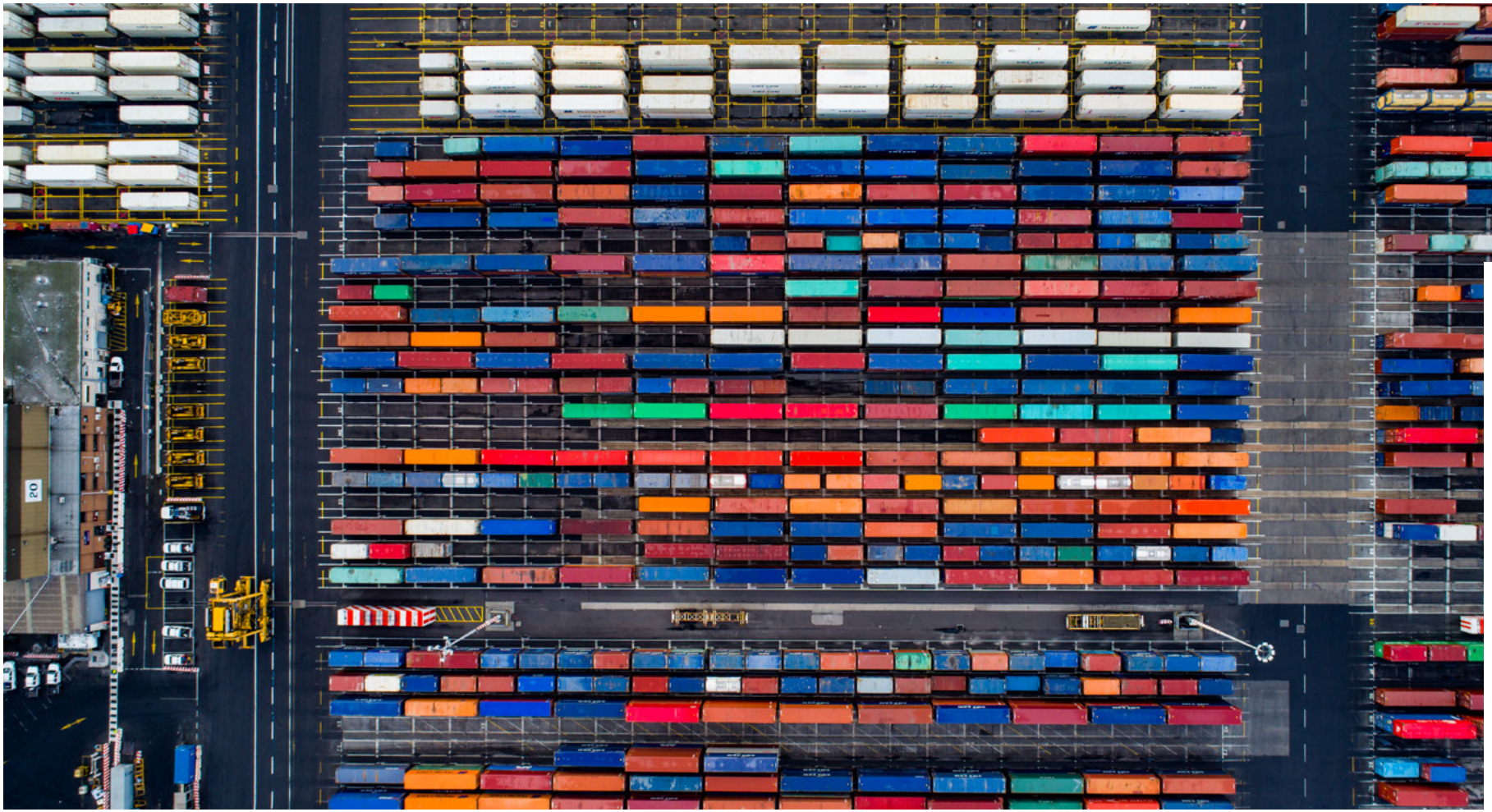
Microsoft

April 2021

# Open Data Campaign:

## Year one in review



# Table of Contents


Imagining What's Possible With Open Data 1

Open Data: Progress and Unrealized Promise 2

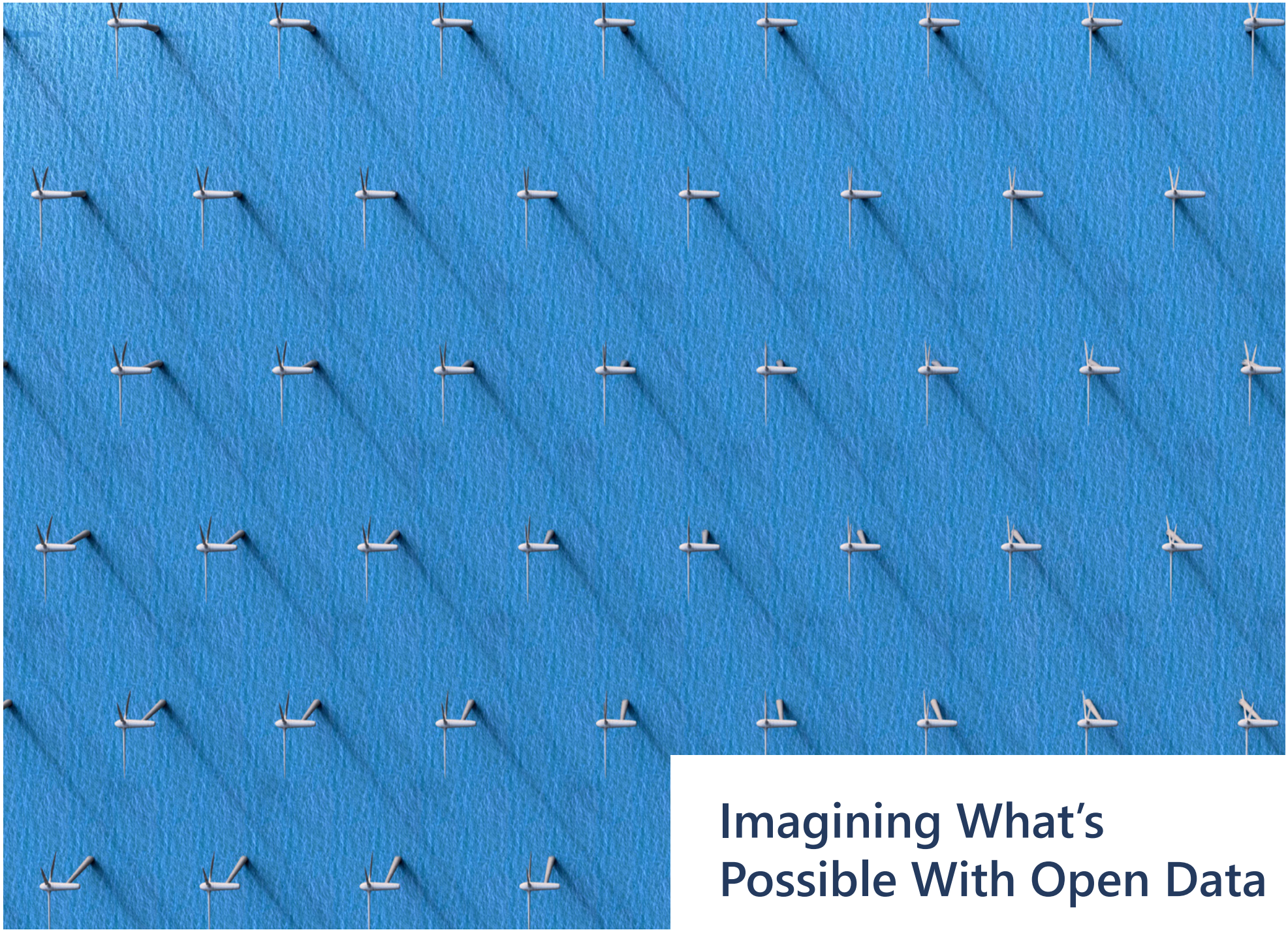
Charting a Principled Path Forward 3

Meet Our Anchor Partners 4

Open Data Campaign at One: Proving Out a New Paradigm 5

Our Focus for 2021 and Beyond 6

Microsoft



# Imagining What's Possible With Open Data

When we announced the Open Data Campaign in April 2020, an integral push to help close the data divide, we knew that launching 20 new data collaborations by 2022 would be an ambitious goal. It needed to be.

The clock is ticking on decisive action on climate change. Intractable inequities—in education, in policing, in incomes—have become more glaring and intolerable. And COVID-19 has revealed vulnerabilities in our healthcare systems we couldn't have imagined.

But we also could not have predicted how a global pandemic would produce a new model for sharing data and taking rapid, informed collective action.

With our anchor partners, the Open Data Institute (ODI) and The Governance Lab (GovLab) at New York University's Tandon School of Engineering, we are shifting to a new paradigm in technology and business: using data to collaborate, not just compete.

We believe open access to data can help us tackle some of the world's most pressing societal challenges. Over the past year, Microsoft helped launch nine data collaborations in **the areas of sustainability, health, and equity and inclusion.** Several more partnerships will follow soon.

These initial collaborations serve as the foundation for our campaign, and they've shown us the wide-ranging and sometimes unexpected benefits of open data. By sharing data concerning COVID-19, climate change, broadband access, criminal justice reform, and beyond, we're seeing that partnership is the key to unlocking innovative solutions across industries, geographies, and focus areas.

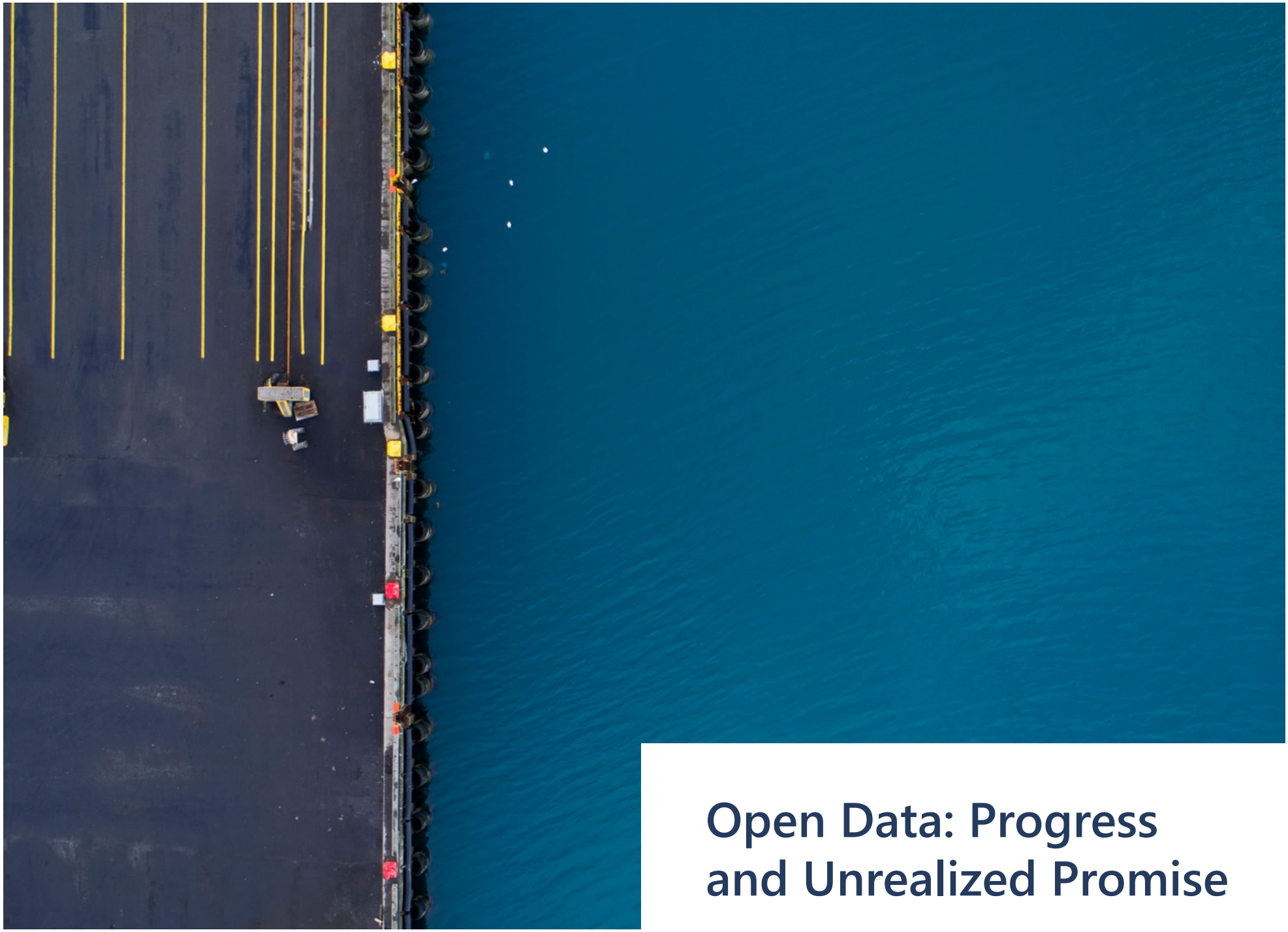
This report recaps the tremendous progress the Open Data Campaign has made over the past year. It provides compelling examples of how bringing diverse datasets together can help researchers, organizations, and governments understand and confront problems that cannot be ignored. And it offers a look at our plans.

Open data can help us do what we imagine, in a future built on collaboration, innovation, and responsibility. We're excited to play our part.

Jennifer Yokoyama

**Vice President and Deputy General Counsel, Intellectual Property Group at Microsoft**

Microsoft



# Open Data: Progress and Unrealized Promise

One year ago, our research found that **fewer than 100 companies** were collecting more than **50 percent of all internet users' interactions,**[1] and **less than 1 percent of companies** had data scientists who could work with their data in a meaningful way.[2] These findings represent a clear data divide, with the COVID-19 pandemic bringing new urgency to closing this gap.

The pandemic also showed just how quickly a viable, effective open data model could emerge and make progress on a societal problem, providing critical insight and otherwise unrealized value.

## Global data sharing during COVID

During the pandemic, various organizations shared countless examples of cross-industry data to:

- Help inform the public on case counts and the trajectory of the disease.
- Provide researchers around the world with direct access to research, including the immune response, genetic information, and novel protein structures, as they worked to develop effective therapies and vaccines.
- Supply updates on vaccine availability and the progress of vaccine rollouts.

Microsoft's AI for Health philanthropic program contributed to this ecosystem through the development of publicly accessible COVID-19 dashboards. Universities likewise released their COVID-19 campus dashboards with data on new and cumulative cases and testing.

Open data around COVID-19 has been both global and local. Once one organization shared digestible forms of data, others quickly followed. But making this open data equitably accessible, at scale, proved to be a new challenge.

## Examples where technology is helping prepare for the next crisis

When we launched the Open Data Campaign, we committed to advancing the tools and technologies to make open data and data sharing easier. We've made important progress on this front. This includes the launch of the first-ever open-source differential privacy platform, SmartNoise.

This technology, pioneered by researchers at Microsoft in collaboration with the OpenDP Initiative led by Harvard University, allows researchers to preserve privacy while fully analyzing datasets. As a part of this effort, we are granting a royalty-free license under Microsoft's differential privacy patents to the world through OpenDP, encouraging widespread use of the platform and allowing anyone to begin using the platform to make datasets widely available to others around the world.

Additionally, Microsoft committed to putting large environmental datasets to work for sustainability and conservation through the Planetary Computer. Conservation depends on very large spatial datasets, especially satellite imagery and climate projection data. This has complicated the task of applying the data to environmental monitoring and sustainability decision-making.

The Planetary Computer hosts key datasets on Azure so that scientists don't have to move all that data around themselves. It provides APIs and computing tools to let scientists work with that data at global scale without having to become experts in parallel computing. This allows conservation practitioners to focus on what's most important: finding answers to critical environmental questions with new applications built on the platform, including forest monitoring, forest-based carbon sequestration, conservation planning, and more.

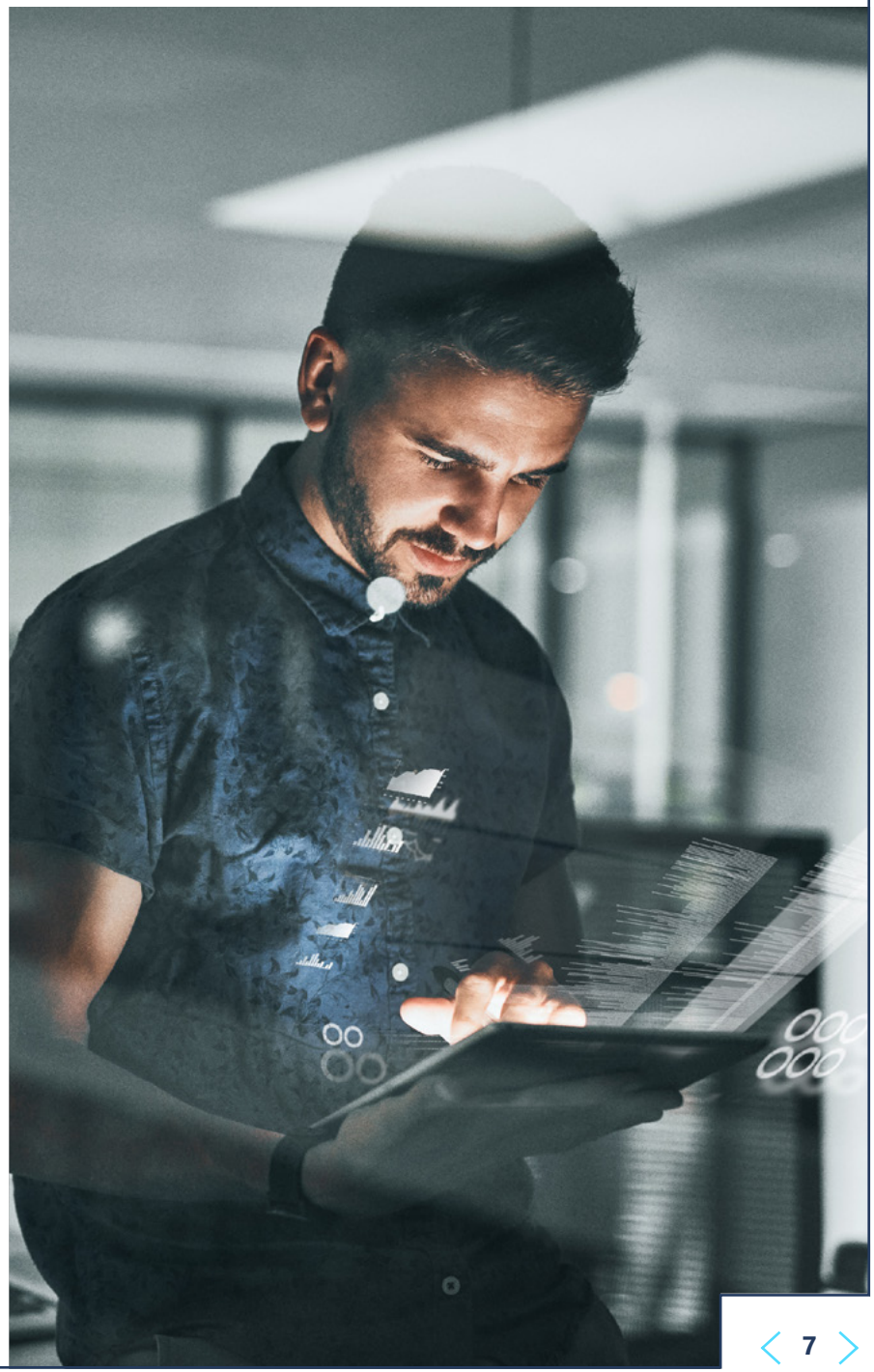
## The next hurdle for open data: capacity building

Countries around the world are looking to harness the tremendous amounts of data being created every day. They understand that unlocking it can lead to economic growth and local economic opportunity through smarter policy and regulation. Governments also see the potential for industries to solve challenges or gain efficiencies by sharing data, and to address complex problems around issues such as sustainability and health research.

At the same time, according to a recent Gartner study,[3] "Executive leaders know that data sharing is a key digital transformation capability, but they lack the 'know-how' to effectively share data at scale and with trust." The study also found that "Organizations that share data externally with their partners generate **three times more measurable economic benefit** than their counterparts that do not."

One of our biggest takeaways from our work this past year is that data collaboration is a spectrum. Varying factors—like the scope of the challenge the data collaboration aims to address, or the trust level of participants, which determines how open the data actually is—will mean different configurations and different goals. But every collaboration can lead to more accessible data, innovative insights, and discoveries.

What's needed is clear: principles that provide a framework for closing the data divide, strategies on scaling data stewardship, added capacity in terms of a qualified data analyst workforce, accessible and scalable tools and technologies, and assistance on advancing policy discussions on open data.



Microsoft

# Charting a Principled Path Forward

**How do we close the data divide and where do we start?** We published five Data Collaboration Principles that guide Microsoft's contributions and commitment to trusted data collaboration. We hope these principles will inform the broader conversation on open data and that others can build on them:

**1) Open:** We will work to make data relevant to important social problems as open as possible, including by contributing open data ourselves.

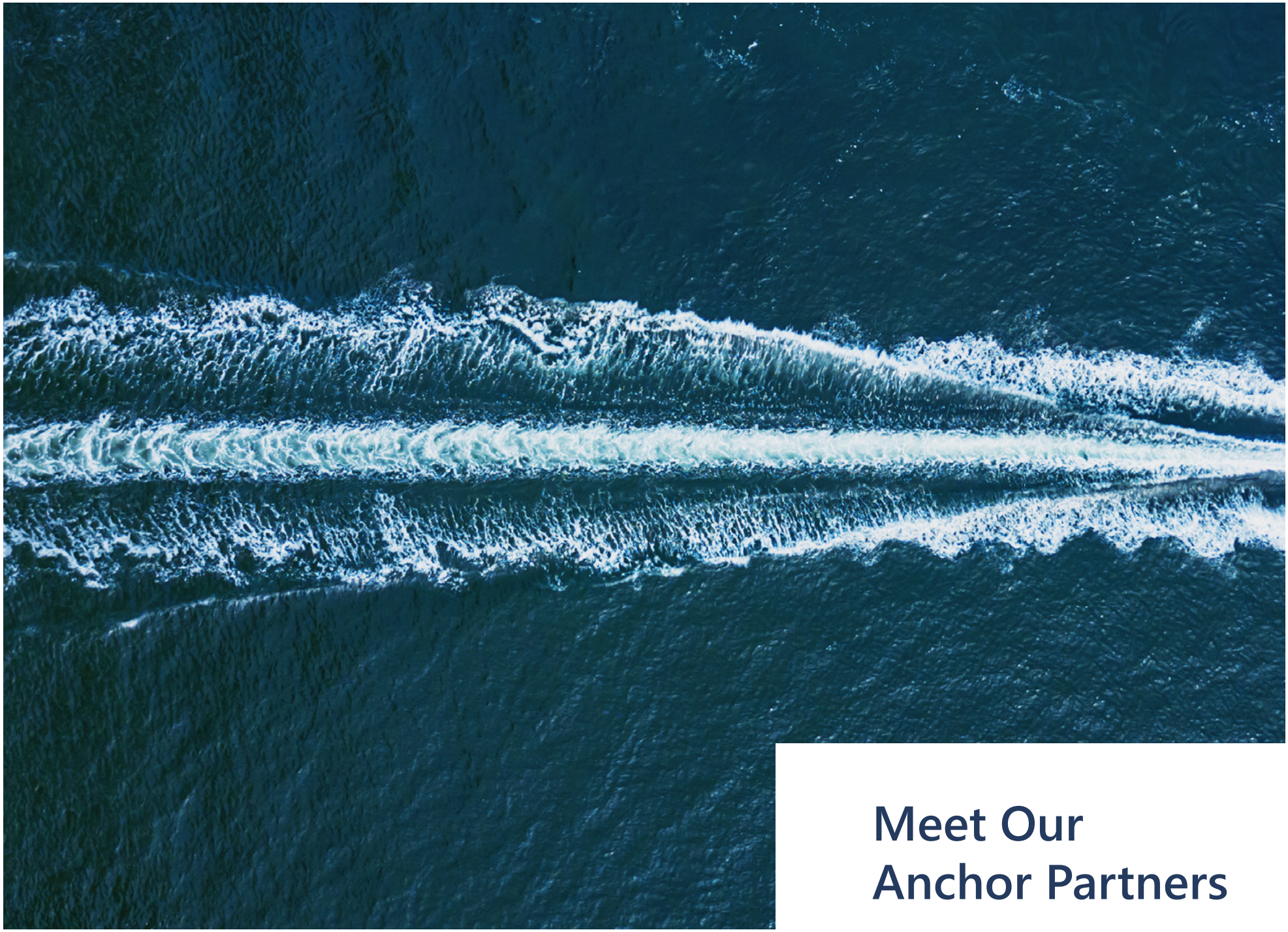
**2) Usable:** We will invest in creating new technologies and tools, governance mechanisms, and policies to make data more usable for everyone.

**3) Empowering:** We will help organizations generate value from their data according to their choices and develop their AI talent to use data effectively and independently.

**4) Secure:** We will employ security controls to ensure data collaboration is operationally secure where it is desired.

**5) Private:** We will help organizations protect individuals' privacy in data-sharing collaborations that involve personally identifiable information.

Microsoft



# Meet Our Anchor Partners

Closing the data divide requires partnership among many other organizations. Microsoft launched the Open Data Campaign with two anchor partners, ODI and GovLab.

The goal of our ODI partnership is to advance the cause of open, trustworthy data sharing and collaboration, so organizations of any size can more easily collaborate around data and realize its benefit. To achieve this, we have a joint agenda to:

**1.** Run an Education Open Data Challenge to better understand and address remote education, including through the use of newly opened Microsoft data.

**2.** Identify, scope, and help launch three data collaborations to tackle issues including climate change.

**3.** Run two peer learning networks that provide data collaborations with guidance, funding, and other support, beginning with a first cohort that explored issues of trust and trustworthiness between participants and other stakeholders.

**4.** Develop case studies that feature open data and data sharing in action.

With GovLab, we supported the launch of a new Open Data Policy Lab. It serves as the premier resource for decision-makers at the local, state, and national levels, as they accelerate the responsible reuse and opening of data, keeping the benefits to society and equitable economic opportunity at the fore.

The Open Data Policy Lab is focused on four areas:

**1. Analysis:** Publishing new research on open data and data reuse movements and policies. The Open Data Policy Lab started this work with a report on *The Third Wave of Open Data*.

**2. Guidance:** Providing resources, training, and tools to help leaders in the public and private sectors develop data reuse strategies, including the launch of a Data Stewardship course and *Third Wave of Open Data Toolkit*.

**3. Community:** Building a community of data stewards through peer engagement and capacity-building via courses, outreach, and events.

**4. Action:** Creating opportunities to collaborate, celebrate, and share in responsible data reuse best practices, to raise awareness of the potential for impact around the world.

Microsoft

# Open Data Campaign at One: Proving Out a New Paradigm

Closing the data divide means putting information and tools within reach of those who are imagining what's possible—a stable climate, more equitable societies, better health for more people. It requires ongoing partnership and access to shared data.

Microsoft has committed to launching 20 data collaborations by 2022 with closing the divide in mind. These projects model how organizations can use open and shared data and combine knowledge to address key societal and business challenges. We are excited to share our progress with summaries of nine collaborative projects, focused on sustainability, health, and equity and inclusion.

Where possible, we opened and shared Microsoft's own data to seed these collaborations. We also contributed other resources, including legal and policy expertise and technology support. These partnerships have been structured to make the learnings and processes developed easy to share—thereby helping others advance their own collaborations.

# Electric Vehicle Charging Infrastructure Pilot

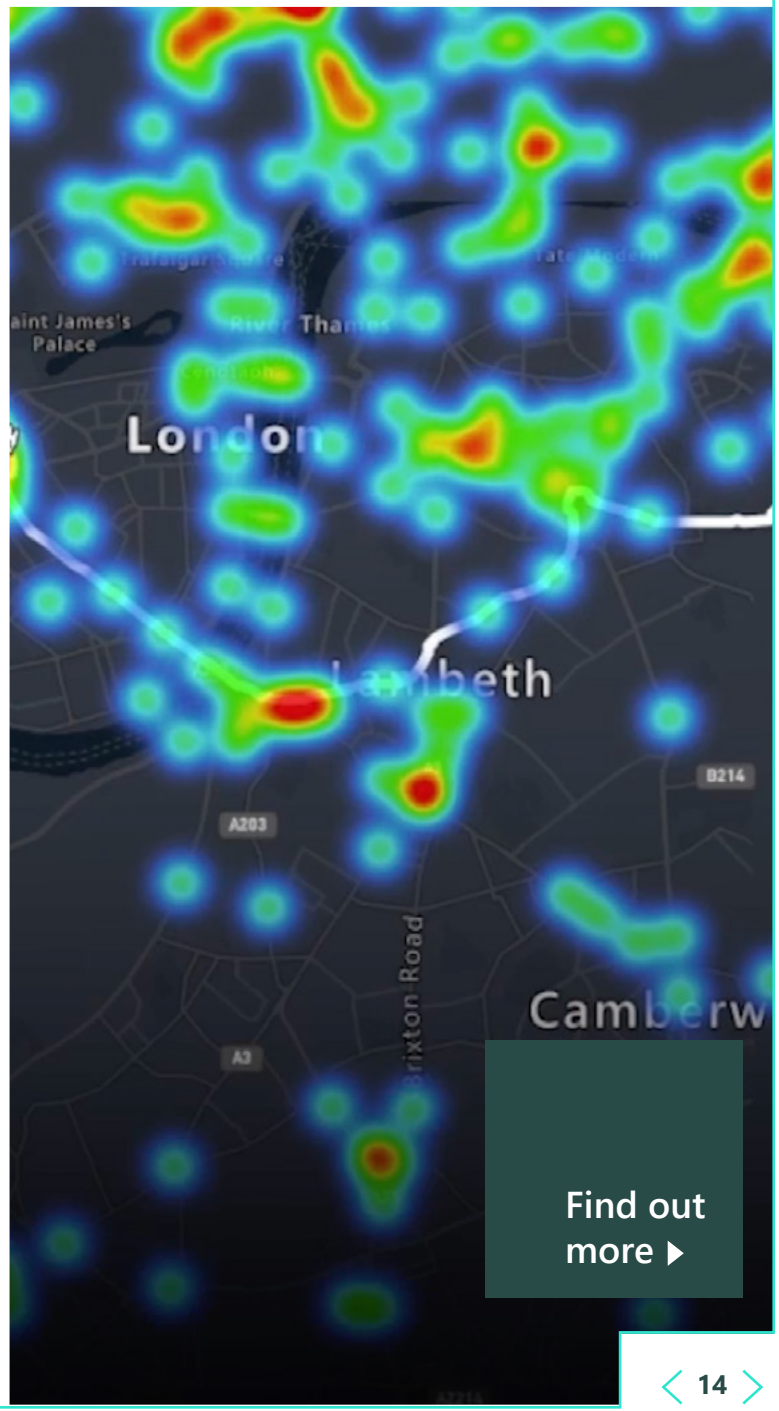
**Imagine:**

Knowing exactly where to build EV infrastructure

**Partners:**

BP, Kainos, London Data Commission, Oliver Wyman, UK Power Networks, Microsoft

The London Data Commission sought to inform decisions on where to build charging stations by integrating multiple data streams. The commission's Electric Vehicle (EV) Charging Infrastructure Pilot mapped and layered public- and private-sector data including EV charging demand, current infrastructure, power capacity, and land availability. Microsoft partner Kainos used the data to develop a dashboard that highlights the optimized placement of EV charging stations. The goal of this ongoing pilot: provide mapping for the Greater London Authority to inform investments that ultimately increase adoption of electric vehicles over time, and reduce carbon emissions.

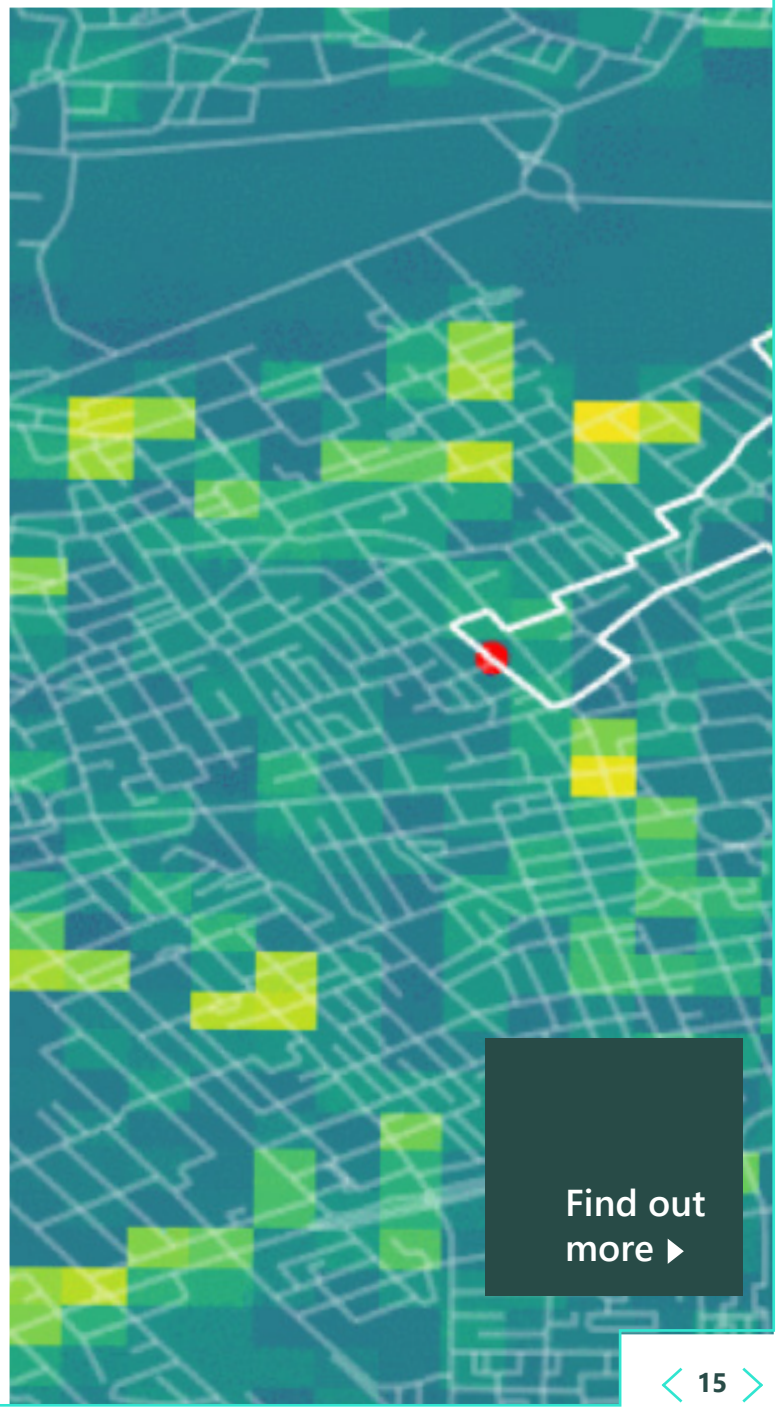


# London Air Quality Project

**Imagine:**
Knowing precisely what will reduce air pollution

**Partners:**
The Alan Turing Institute, Greater London Authority, Kainos, Mayor of London, Transport for London, Waze, Microsoft

The Alan Turing Institute launched the London Air Quality Project to understand and improve air quality over London. The project uses various heterogeneous air quality sensors across the city to estimate and forecast air pollution. It also includes developing machine learning algorithms and data science platforms for better insights using cloud computing supported by Microsoft and The Alan Turing Institute. A dashboard to visualize the insights was developed by Kainos as part of a Microsoft AI for Earth grant. The findings will be used to help inform and evaluate government policy, but this collaboration also had an unintended benefit: The data streams have been repurposed to look at the effects of lockdown easing during COVID-19.



# OS-Climate Initiative

### Imagine:

Being able to see specific financial opportunities and economic impacts of climate change

### Partners:

Allianz, Amazon, Federated Hermes, Goldman Sachs, S&P Global, Linux Foundation, Microsoft

In September 2020, Microsoft joined with Allianz, Amazon, and S&P Global to announce plans to launch a new Open Source Climate Initiative to address one of the most pressing challenges of the 21st century. Led by the Linux Foundation, this initiative enables the investment community to build accurate and reliable economic models around corporate climate-related risk and opportunity. It will leverage high-quality open and shared corporate sustainability data, including Microsoft's own relevant data, which will be shared on the OS-Climate platform supported by this effort.

We are heavily invested in sustainability and recognize the success of this platform relies on the participation and data of many organizations. In February 2021, Goldman Sachs and Federated Hermes became members of OS-Climate, and we expect several new members from banking, institutional investing, technology, and market infrastructure to be announced in the weeks and months ahead.

# Purdue Food and Agricultural Vulnerability Index

**Imagine:**

Mitigating disruptions to food supply chains and identifying potential hotspots to deploy healthcare resources during COVID

**Partners:**

Purdue University, Microsoft

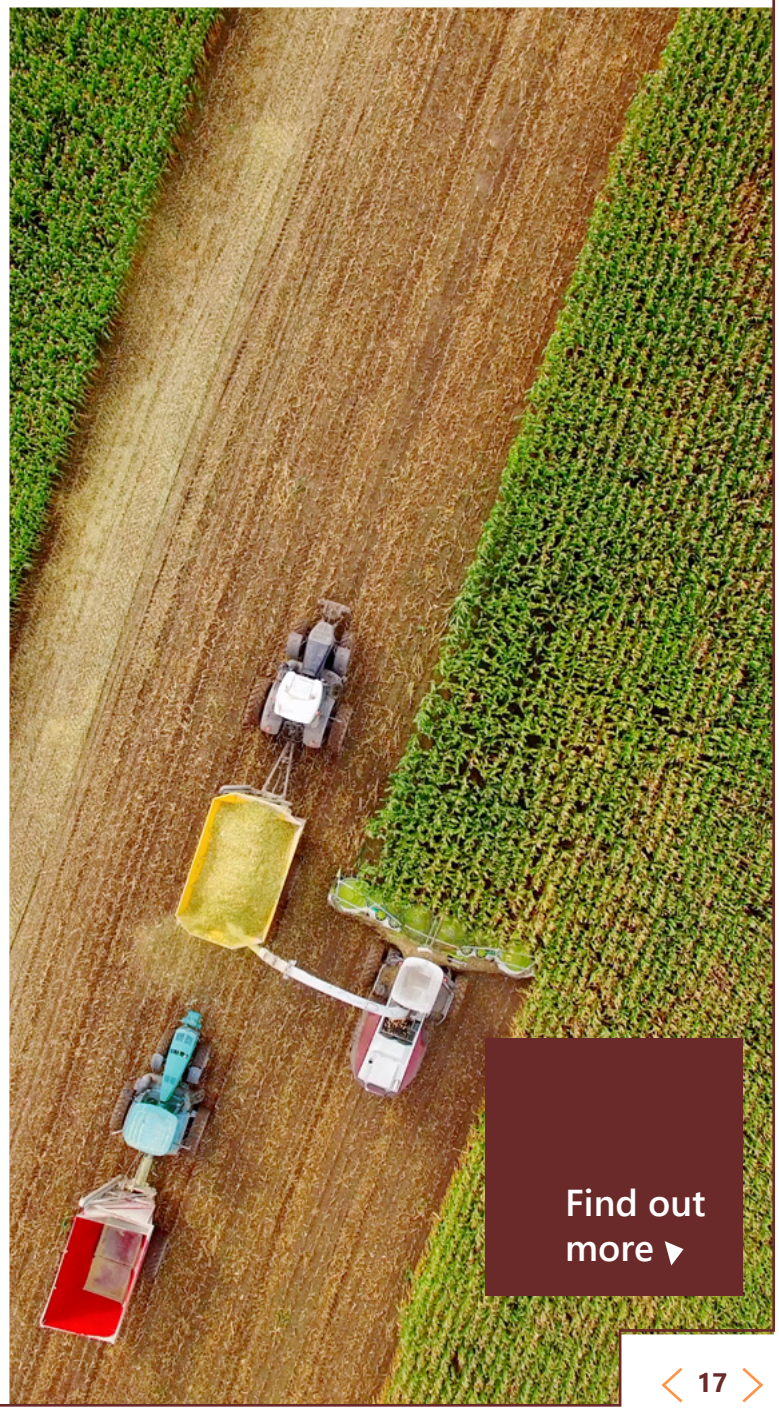
The Purdue Food and Agricultural Vulnerability Index online dashboard, built on top of Microsoft Azure and Power BI platforms, estimates the share of agricultural production at risk from pandemic-related illnesses among the farming workforce. By combining public datasets from Johns Hopkins University of Medicine, the U.S. Department of Agriculture, and the U.S. Census Bureau, users can estimate the risks to production and productivity from COVID-19-related farmer and farm worker illness. Although initially built to identify the vulnerability of agricultural commodity supply due to COVID-19, the dashboard also identifies potential pandemic hotspots, which can be used to decide where to deploy health care resources. In this case, mostly in underserved communities. This collaboration demonstrates that when making data available for one purpose, benefits not initially envisioned are uncovered.

**Additional resources**

COVID-19 United States Cases by County - Johns Hopkins Coronavirus Resource Center (jhu.edu)

County Population Totals: 2010-2019 (census.gov)

USDA/NASS QuickStats Ad-hoc Query Tool



# London Busyness

**Imagine:**
A smooth transition to normal following a pandemic

**Partners:**
The Alan Turing Institute, Greater London Authority, London Data Commission, Microsoft

London Busyness looked at how the city can ensure its populace emerges from the COVID-19 pandemic safely while expediting economic recovery. The Alan Turing Institute and the Greater London Authority, supported by Microsoft and the London Data Commission, mapped London "busyness" or movement around the city as restrictions were relaxed, to monitor how people were responding to the changes. The data insights are allowing for more nuanced planning and a better understanding of the extent to which London is returning to normal.

Find out more ▶

# Education Open Data Challenge

**Imagine:**

More effectively targeting educational resources for remote learners

**Partners:**

Open Data Institute, BroadbandNow, Microsoft

At the time of project launch, according to UNESCO monitoring, 144 countries were implementing nationwide school closures due to the pandemic, and 38 countries were implementing local closures, affecting more than two-thirds of the world's learners.

Together with the Open Data Institute, we initiated an Education Open Data Challenge to look at the impact of the transition to remote learning on young students' education. Microsoft published United States Broadband Usage Percentage Datasets, both at a county level and at a ZIP code level, derived from anonymized data we collect as part of our ongoing work to improve the performance and security of our software and services.

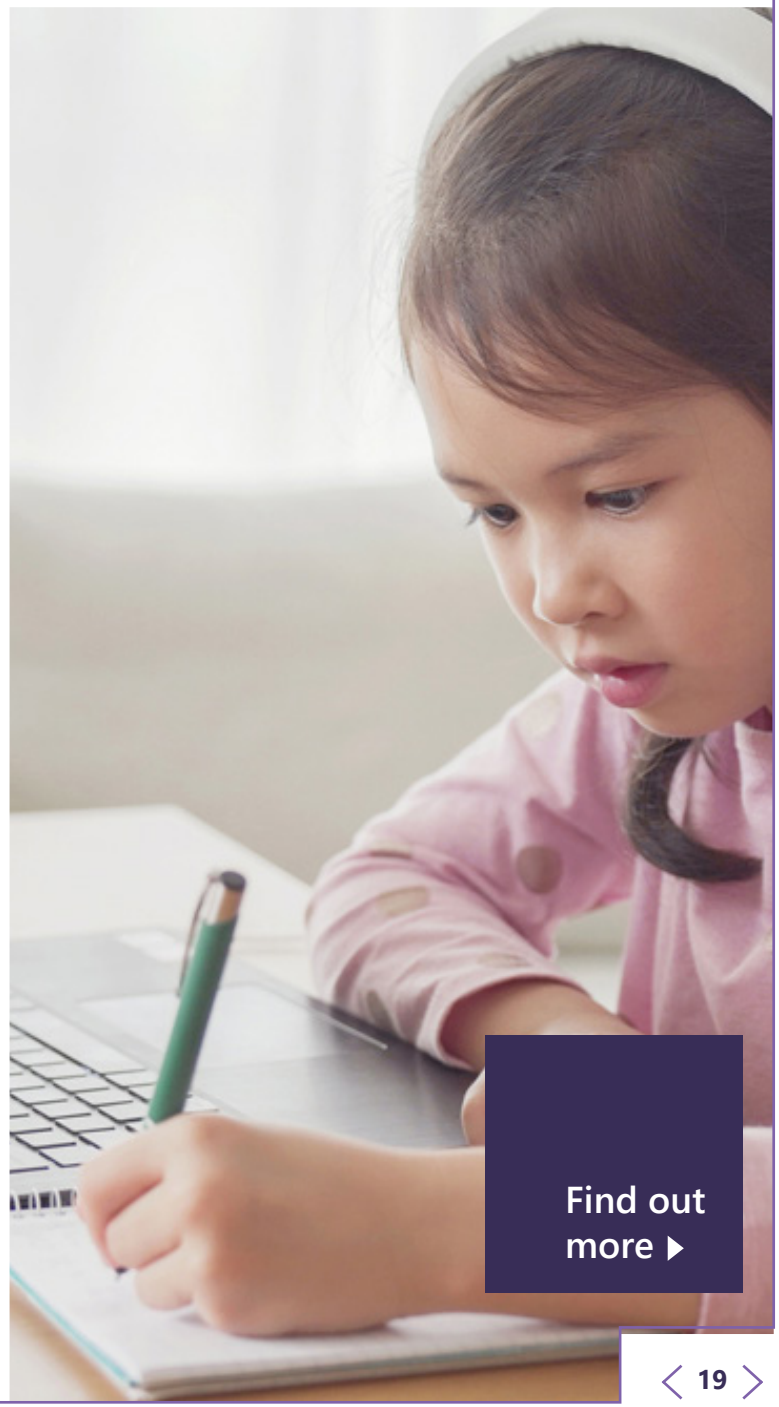
The ZIP code level dataset provides a granular view of broadband usage percentages by households within a ZIP code, so we took an additional step to ensure data privacy guarantees. We applied differential privacy, a technique that adds noise to the data aggregations and prevents leakage of personally identifiable information. BroadbandNow also participated, making its county-level pricing and broadband provider data available for the first time.

The Education Open Data Challenge is generating insightful analyses with combinations and visualizations of data. These can be used to help governments, policymakers, nonprofits, and organizations around the world better target resources so students have access to the education they need.

**Additional Resourses**

United States Broadband Usage Percentage Datasets

BroadbandNow



# NeurIPS 2020 Education Data Challenge

**Imagine:**
Improved online learning platforms, personalized to the student

**Partners:**
Eedi, Microsoft

The NeurIPS 2020 Education Data Challenge aimed to advance research in two areas via machine learning:

1. **Provide insights for teachers on how students respond to questions.**
2. **Personalize diagnostic questions for online learning platforms for students.**

Challenge participants used anonymized open educational data made available from the online education platform Eedi to accomplish a series of tasks. These tasks helped identify which questions would be most suitable for a particular student at a particular point in their learning journey. Microsoft Research-Cambridge supported the challenge by providing sample code, structuring the project on Azure, and contributing to the award for winning teams.

**Additional Resourses**

Post-competition research paper from the organizers

Find out more ▶

# Policing in America Survey

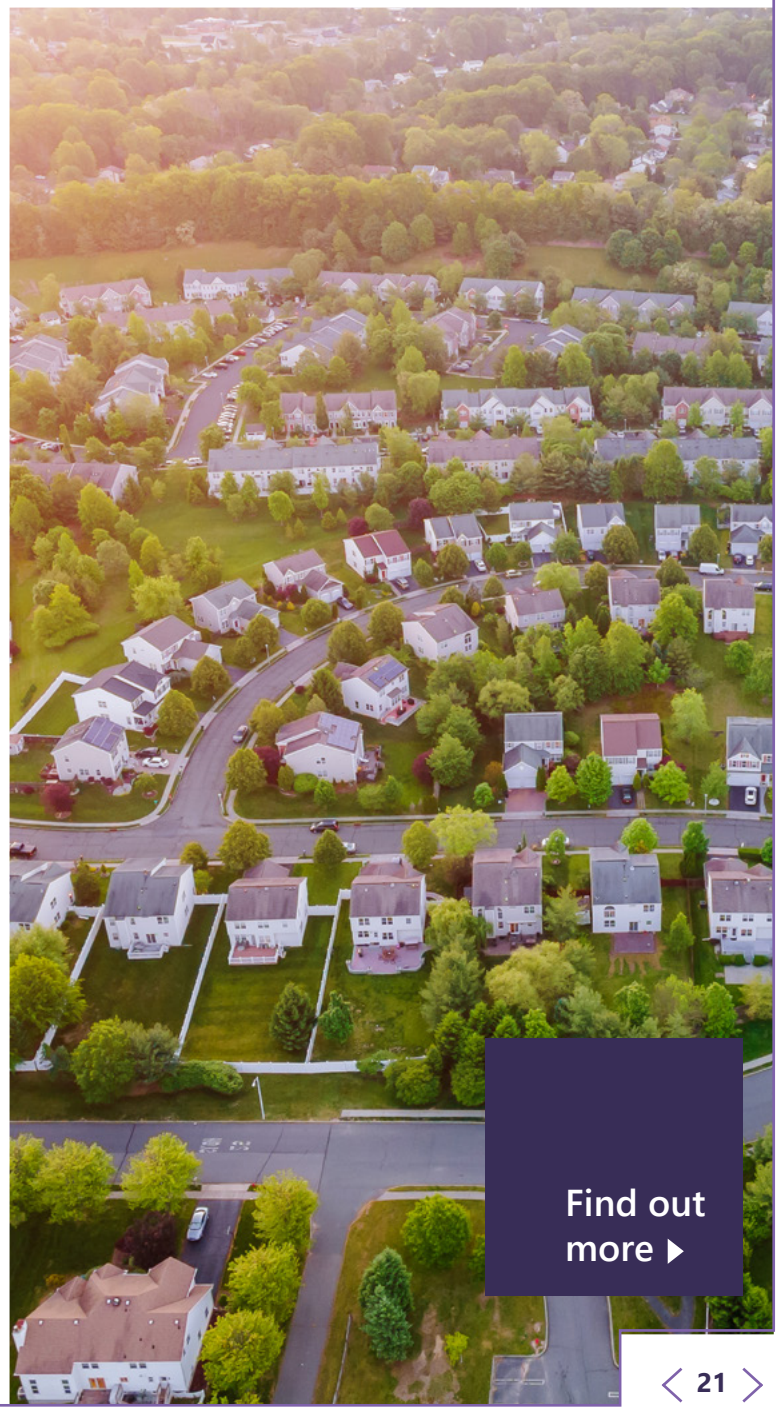
**Imagine:**

Knowing what policing reforms truly work

**Partners:**

Data Foundation, data.world, NORC/University of Chicago, SpotCrime, Microsoft

The Data Foundation, with the support of Microsoft, data.world, NORC at the University of Chicago, and SpotCrime, launched a new Policing in America Survey to fill a long-overdue gap in systematically understanding the American people's views of the criminal justice system and police forces. A free and open Data Resource Hub will be hosted on data.world, enabling widespread access to the public, researchers, and policymakers. The hub will combine information collected by the Policing in America Survey with existing data assets produced by federal, state, and local agencies and other nonprofit initiatives. City leaders and communities will have a powerful new tool for informing policy reforms and law enforcement efforts.



# OpenSidewalks

**Imagine:**

Stewarding healthy, inclusive, and resilient cities around the globe

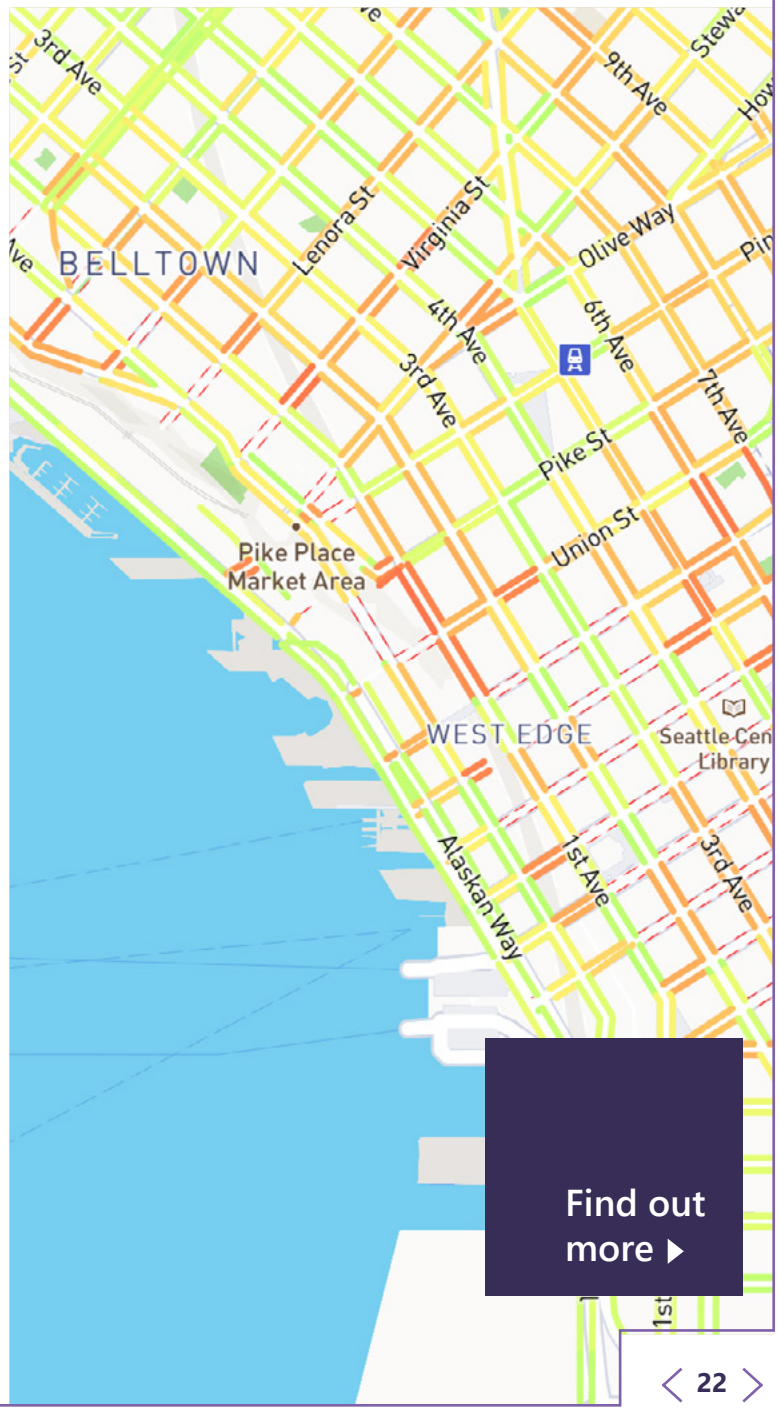
**Partners:**

The University of Washington Taskar Center for Accessible Technology, The Global Initiative for Inclusive Information and Communication Technologies (G3ict), Microsoft

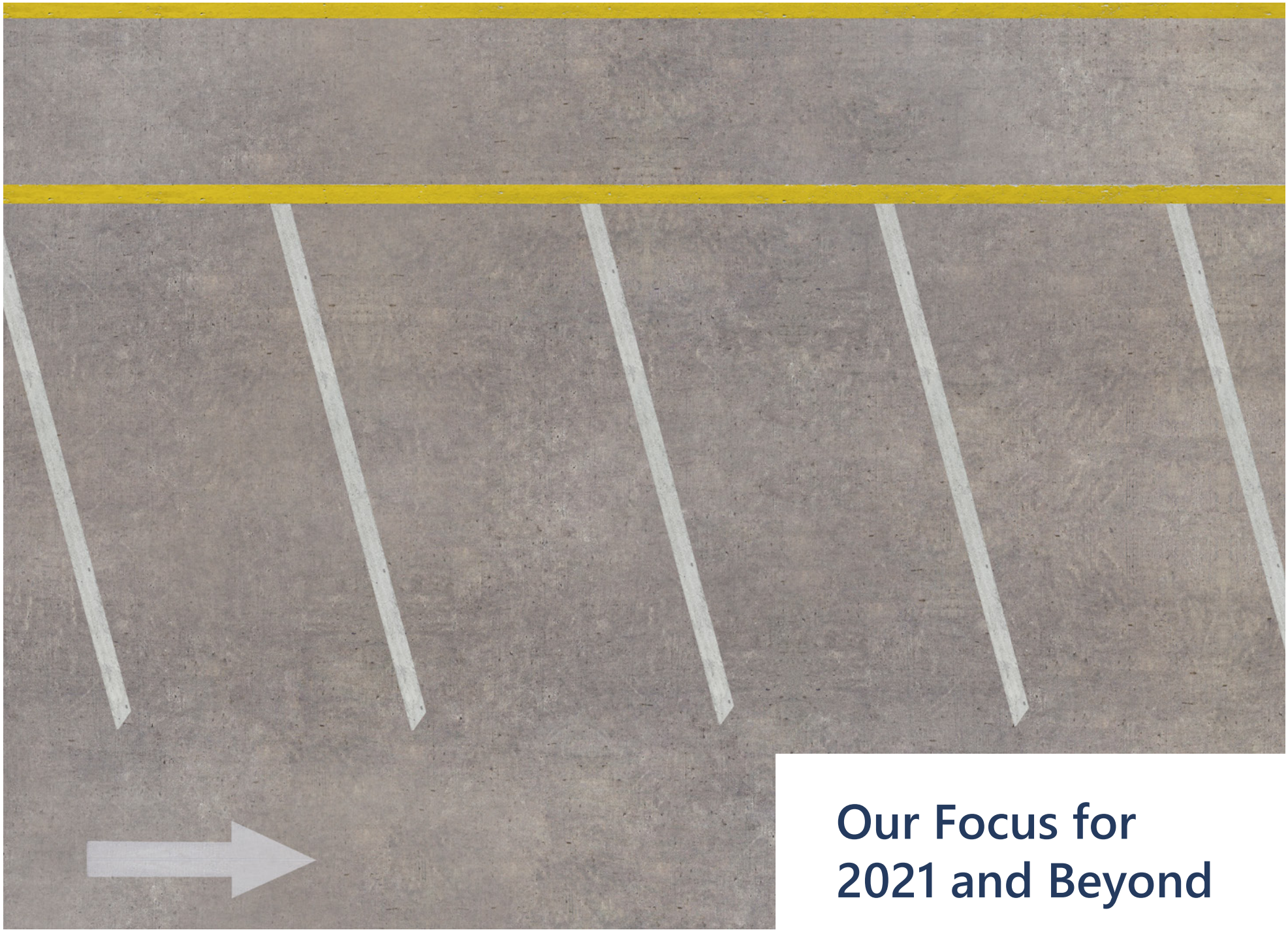
Everyone can be a pedestrian, and the population has a wide variety of needs and preferences for accessible navigation of sidewalks and pathways. For example, a person using a wheelchair may need to avoid steep hills whereas another person will attempt any hill. This project aims to empower disability communities by working with global cities to collect quantitative data about sidewalks—such as width, surface composition, steepness, and shared traffic—and provide applications to make use of this data.

The Taskar Center, hosted by the Paul G. Allen School of Computer Science & Engineering, will provide industry-leading academic expertise in collecting, modeling, and visualizing this data. G3ict will enable the effort to successfully partner with cities and communities to meet their unique needs. In addition to funding and technical expertise, Microsoft is providing Bing Maps aerial imagery from which the Taskar Center will extract sidewalk data at scale using AI.

People in the disability community will be able to leverage the AccessMap app for personalized pedestrian trip routing, currently supported in Seattle, Bellingham, and Mt. Vernon in Washington state. This collaboration will create new datasets for pedestrian routing in Sao Paulo, Brazil; Gran Valparaiso, Chile; Santiago, Chile; Quito, Ecuador; and Los Angeles, United States.



Microsoft



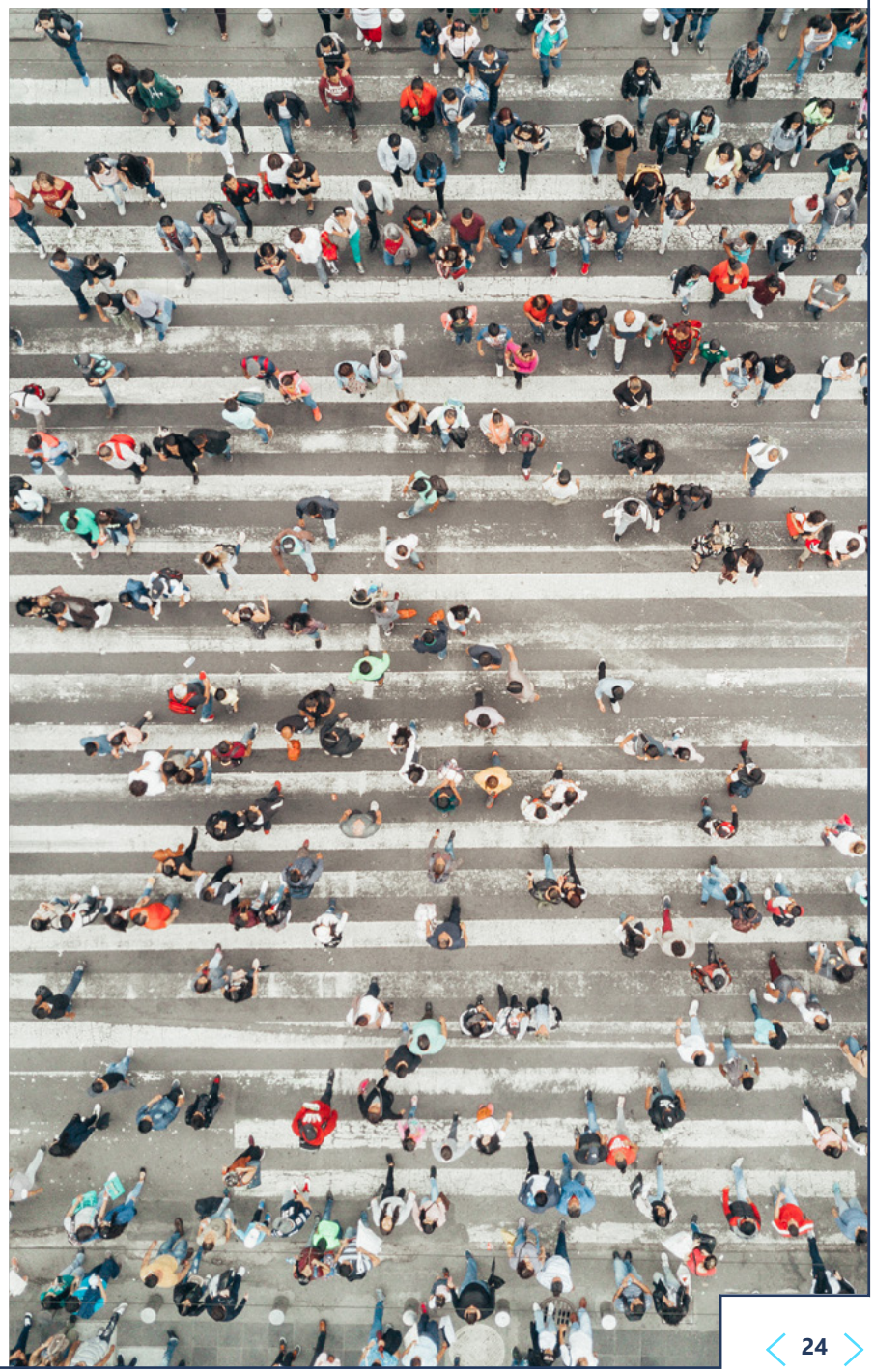
# Our Focus for 2021 and Beyond

In addition to our continued commitment to launching new data collaborations, we are focused this year on the practical aspects of data sharing and how to make it easier for public and private organizations of all sizes. Specifically, we're working with our partners to explore:

- **How do we advance the understanding and practice of data stewardship?**
- **How do we help close the data skills gap?**
- **How can we make data sharing easier through scalable tools and technologies?**
- **What policy steps can governments take to promote open data and data collaboration?**

Our aim with the Open Data Campaign is to empower those who have visions of a better, different future to achieve what they imagine through data. We believe that countries and organizations must have equal opportunity with data—and that those that consider data sharing and collaboration as fundamental to innovation will be the ones that succeed in this future. We're committed to leading and partnering within this new paradigm, through our Open Data Campaign.

Learn more about the Open Data Campaign at news.microsoft.com/opendata.

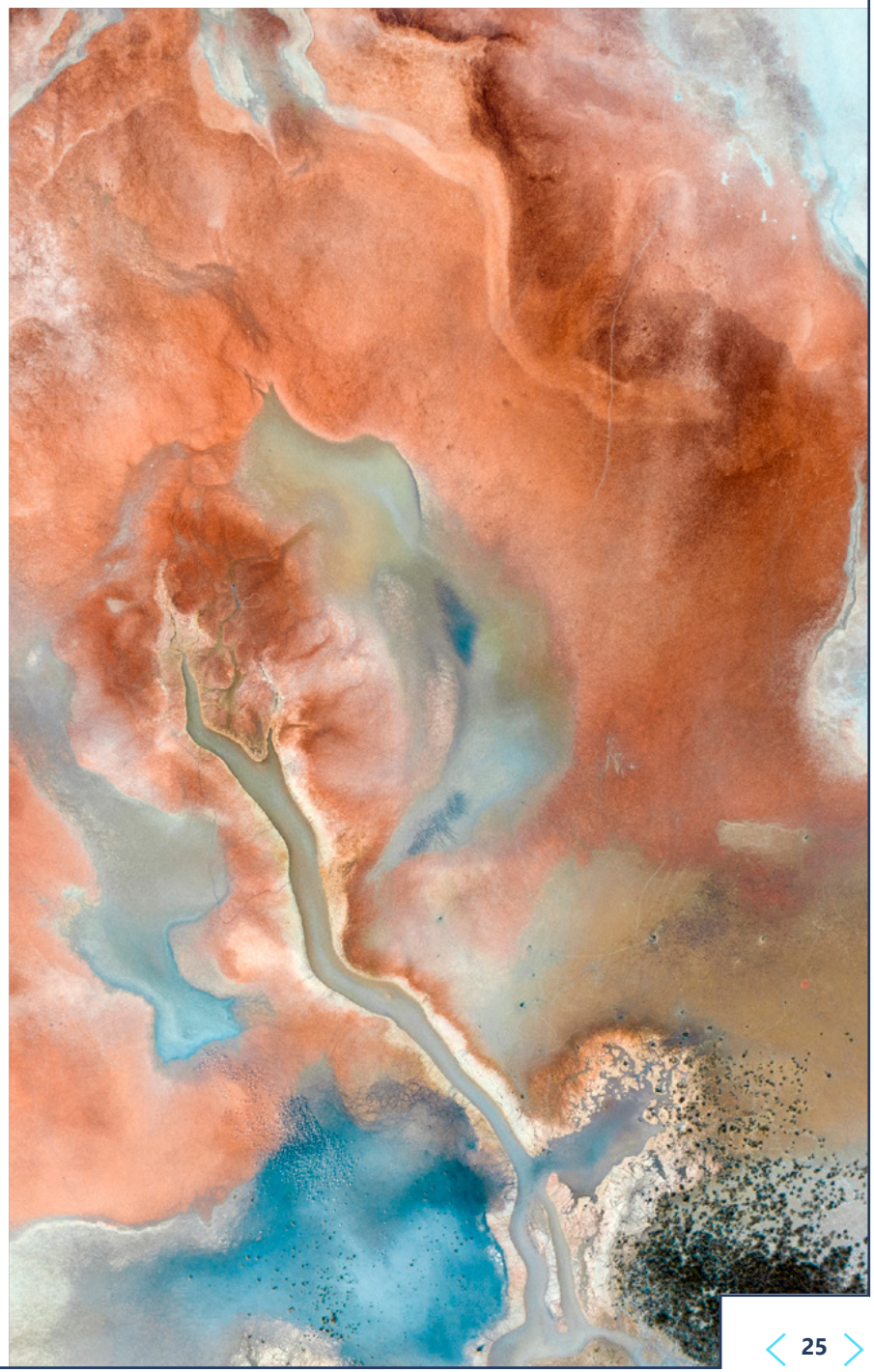


# Footnotes

[1]Based on Microsoft CELA Data Science analysis of similarweb.com, appfigures.com, and alexa.com.

[2]Microsoft CELA Data Science analysis of LinkedIn data.

[3]Gartner, "Flip 'Don't Share Data' Mantras — Introducing Gartner's 'Must Share Data Unless' Data Sharing Model", September 1, 2020.



Microsoft