accessible to more than just the “usual suspects” — journalists, lawyers, and civic technologists — and additionally seeking ways to make data useful to community-based organizations, NGOs, academics, and small businesses.

**Wednesday, 29 July 2020: [Data Re-Use from Local Government to the Corporate Sector](#)** (Panel 2 [Blog](#) and [Video](#))

- ▶ Head of Economic Policy Research & Insights at LinkedIn [Paul Ko](#)

- ▶ Professor of Economics and International and Public Affairs at Brown University and Founding Director of Research Improving People's Lives [Justine Hastings](#)
- ▶ Chief Innovation Officer for the City of South Bend, Indiana [Denise Linn Riedl](#)

The panel focused on the differences of data re-use across sectors, lessons learned from open data initiatives, and potential innovations for data collaboration. Panelists discussed topics within the broader context of the **Third Wave of Open Data**, a purpose-driven and responsible approach to data collaboration through cross-sector partnerships.

### **Wednesday, 5 August 2020: Keynote Conversation with Taiwan's Audrey Tang**

(Keynote [Blog](#) and [Video](#))

- ▶ Taiwanese Digital Minister [Audrey Tang](#)

In an 45-minute conversation, host Stefaan Verhulst and Audrey spoke on a variety of issues, including Taiwan's response to the COVID-19 pandemic, Taiwan's application of data collaboration, the real and potential value of emerging technologies, the need to engage the public on data use, and ways to develop digital skills.

### **Wednesday, 12 August 2020: Data Responsibility and New Forms of Collaboration**

(Panel 4 [Blog](#) and [Video](#))

- ▶ Open North Executive Director [Jean-Noé Landry](#)
- ▶ Vice President of Data & Insights, Mastercard Center for Inclusive Growth [Arturo Franco](#)
- ▶ OECD Head of Unit of Digital Government and Open Data [Barbara Ubaldi](#)

In a 45-minute conversation, moderator Stefaan Verhulst and the panelists spoke on a variety of issues, including the importance of data governance, new data roles, and effective collaboration across sectors.

### **Wednesday, 19 August 2020: Data Reuse, Service Delivery, and Horizontal Silos** (Panel

5 [Blog](#) and [Video](#))

- ▶ British Columbia Chief Digital Officer [Jaimie Boyd](#)
- ▶ World Bank Senior Data Scientist [Malarvizhi Veerappan](#)
- ▶ Inter-American Development Bank Senior Specialist for Modernization of the State [Arturo Munte Kunigami](#)

In a 45-minute conversation, moderator Stefaan Verhulst and the panelists spoke on a variety of issues, including the dangers of horizontal data silos, diverse models for data collaboration, and the perspective that subnational actors can bring to data reuse.

### **Wednesday, 26 August 2020: Subnational Data, Sustainability, and Skills**

**Development** (Panel 6 [Blog](#) and [Video](#) of Panel 6)

- ▶ Maxar Technologies Director of Sustainable Development Practice [Rhiannan Price](#)
- ▶ Co-Founder Tanzania Data Lab and Africa Philanthropic Foundation [Stephen Chacha](#)
- ▶ London Chief Digital Officer [Theo Blackwell](#)

In a 45-minute conversation, moderator Stefaan Verhulst and the panelists spoke on a variety of issues, including the state of subnational data, developing data collaboratives at local and community level, promoting sustainability of data projects, advocating for open data's value, and developing appropriate data skills.

### **Wednesday, 2 September 2020: The Impact of COVID-19 on States, Localities, and Business** (Panel 7 [Blog](#) and [Video](#))

- ▶ Open Government Partnership Deputy Director of OGP Local [Rudi Borrmann](#)
- ▶ State Chief Data Officers Network Director [Tyler Kleykamp](#)
- ▶ StreetLight Data Vice President of Commercial Development and Privacy [Kara Selke](#)

### **Wednesday 9 September 2020: Focus on Public Communication, Legal Mandates, and Data Ethics** (Panel 8 [Blog](#) and [Video](#))

- ▶ BSA | The Software Alliance Senior Director of Policy [Christian Troncoso](#)
- ▶ New York City Open Data Program Manager [Zachary Feder](#)
- ▶ United Kingdom Department of Digital, Culture, Media and Sport Head of Data Ethics Policy and Open Government [Natalia Domagala](#)

In a 45-minute conversation, moderator Stefaan Verhulst and the panelists spoke on a variety of issues, including the evolution of the open data movement, the importance of legal mandates for directing energy, and the need for transparency and responsible use of data amid the ongoing pandemic.

**Wednesday, 16 September 2020: Incentives for Data Reuse, Frameworks for Collaboration, and Centering Data Responsibility** ([Blog](#) and [Video](#) of Panel 9)

- ▶ Head of Strategic Partnerships, Data Marketplace [Patrick McGarry](#)
- ▶ Sage Bionetworks Chief Commons Officer [John Wilbanks](#)
- ▶ Luminate Group Principal for Data & Digital Rights [Swee Leng Harris](#)
- ▶ European Laboratory for Learning and Intelligent Systems Board Member [Nuria Oliver](#)

In a 45-minute conversation, moderator Stefaan Verhulst and the panelists spoke on a variety of issues, including the strategies for forging cross-sector data collaboration, the pandemic's influence on the data space, and mechanisms for advancing responsible data reuse in the social interest.

**Wednesday, 23 September 2020: Defining the Value Proposition, Building Common Infrastructure, and Avoiding Missed Use** (Panel 10 [Blog](#) and [Video](#))

- ▶ Infinite Campus Head of Learning Science Technologies [Daniel Jarratt](#)
- ▶ National Student Clearinghouse Managing Director of Strategic Initiative [Vanessa Brown](#)
- ▶ Commonwealth of Virginia Workforce Policy Analyst [Felix Shapiro](#)

In a 45-minute conversation, moderator Matt Gee and the panelists spoke on a variety of issues, including incentives for data collaboration and reuse, the need for repurposable legal and technical infrastructure, and avoiding missed use of potentially valuable data.

## AUTHORS

**Stefaan G. Verhulst** is Co-Founder and Chief Research and Development Officer of the Governance Laboratory (GovLab) at NYU where he is building an action-research foundation on how to transform governance using advances in science, data and technology. Verhulst's latest scholarship centers on how technology can improve people's lives and the creation of more effective and collaborative forms of governance. Specifically, he is interested in the perils and promise of collaborative technologies and how to harness the unprecedented volume of information to advance the public good.

**Andrew Young** is the Knowledge Director at The GovLab, where he leads research efforts focusing on the impact of technology on public institutions. Among the grant-funded projects he has directed are a global assessment of the impact of open government data; comparative benchmarking of government innovation efforts against those of other countries; a methodology for leveraging corporate data to benefit the public good; and crafting the experimental design for testing the adoption of technology innovations in federal agencies.

**Andrew J. Zahuranec** is Research Fellow at The GovLab, where he is responsible for studying how advances in science and technology can improve governance. In previous positions at the NATO Parliamentary Assembly and National Governors Association, he worked on issues as far-ranging as election security, the commercial space industry, and the opioid epidemic. He has a Master of Arts in Security Policy Studies from the George Washington University and a bachelor's degree in Political Science and Intelligence from Mercyhurst University.

**Susan Ariel Aaronson** is Research Professor of International Affairs and Director of the Digital Trade and Data Governance Hub. Aaronson conceived of and directs the Hub, which aims to educate policymakers, the press and the public about domestic and international data governance issues from digital trade to public data governance.

**Ania Calderon** is the Executive Director of the Open Data Charter, a collaboration between governments and organizations working to open up data based on a shared set of Principles, whose goal is to embed open data as a central ingredient to achieving better solutions to the most pressing policy challenges of our time. She has previously led the national open data policy in Mexico between 2013-2016, delivering a key presidential mandate on opening up government data in more than 200 public institutions and a network of over 40 cities in Mexico and strengthening open data commitments globally.

**Matt Gee** is co-founder and CEO at BrightHive, a public benefit corporation building data collaboratives that power smarter government and more effective social service delivery. He is also a Senior Research Scientist at the University of Chicago's Center for Data Science and Public Policy. He is the co-founder of the Eric and Wendy Schmidt Data Science for Social Good fellowship, which over the last five years has paired 250 data science fellows with over 85 national, state, and local government organizations and NGOs to build data-driven solutions to social problems.



*This piece was developed from the input and support of all the participants of the Open Data Policy Lab's Summer of Open Data initiative. We'd like to thank all its participants, Arturo Franco, Arturo Muent Kunigami, Barbara Ubaldi, Christian Troncsco, Daniel Jarratt, Denise Linn Riedl, Felix Shapiro, Jaimie Boyd, Jean-Noé Landry, Jeni Tennison, John Wilbanks, Justine Hastings, Kara Selke, Malarvizhi Veerappan, Natalia Domagala, Nuria Oliver, Patrick McGarry, Paul Ko, Rhiannan Price, Rudi Borrmann, Stephen Chacha, Swee Leng Harris, Theo Blackwell, Tyler Kleykamp, Vanessa Brown, and Zachary Feder. We'd also like to thank our colleagues at The GovLab, Digital Trade & Data Governance Hub, the Open Data Charter, BrightHive, and Open Data Institute, including especially Akash Kapur, Danuta Egle, Mary Ann Badavi, Aditi Ramesh, and Michelle Winowatan for their editorial, conceptual, and design contributions. We greatly appreciate their support and the support of all our other colleagues.*