



INDUSTRY DATA FOR SOCIETY PARTNERSHIP

2023 YEAR IN REVIEW

Overview

LETTER FROM THE PARTNERS 3

FIRST-YEAR MILESTONES AND KEY FINDINGS 4

Data Challenge: Open Data for Local Solutions 5

Impact Beyond the Data Challenge 6

Lessons Learned for Organizing a Data Challenge 7

U.S. Research Data Summit 8

Engaging with Open Data and Policy Communities 9

INDUSTRY INSIGHTS AND IMPACT 11

GitHub’s Innovation Graph 12

LinkedIn’s Data for Impact Program 13

Northumbrian Water’s Open Data Strategy 14

Open Energy Maturity Framework hosted by UKPN 15

FEATURES FROM OUR AFFILIATES 16

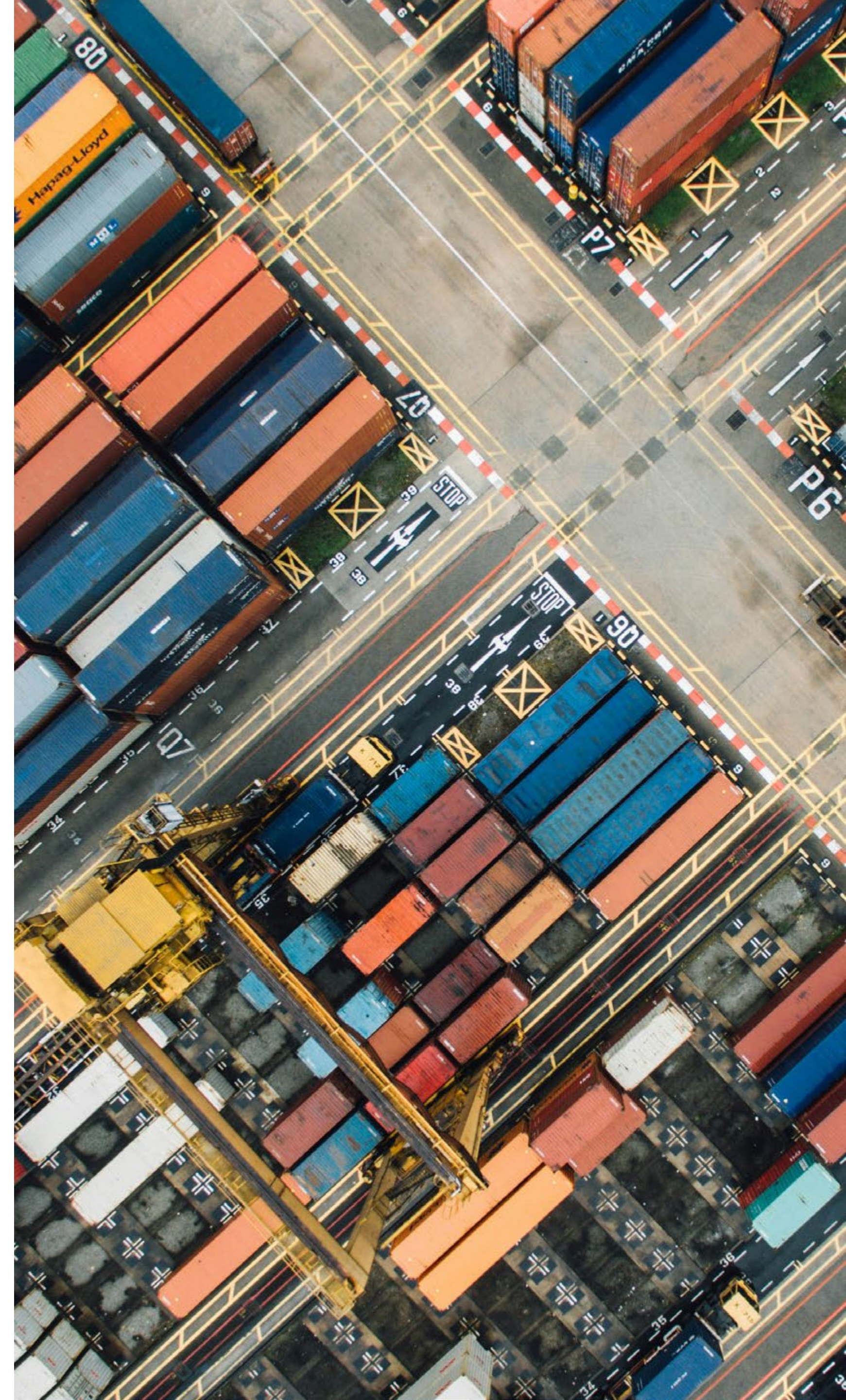
Perspectives from The GovLab by Dr. Stefaan G. Verhulst at The GovLab 17

Perspectives from the Open Data Institute by Josh D’Addario at the Open Data Institute 19

FUTURE DIRECTIONS 22

Future Directions and Opportunities 23

Join Us in Shaping a Data-Driven Future 24



Letter from the Partners

We are pleased to present the inaugural Year in Review of the Industry Data for Society Partnership (IDSP). Launched in December 2022, the IDSP is a network of leading companies committed to sharing private sector data to address societal issues. This report provides an overview of the partnership's activities, achievements, and insights over the past year.

Our belief is that data is a powerful resource for accelerating the development of solutions to benefit society. By sharing datasets, insights, and expertise with researchers, policymakers, civil society, and other data practitioners, we aim to demonstrate the value of using private sector data to tackle critical global challenges including climate change, health, education, and inequity.

The context for our work is rapidly evolving and expanding. Data is increasingly recognized as a key resource and driver for economic growth, social development, and scientific discovery. However, data also poses significant challenges and risks, relating to privacy, security, quality, and governance. Moreover, data is often siloed, inaccessible, or underutilized, limiting its potential for public benefit. Therefore, there is a need for more open, collaborative, and responsible approaches to unlock valuable outcomes from data.

Collaboration is key to achieving our mission, as we cannot achieve our goals alone. We are grateful for the support and guidance of two organizations: The GovLab at New York University and the Open Data Institute (ODI). As affiliates of the partnership, they have provided invaluable advice, resources, and connections to shape and implement our data initiatives. Their contributions also expanded our reach within the broader data community, enhancing our impact.

Our members represent a diverse range of industries and share a common vision and commitment to more open and accessible data. Each has a unique story to tell about their journey towards opening up data and the benefits and challenges they have encountered along the way.

In this report, you will find out more about our backgrounds, motivations, and contributions to the IDSP, as well as our aspirations and plans. We'll share how convening the IDSP has helped bring forward new insights, skills, networks, and opportunities.

To learn more, please contact us at info@industrydataforsociety.com or visit our Industry Data for Society Partnership [website](#) for more information.

Together, we can harness the power of data to create a better world for everyone.

Peter Cihon, Senior Policy Manager and Abby Rieflin, Senior Corporate Communications Manager, GitHub

Grant Ritz, Director & Associate General Counsel, Intellectual Property Transactions, Hewlett Packard Enterprise

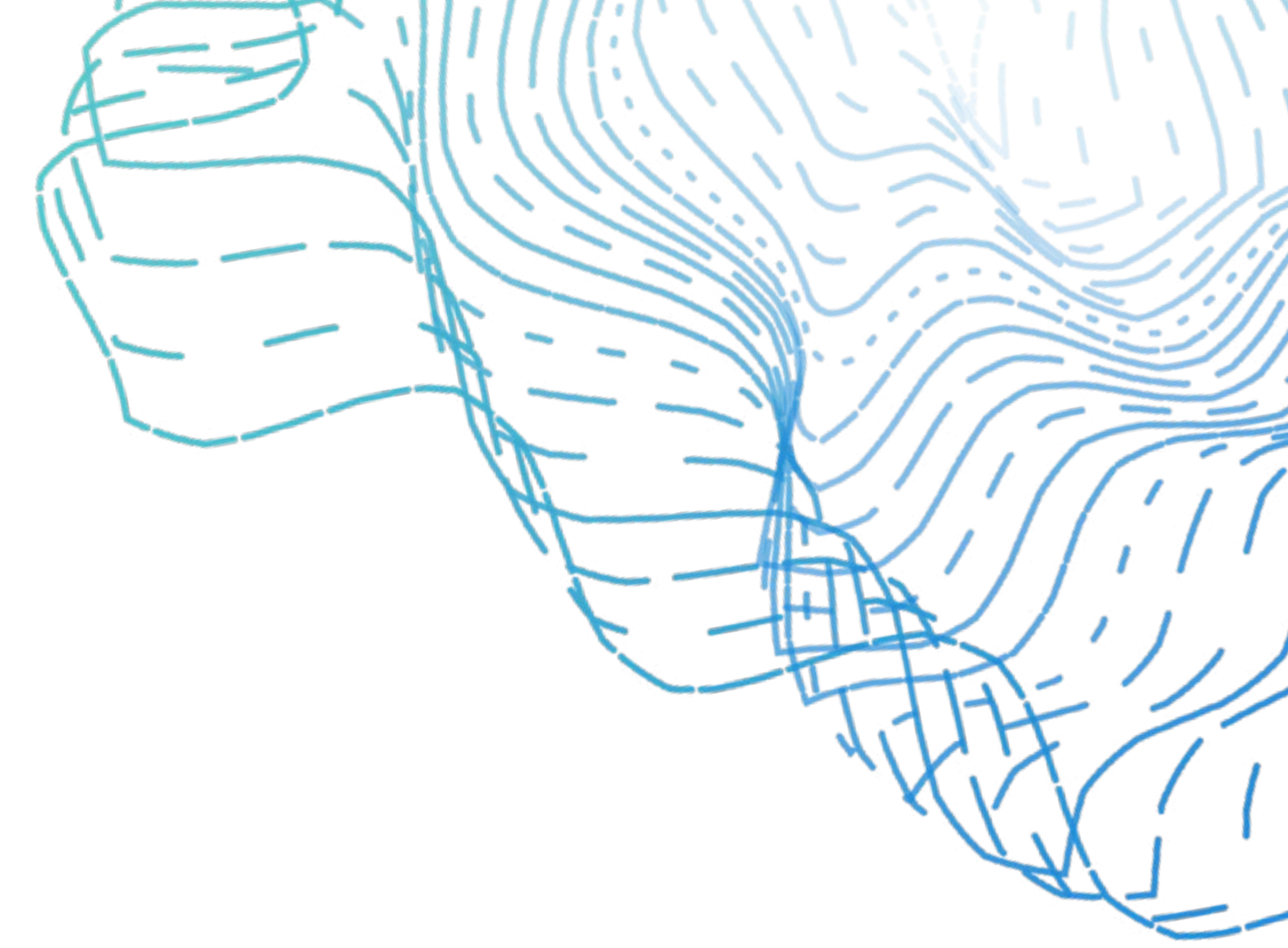
Casey Weston, Data for Impact Co-Lead, LinkedIn

Sonia Cooper, Associate General Counsel, Open Innovation, Intellectual Property Group, Microsoft

Nigel Watson, CIO and Melissa Tallack, Open Data Lead, Northumbrian Water Group

Bridget Erb, Ookla for Good Analyst, Ookla

Matt Webb, Head of Enterprise Data Management and Yiu-Shing Pang, Open Data Manager, UK Power Networks



IDSP PARTNERS

GitHub

**Hewlett Packard
Enterprise**

LinkedIn

Microsoft

NWG
living water

OOKLA

**UK
Power
Networks**

**FIRST YEAR
MILESTONES AND
KEY FINDINGS**



Data Challenge: Open Data for Local Solutions

The IDSP and the ODI launched the Data for Local Environments Challenge in February 2023, with the aim to highlight the role that publicly available data can play in helping local governments to improve their local environments. Data scientists, citizen scientists, researchers, and other experts submitted a range of projects that explored the potential to apply open data to local environmental challenges. The [winning applications](#) highlighted the following themes and opportunities that are applicable to local governments globally:

1. Datasets from both the public sector and the private sector can be leveraged to better understand environmental factors

The overall winning challenge team, Estimated Energy Attribution Certificates based in the UK, used data from Microsoft's Planetary Computer and UK Power Networks — for historic weather patterns and capacity registers, respectively — to calculate the total renewable generation in each local authority. While public sector datasets are often looked to as a source for energy analysis, the private sector is making strides to make sustainability data more accessible to advance research in addressing climate-related challenges. Local governments can leverage these datasets in combination with others for deeper analysis.

2. Informative resources for local governments are increasingly open sourced — this includes code for renewable generation estimation and CO2 analysis, and a framework for collating datasets

One of the biggest challenges faced by local governments can be collating and standardizing across datasets, including linking similar data points across

data sources. The Estimated Energy Attribution Certificates challenge team developed code for renewable energy estimation and CO2 analysis, alongside a [framework for collating data](#) across the power sector. The resulting package was released as open source on GitHub, and can be leveraged by local governments globally. Additionally, the 2nd runner up, Climate Emergency UK, released their [Council Climate Action Scorecards](#) in the public domain.

3. Bringing diverse datasets together creates greater clarity for both citizens and local governments on the goals and progress

The Climate Emergency UK team identified the challenge posed by siloed, sector-specific data, resulting in local residents, councilors, and officers struggling to understand their council's progress to net-zero and develop impactful solutions to reach their goals. Through the development of Council Climate Action Scorecards, a one-stop shop was created on all actions councils can take and have taken — thereby, helping to unite the local climate action movement. Greater clarity and access to information increased citizen knowledge and allowed residents and councils to work together. Additionally, councils could see what others were doing and benchmark against best practices.

4. Data analysis skills can be leveraged from students, citizen scientists, researchers, and other sources of expertise, without necessarily needing to hire a full-time position in-house

Applicants for the Data for Local Environments Challenge had a broad range of skills and diverse backgrounds, all pursuing projects in their own

capacities outside government. This highlights the wide talent pipeline from which local governments can tap into for data analysis expertise. Many of the applicants were eager to work with local governments and the broader community to partner on the challenges and potential solutions. The Climate Emergency UK team not only worked with local councils, but aims to provide trainings to councils on policy and actions found within their scorecards.

5. Advancing new ways to gather and disseminate data are critical to empowered communities

The team winning 1st runner up, Infrastracker — Open Air Data based in Kenya, is pursuing a project that integrates public participation and community involvement in environmental conservation. The project relies on members of the community to monitor climate indicators, particularly carbon emissions during road construction projects, through local weather monitoring stations. By involving the community directly in the gathering of data, this approach serves to inspire involvement, building the capacity of communities to gather the data and see first-hand how data can lead to greater awareness and influence change — in this case, bringing awareness to vulnerable members of the community from the negative effects of poor air quality, with an aim towards support for expanding the scope of air quality regulations. The Climate Emergency UK team also operates with critical community engagement, including the involvement of over 200 volunteers.

Additional resources for local governments and cities are provided by The GovLab: [An agenda for advancing trusted data collaboration in cities.](#)

Impact Beyond the Data Challenge

An essential part of the Data Challenge was that participants' contributions shouldn't necessarily begin and end with the competition. The winners of the challenge have extended their efforts beyond the competition. Following are some examples:

Estimated Energy Attribution Certificates, linked to Councils (UK) by Ayrton Bourn (Overall Winner)

Ayrton Bourn's submission for the Data Challenge focuses on creating Estimated Energy Attribution Certificates (EEACs) that are calculated on a half-hourly basis and tied to individual assets. Since the challenge, he has taken additional steps to develop the Power Station Dictionary website.

Bourn shifted his focus to the project's back-end features, concentrating on building automations for data feeds and pipelines to connect various data sources. He continues to work on standardizing representations for common energy assets like power stations and substations while integrating text-based information into his platform. Additionally, he is exploring collaborations with other external organizations with connections to local councils.

Infrastracker, Open Data for Local Governments to Reduce Infrastructure Development-Co2 Emissions (Kenya) by the Demography Project (1st Runner Up)

The Demography Project continued to develop their Infrastracker project by increasing outreach to key stakeholders in the road and infrastructure sector, as well as started new projects around air quality and water conservation. They are in the process of creating a comprehensive database of the classification of all roads in Kenya, listing all road work projects dating back to 2013. The Demography Project aims to calculate the overall carbon footprint during and after road construction projects and highlight the lifetime of the road itself.

The aim of collecting emissions data is not limited to roads; they are looking to upgrade equipment in their offices which will additionally enable them to capture emissions data from aircrafts. Raising awareness around water conservation is another focus as they look to educate communities about the harmful impacts of residue left behind after construction projects that lead to water pollution.

Council Climate Action Scorecards (UK) by Climate Emergency UK (2nd Runner Up)

After the Data Challenge, Climate Emergency UK launched their proposal idea, the Climate Action Scorecards (Scorecards) which evaluates local government progress towards net zero goals. These Scorecards have garnered media attention and have drawn the interest of numerous councils. Before the official launch of the Scorecards, the team made the Freedom of Information (FOI) requests to all local councils, garnering coverage on the postcode lottery for energy efficient social housing.

Notably, the FOI request data resonated beyond its immediate domain, finding utility in various sectors. For instance, the Green Jobs Foundation used FOI data gathered by one of the FOI requests on green skills training courses offered by combined authorities and the number of attendees, in their State of the Nation report. The Scorecards enable stakeholders to identify the highest-scoring councils and delve into the best practice from each council. Climate Emergency UK is now commissioning a research project, in collaboration with Anthesis and Environmental Consultancy, utilizing the vast amount of information collected during the Scorecards project. They aim to publish the results of this project in the upcoming year.



Lessons Learned for Organizing a Data Challenge

The [Data for Local Environments Challenge](#) was a key project of the IDSP in its inaugural year of partnership. This challenge focused on harnessing publicly available data to empower local governments in enhancing their environments.

Key Highlights

The challenge was supported by rich and diverse datasets made available by IDSP members and was hosted in partnership with the ODI. These datasets provided a basis for participants to craft solutions, showcasing the potential of open data. The challenge attracted global attention as it drew involvement from 24 countries, spanning from Malawi to Sri Lanka.

Among the seven shortlisted candidates, we found not just impressive solutions but, more significantly, actionable strategies that local governments can readily embrace to improve their environmental efforts. For instance, participant Oliver Morris's solution centered on the promotion of local biodiversity by utilizing Ordnance Survey's Green Space dataset to identify suitable locations for green roofs and solitary bee habitats. Another compelling solution by ReData revolved around supporting the UK Cambridgeshire County Council in its journey toward decarbonizing schools.

The array of solutions, each making innovative use of openly available datasets, underscored the many ways in which open data can be harnessed for the betterment of local environments. More detailed information about these shortlisted proposals can be found [here](#).

Lessons Learned hosting the Data Challenge

Conducting the Data Challenge served an invaluable learning experience for the IDSP as organizers. Here are some of the lessons we learned from the Data Challenge:

1 Formulating the Challenge Question

One of the key takeaways from hosting the Data Challenge was the significance of crafting a clear and compelling challenge question. The question guides participants in their data analysis journey. We learned that a narrower, more focused question is likely to provide clearer guidance for the participants, so there is a common approach to answering a challenge question that is less open to interpretation. More broadly framed questions will invite a greater variety of solutions.

3 Communication with the Participants

Insights from participant feedback prompted us to consider the implementation of a chat function. Such a feature could facilitate collaborative discussions among participants about the challenge questions, potentially fostering connections and encouraging individuals to form teams for a more collaborative experience.

2 Spreading the Word

Allocating dedicated time for pre- and post-challenge promotion can help drive greater engagement. An extended promotion phase enables organizers to reach a wider audience and amplify the impact of the challenge, engaging more participants.

4 Timing and Flexibility

Adaptability in the timeframe is important for any challenge team. As the challenge unfolds, building in the ability to adapt to different factors, such as the number of applications received, can help ensure an optimal experience for applicants, and ultimately, a robust set of applications and solutions.

U.S. Research Data Summit

A major milestone of the IDSP's inaugural year was our involvement in the [U.S. Research Data Summit](#), an event that brought together leaders and experts from various sectors and disciplines to discuss the opportunities and challenges of research data in the United States and beyond.

The Summit, held at the National Academies in Washington, DC, on October 10 and 11, 2023, aimed to foster greater coherence, communication, and collaboration among the diverse stakeholders and initiatives related to research data, as well as to position the United States as a strong and active participant in the global dialogue on this topic.

As the partnership's program of work was developed in early 2022, we saw an opportunity for the IDSP to serve as a conduit for researchers to connect with industry and for the private sector to establish new connections and build relationships with the global research community.

One of those opportunities presented itself through outreach from the Association of Research Libraries (ARL) and the United States National Committee for CODATA, as they started to explore the interest of U.S. research data organizations to strengthen coordination and act on a shared agenda. The IDSP recognized that it was a substantial opportunity to demonstrate the partnership's commitment to building relationships with the global research community to use private sector data for societal benefit.

Determining the Road to an In-Person Summit

The journey toward the Summit started with the Association of Research Libraries (ARL) and the United States National Committee for CODATA (USNC/CODATA) surveying 80 representatives from U.S. research data organizations with national and international research data goals, including organizations that are cross-functional, disciplinary, representative of scholars, government, industry, or nonprofit organizations.

The ARL and USNC/CODATA hypothesized that U.S. research data organizations would find greater coordination and collaboration beneficial, and the questionnaire sought to identify what level of interest there was among the respondents and learn more about the areas that would be of interest to explore together. The survey results confirmed the hypothesis, and the National Academies appointed a planning committee to convene the U.S. Research Data Summit.

Planning and Convening an In-Person Summit

To inform the Summit agenda, the planning committee held virtual focus groups to explore themes uncovered by the ARL and USNC/CODATA survey. They included organizational approaches to AI, cross-sector access to research data, decarbonization data needs, disruption data, Indigenous data governance, and justice, equity, diversity, and inclusion considerations.

The Summit itself convened national organizations from various sectors to strategize on increasing coherence, communication, and collaboration in research data policies and practices within the United States and fostered dynamic discussions and collaborations, resulting in a draft of data sharing principles focused on leading with empathy, empowering data workers, building trust, prioritizing equity, making evidence-based decisions, and ensuring sustainable compliance.

Potential collaborations were also identified, with one proposed initiative aiming to use a data-driven approach to strategically allocate resources to decarbonization and green energy transitions across environmental, social, and economic dimensions. This initiative would unite stakeholders from policy, corporate, governmental, private, and academic sectors, capitalizing on existing expertise and initiatives within these sectors. Using LinkedIn data was recommended as an initial step for identifying current initiatives.

Providing an Industry Perspective

IDSP member, Jennifer Hansen from Microsoft, was appointed as a Co-Chair of the planning committee, which enabled private sector input from the outset and a strong research-industry partnership in the development of the direction and outcomes of the summit.

Melissa Tallack, Northumbrian Water Group; Matt Webb and Yiu-Shing Pang, UK Power Networks; and Kathy MacDonald, Ookla provided industry perspectives by participating in focus groups which helped shape the Summit's agenda.

Casey Weston from LinkedIn and Stefaan Verhulst from The GovLab, an affiliate of the IDSP, provided valuable insights as panelists during the session as well as contributing their expertise to the breakout sessions, Charting a Path for AI Policies, Ethics, and Cross-Sector Adoption and Collaborative Approaches for Data-Driven Application Decarbonization and Resilience.

Looking ahead, the National Academies will publish a Proceedings of a Workshop-in Brief in early 2024 and the planning committee will continue to engage with the Summit participants to identify areas for collaboration.

This summit marked a meaningful advancement in fostering relationships with the research community, promoting greater openness in private sector data, and identifying collaborations for positive societal progress. Moving forward, the IDSP will continue to seek opportunities to join and bring these communities together.

Engaging with Open Data and Policy Communities

The IDSP is committed to sharing our knowledge, experience, and perspectives with the global research community, public sector, intergovernmental organizations, and other industry leaders. In the past year, we've participated in events and authored publications that showcased our work and vision for the future of open and accessible data in the private sector. Below are highlights of our activities, which underscores the partnership's commitment to fostering collaboration across sectors for societal benefit.

Data for Policy 2022

The Data for Policy conference, held annually, serves as a global platform for interdisciplinary discussions on data science innovation's concepts, uses, and impacts in policy and the public sector. In December 2022, participants from academia, nonprofit, and government sectors convened to explore the ecosystems of innovation and virtual-physical interactions.

IDSP member Jennifer Hansen from Microsoft participated in the *Open Data: Privacy, Ownership, Trust* panel, where the panelists discussed the need to establish trust and incentivize data sharing, proposing solutions like a Data Ombudsman role and an Open Data Ambassador program.

UK Water Open Data Forum

Northumbrian Water Group led a small group to establish the first Open Data Forum for Water, hosted by Microsoft in September 2023. In addition to Microsoft, Northumbrian Water was supported by some of the IDSP members and affiliates GitHub, UKPN and the Open Data Institute. They joined representatives from various sectors within the water industry, including water companies and industry bodies to define, document, and publish a [draft industry roadmap](#) to drive tangible progress.

Data & Policy Journal Commentary

In June 2023, the Data & Policy Journal published an article authored by members, Jennifer Hansen of Microsoft and Yiu-Shing Pang from UK Power Networks, titled [Opening industry data: The private sector's role in addressing societal challenges](#). The piece explores how private company data can drive scientific progress and address societal issues. It emphasizes the benefits of open data in expediting discoveries, fostering collaboration, and ensuring business longevity. While data privacy and security are valid concerns, companies can employ practical measures like robust data governance and stakeholder engagement.

The commentary underscores pressing societal issues, including climate change, poverty, and limited access to education and employment. It highlights the significant contribution the private sector can make by making their data more accessible. This approach drives innovation and enables companies to actively participate in finding meaningful solutions.

Additionally, the article provides actionable steps for companies dealing with privacy and security issues, highlighting the tangible advantages that open data can offer in advancing both scientific progress and societal well-being.



Engaging with Open Data and Policy Communities

State of Open Data Policy Summit 2023

In June 2023, Gretchen Deo of Microsoft participated as a panelist at the [State of Open Data Policy Summit](#), hosted by the Open Data Policy Lab. This annual conference convenes stakeholders from across government, industry, and civil society to explore practices for making open data policy impactful.

The *Open Data and Data Collaboration: Private Data for Public Good* panel delved into the unique considerations of the private sector in advancing open data initiatives, and highlighted current innovations in the field.

State of Open Con 23

In February 2023, IDSP members Matt Webb of UK Power Networks, Melissa Tallack from Northumbrian Water, and Sonia Cooper from Microsoft, participated as panelists at the State of Open Con, an event hosted by Open UK.

They contributed to the session titled *Open Data for Climate Action*, which examined the role of public-private collaboration in utilizing open data to address the challenges posed by climate change. The discussion emphasized how joint efforts between governments and the private sector can harness open data to bolster climate action.

Expanding Open Data Infrastructure: A virtual roundtable by the ODI, LSEG and Microsoft

In October 2023, ODI, Refinitiv and Microsoft hosted a roundtable to discuss the importance of open data and open infrastructure to help build markets, deliver value to individuals and organisations, and improve society and the environment.

The roundtable convened private sector companies across multiple industries. All participants worked in some part of the ESG and net zero ecosystems. The discussion produced insights around how further open data collaboration and innovation can help us all achieve our missions.

ODI Summit 2022 & 2023

The Open Data Institute, an affiliate of the IDSP, hosts an annual summit to convene representatives from civil society, the public and private sectors, and academia to discuss current events happening in the open data space.

Demonstrating its commitment to continued engagement, IDSP members from Microsoft, UK Power Networks, Ookla, and LinkedIn joined panels for the 2022 and 2023 summit to talk about the opportunity for the private sector to open and share its data to help solve societal challenges and why efforts like the IDSP play a role.

Through engagement activities such as these, the IDSP strives to contribute to the advancement of the open data agenda for the private sector to benefit society, and to work with actors in the data ecosystem to develop solutions together for societal good. We look forward to continuing our engagement and collaborating with other like-minded organizations and individuals in this space.





INDUSTRY INSIGHTS AND IMPACT

GitHub's Innovation Graph

In September of 2023, GitHub announced the launch of the GitHub Innovation Graph, an open data and insights platform on the global and local impact of software developers. In the past, measures of innovation have focused solely on resources like patents and research papers, while policymakers and researchers struggled to find reliable data on global trends in software development. GitHub created the Innovation Graph as a solution.

The Innovation Graph includes longitudinal metrics on software development for economies around the world. This open data initiative was launched with a dedicated [webpage](#) and repository, and provides quarterly data on eight metrics dating back to 2020: Git pushes, developers, organizations, repositories, languages, licenses, topics, and economy collaborators. The platform offers several data visualizations, and the repository outlines GitHub's methodology. Data for each metric (licensed) is available to download ([CCO-1.0](#)).

GitHub's Innovation Graph will be useful for researchers, policymakers, and developers alike. In [research](#) commissioned by GitHub to help design the platform, consultancy Tattle found that researchers in the international development, public policy, and economics fields had interest in using GitHub data but faced many barriers while trying to obtain and use the data. The Innovation Graph aims to lower those barriers. Researchers in other fields will also benefit

from convenient, aggregated data that may have previously required third-party data providers if it was available at all.

Promoting digital transformation and well-paid jobs is a key goal for many policymakers. GitHub was encouraged to see research indicating that open source contributions on GitHub were associated with [more startups](#), [increased innovation](#), and [tens of billions of euros in GDP](#). It is anticipated that more readily accessible data will contribute to more (and compelling) research, and ultimately an increase in policies that foster developer opportunity, as well as an increased opportunity for someone to become a developer in the first place.

Developers will be able to see and explore a broader context for their contributions, for example, the ways in which developers collaborate across the global economy, or how a particular language or topic they may be interested in is trending in their local economy or around the world.

GitHub released the Innovation Graph as a data resource for community reuse and is excited to see how policymakers, researchers, and companies explore data trends, use the data to inform research and make beautiful visualizations, and for developers to show how their contributions relate to broader trends.



LinkedIn's Data for Impact Program



Over the past year, LinkedIn's [Data for Impact](#) program has helped [inform digital training programs in Lebanon](#), [AI workforce strategies in Latin America](#), [economic policy in Germany](#), and [cybersecurity dialogues in the US](#). As these contributions demonstrate, LinkedIn takes seriously its commitment to partner with public benefit entities, at no cost, to achieve its vision of generating economic opportunity for every member of the global workforce. Governments, multilaterals, and non-profits all leverage LinkedIn's unique insights on [digital skills and AI](#), [the green transition](#), and [gender equity](#) to improve labor market function and equity.

As a founding member of IDSP, LinkedIn has sought to leverage this community of industry partners to innovate around data sharing agreements, project prioritization, and private sector incentives. The story of LinkedIn's collaboration with non-profit workforce innovator, International Youth Foundation (IYF), illustrates three lessons learned over this year: create data menus, find the impact, tell the story.

Create data menus: IYF sought to boost the supply of skilled professionals for the electric vehicle (EV) manufacturing sector in Mexico, a critical sub-industry for Mexican economic development and a just green transition. IYF had a variety of thoughtful questions for LinkedIn about EV skill supply and demand, but not all of them could be answered by the no-cost Data for Impact program. By developing a clear menu of data and insights on offer, the teams were able to co-craft a data request that could produce the necessary insights in a timely fashion.

Find the impact: LinkedIn's Data for Impact program gets approached by many researchers and practitioners seeking insights for exploratory purposes. While admirable and necessary, exploratory exercises alone cannot lay the foundation for a private sector data-sharing operation; to justify investment beyond what regulations require, programs like Data for Impact must demonstrate that some portion of their work drives real-world impact. IYF stood ready to take LinkedIn's insights directly to market, incorporating them into training curricula with promise to advance economic development and a just green transition for members of Mexico's workforce.

Tell the story: IYF and LinkedIn both sought to celebrate this data sharing collaboration as a means of encouraging similar efforts and drawing industry members into conversations around what skills the green transition requires. The collaborators co-authored a [blog](#) on their work to highlight the nuts and bolts of how their data-sharing operation functioned and request advice from peers on how to improve it moving forward.

Data for Impact looks forward to collaborating with the IDSP community to deepen its commitment to privacy-preserving data sharing and build on these lessons learned in the year ahead!

Northumbrian Water's Open Data Strategy

Northumbrian Water believes strongly in the power of data to fuel innovation. They have a long track record of facilitating successful innovation through their annual innovation festival, which has been running since 2017. Within the water industry, many of the challenges they face are complex and involve many stakeholders. Through sharing and opening their data, it can be utilized by others alongside other data to unlock value and greater insights for their customers.

The principle behind the festival was simple — set a challenge, convene a diverse group of people, input as much relevant data as possible, and create the right environment for creativity and collaboration to flourish.

Given the company's desire to innovate and push boundaries, it comes as no surprise that Northumbrian Water wanted to take their open innovation approach to a broader stage through opening up its data. The company felt it was important to publicly set out their ambitions and commitments in an Open Data Strategy which can be found [here](#).

The strategy is published under open license (CC-BY-SA) to help support and encourage other companies to do the same. Northumbrian Water's mission is to make a wide range of useful datasets available, along with contextual information that can help others use and interpret their data.

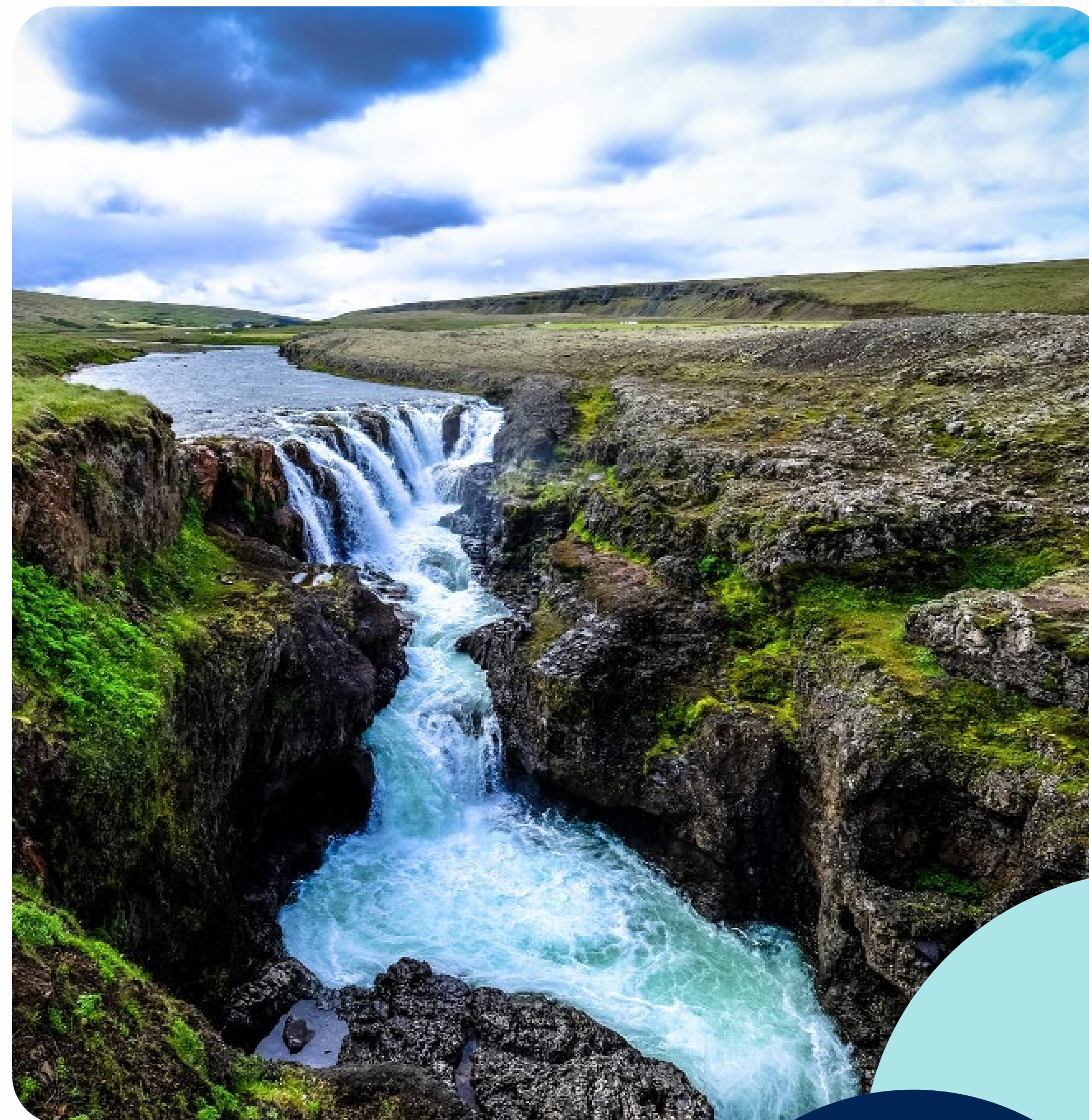
They aim to boost the discovery of innovative solutions that will improve their performance, as well as allow for the creation of new products and services. In order to achieve these goals, Northumbrian Water has created a

six-pillar strategy based on the following pillars: Northumbrian Water will:

- Continue to lead the water industry as it moves together towards Open Data.
- Publish data openly in an easy-to-consume way.
- Focus efforts on improving the lives of its customers and the precious environment in which we live.
- Collaborate with other primary, secondary and third sector organizations such as utilities, higher education, social enterprises, and other businesses in the region to enable better cross-sector outcomes.
- Convene a vibrant local ecosystem, promoting citizen data science and building valuable skills in the region.
- Build out the enabling people, process, and technology capabilities for open data within the organization.

In addition, Northumbrian Water, along with 11 other water company partners, were successful in securing innovation funding from the Ofwat Water Breakthrough Challenge (WBC) competition in May 2022. This initiative is called Stream and involves 13 of the 17 English, Welsh and Scottish water companies, which collectively serve over 60% of the UK's population. The vision of Stream is to harness the power of water data to serve customers, society, and the environment.

Thanks to WBC funding, Stream has been able to create a high-level design for all components required to develop a common framework that will facilitate data sharing within and across sectors. This scalable Open Data framework aims to make it as simple as possible for water companies to publish data and for innovators to tackle key sector and cross-sector challenges. The first datasets will be published at the end of 2023.



Open Energy Data Maturity Framework hosted by UKPN



The [Open Energy Data Maturity Framework \(OEDMF\)](#) hosted by UK Power Networks (UKPN), and created by Matt Webb — its Head of Enterprise of Data Management, is a comprehensive resource that provides a definition of the enabling capabilities an organization needs to implement open data in energy. It also provides an assessment of corresponding maturity.

While the framework is specific to energy, most of it is applicable to other sectors. There is an increasing demand for energy data, driven by the need to tackle global warming, demand for renewable energy generation, electrification of transport, and changes in lifestyles. Open energy data helps in making quicker and better-informed investment decisions and improving the accuracy of modelling load or generation.

UK Power Networks' Open Street Works dataset is an example of a challenge but also an exciting opportunity. It is updated every two hours, showing where live street works activities are taking place. This requires good data governance, data quality, and automation. The dataset provides visibility to other utilities and local authorities. Real synergies could be achieved if proposed street works are surfaced, enabling opportunities for collaboration.

The OEDMF sets out seven themes, from strategy to data publication, and 38 dimensions that define the specific arrangements, capabilities, and processes that an organization needs to develop and implement.

Maturity is assessed based on these dimensions, with each dimension being assigned a maturity rating of one to five. Maturity indicators detail the specific characteristics within each of the individual dimensions. At the start of the Open Data program in 2021, a baseline assessment with a score of 45.11% was achieved. UK Power Networks set an aspirational overall end score of ~92%. At the end of 2022, the score was 67.56%, with strong progress made across all themes and dimensions. At the end of 2023, the score advanced to 73.60%.

UK Power Networks has found the OEDMF to be highly beneficial in informing team and individual target setting and has helped develop and deliver open data services. They encourage other organizations to utilize the framework to help inform their own strategies and development plans, establish a baseline, and assess their organization periodically to measure progress and identify strengths and weaknesses.

As more organizations use this framework and publish their scores, the industry can then benchmark organizations, driving a competitive state which benefits open data users.

FEATURES FROM OUR AFFILIATES

The GovLab at NYU and the Open Data Institute, affiliates of the IDSP, share their perspectives on private sector data availability in addressing societal issues and provide a look at the opportunities ahead for collaboration.

Perspectives from The GovLab

By Dr. Stefaan Verhulst at The GovLab

Strategic Data Stewardship: The Role of the Private Sector in Shaping the Future of Data Collaboration

In recent years, [the private sector](#) has recognized the potential of leveraging its vast data resources for the public good. This endeavor, while worthwhile socially and able to contribute to a company's bottom line, poses both opportunity and risk. Until now, fear of these [risks](#) has outweighed a desire to seize the opportunities. This has in turn limited the number of collaborations for data reuse initiatives in the private sector.

[Data stewardship](#) is a key principle and concept that can help overcome industry's (often understandable) risk aversion. Data stewardship, when operationalized, represents a critical strategy or role within an organization, focusing on maximizing the utility and impact of data assets for both the organization and the public. This role, which can be held by an individual or a group, involves developing and implementing responsible data practices and ensuring ethical and legal compliance in data management. Data stewards are key in identifying and fostering internal and external collaboration opportunities, thereby enhancing the sharing and utilization of data.

In this essay, I explore how the private sector can nurture and encourage data stewardship, unlocking the social and business potential of data. First, I examine four key steps that companies can take, focusing on making the practice of data stewardship more [systematic, sustainable, responsible, and networked](#). I then discuss the critical role that can be played by [data collaboratives](#), an emerging form of inter-sectoral partnership, typically between the private and public sectors, that can help advance the practices of data sharing and data reuse. Finally, I conclude with some thoughts on building what I call a Data4Good Stack—an integrated set of tools, resources, and practices to help promote data stewardship and the reuse of privately held data for the public good.

Key Steps to Advance Data Stewardship

Data stewards have been a feature of the corporate landscape for several years now, sometimes and mostly operating under a different nomenclature. Despite increased recognition among industry leaders of their value, data stewards have often had trouble advancing their goals. This difficulty can be attributed to a number of factors, including a lack of awareness of their critical role in both the first (providing access) and last (translating insight into action) mile of data initiatives, the absence of a clear definition or authority when it comes to their responsibilities, as well as the previously mentioned risk aversion that often leads companies to avoid data sharing and reuse.

At [The GovLab](#) (New York) and the [DataTank](#) (Brussels), both of which I co-founded, we have supported numerous organizations in advancing data stewardship in both the private, public and nonprofit sectors. Based on this practice and additional research, we have identified four key steps that companies can take to make data stewards more successful.

Systematic: The first step towards advancing data stewardship is to make its practice more systematic, for example by moving beyond pilot projects to integrate data sharing and collaboration into core business processes. These goals can be achieved by establishing a dedicated data stewardship office or function within organizations. Such a position or office would [specialize in requesting or providing access to data](#) and expertise to address public needs, ensuring that data sharing and collaboration are not just ad-hoc projects but integral parts of an organization's structure.

Sustainable: Sustainability, another key step, refers to the financial and environmental viability of data stewardship. Financial sustainability can be ensured by developing and strengthening

the [business case for data collaboration](#). This involves demonstrating how sharing data can lead to new business opportunities, innovations, and enhanced brand reputation, in addition to advancing societal goals. On the environmental front, data stewards should also work toward minimizing carbon footprints when leveraging data centers, aligning with global efforts to combat climate change.

Responsible: Developing policies and frameworks for responsible data stewardship is also essential. Responsible data stewardship encompasses preventing misuse and promoting beneficial use of data. It requires robust [data responsibility policies](#) to address privacy, security, and ethical considerations. Responsible data stewardship also involves developing standards and protocols that ensure data is used in a way that benefits society while respecting personally identifiable information (PII) and community's interests.

Networked: Establishing networks between and among data stewards can play a vital role in encouraging data stewardship and data sharing. Such [networks](#), which may take a physical or virtual form, are akin to professional trade associations or other similar groups and could include as members both data stewards themselves and those tasked with overseeing data stewards within organizations. Among other benefits, networks of data stewards can serve as platforms for exchanging best practices, exploring partnership opportunities, and collectively addressing challenges and limitations.

Perspectives from The GovLab

Data Stewards and Data Collaboratives

As data stewardship becomes increasingly institutionalized, a new operational model for accessing data is emerging as central to its practice: the [data collaborative](#). As mentioned, data collaboratives refer to a cross-sectoral partnership that enables the sharing of data held by an entity (typically in the private sector), to be used by a third party (often in the non-profit or public sectors). If data stewards are the individuals or groups that initiate and oversee data sharing, then data collaboratives are the vehicles for the exchange: they effectuate the policies and directions of data stewards. As such, the form and structure of data collaboratives, about which we have written [extensively elsewhere](#), are central to the practice of data stewardship.

As with data stewards, the evident potential of data collaboratives is often [limited by a range of factors](#). These include a lack of awareness as well as concern over perceived costs. Three key steps can help organizations more effectively implement data collaboratives and thus advance the role of data stewards.

Lower Transaction Costs: Lowering the transaction costs of establishing partnerships is crucial for encouraging data collaboration. This goal can be achieved in a number of ways, for example, by developing template licensing or [data sharing agreements](#) that simplify the legal and administrative processes involved in data collaboration. In addition, facilities and resource sharing (e.g., cloud computing or technical capacities) can also help lower costs.

Raise Awareness: The potential of data collaboratives is often unrealized simply due to a lack of awareness. When individuals, organizations, governments, and other stakeholders are more informed about the positive outcomes and opportunities offered by data collaboratives, they are more likely to actively engage in such initiatives. Greater awareness can also mitigate concerns about data privacy and security, as well as foster a culture of transparency and trust among collaborators.

Data Collaboratives as a Key Pillar of ESG: Data collaboratives, and data stewardship in general, can [significantly enhance](#) a company's Environmental, Social, and Governance (ESG) profile by fostering responsible and sustainable business practices. Through collaboration with external stakeholders, especially in the non-profit sector, companies can share (and access) valuable data resources that allow them to make informed decisions and design more environmentally friendly and socially responsible strategies. By participating in data collaboratives, businesses can therefore not only improve their ESG performance but also demonstrate their commitment to a more sustainable and responsible corporate agenda, which is increasingly valued by investors, customers, and other stakeholders.

Conclusion: Toward a Data4Good Stack

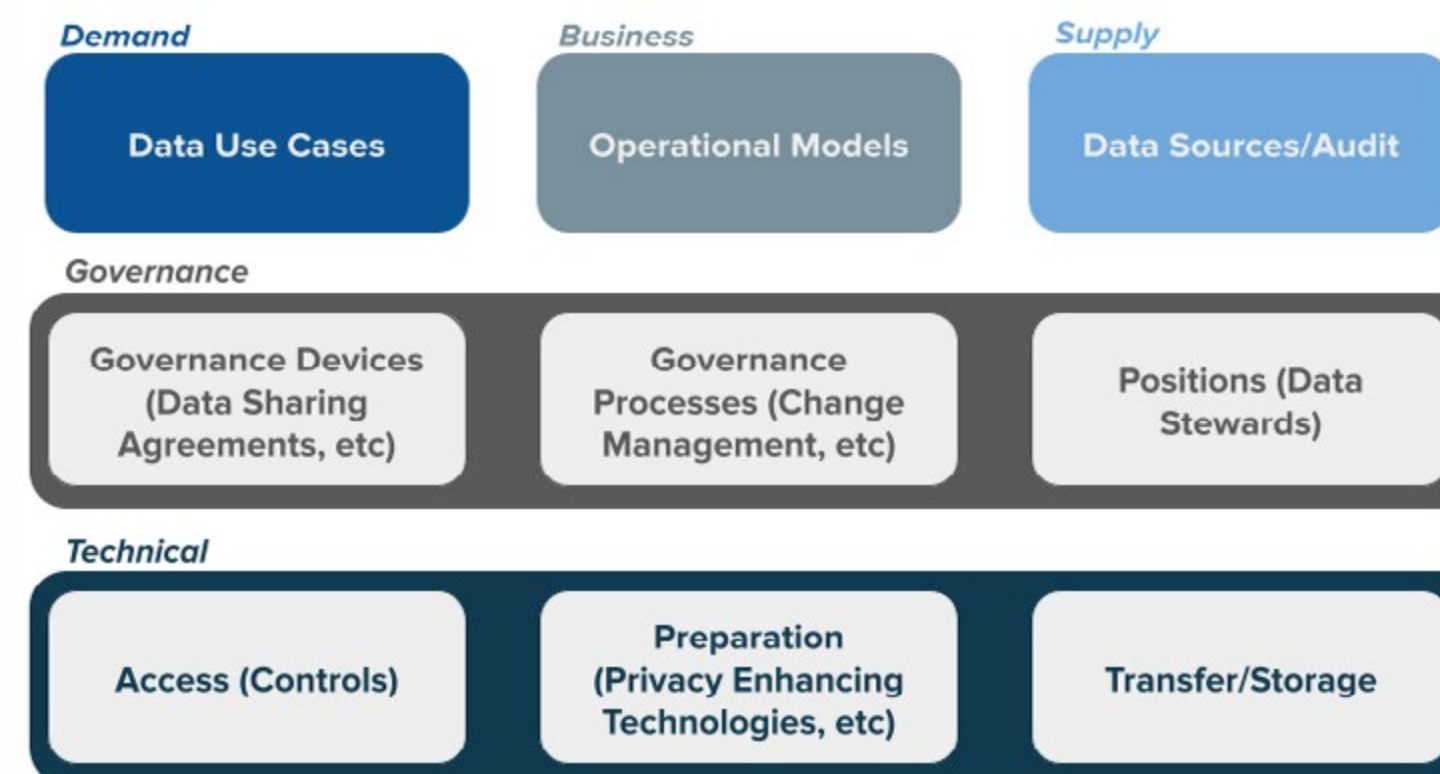
The private sector plays a crucial role in advancing data stewardship for the public good. By adopting a systematic, sustainable, and responsible approach to data stewardship, companies can not only enhance their own business operations but also contribute significantly to societal well-being. Establishing dedicated data stewardship functions, lowering transaction costs, developing a strong business case for data collaboration, and integrating data stewardship into corporate ESG strategies are pivotal steps towards realizing this vision.

In conclusion, we suggest that many of these goals can be implemented through an integrated Data4Good stack. The concept of a Data4Good stack involves creating a suite of tools, best practices, and frameworks that would encourage collaboration and leverage data for the public good. This stack, which would include both technical and non-technical components, would guide organizations in implementing effective data stewardship strategies, learning from best practices, and ensuring they are practical, impactful, and aligned with broader

societal and corporate goals. For example, the stack could include guides on setting up different types of collaborations, non-traditional data sources that can be tapped into, and best practices from the field. It could also include tools for navigating specific barriers to data collaboration and other governance processes.

These are ideas toward an integrated implementation strategy to advance data stewardship. One thing is certain: as data continues to grow in both volume and social and economic importance, the role of the private sector in stewarding this valuable resource to fulfill its public interest potential will become increasingly critical. Data stewardship and data collaboratives offer two of the most promising avenues for fulfilling this potential.

TOWARD A DATA4GOOD STACK



Perspectives from the Open Data Institute

By Josh D'Addario at the Open Data Institute

Going far by going together: The IDSP's role in driving private sector data sharing

By some accounts, open data in the UK is almost 14 years old now, following the launch of data.gov.uk in January 2010, though scholars and activists in the movement had been working on 'open' longer than that. But as the date we've chosen to start counting from entails, the original target of open data in the UK, and indeed the ODI, was opening government data. This seems an obvious choice, our taxes pay for the collection, management, and use of this data — why shouldn't we be able to access it?

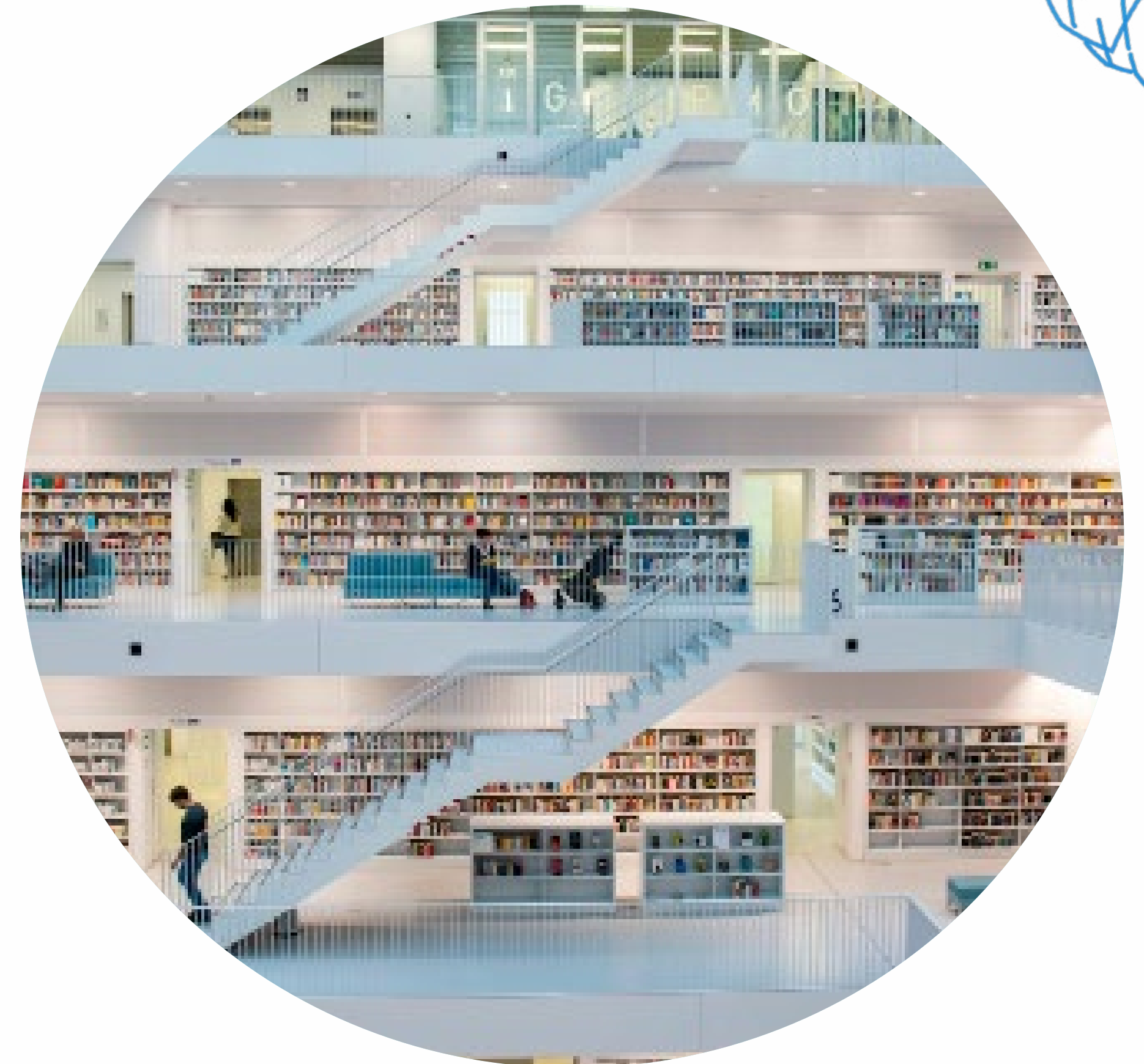
But we know that there is data out there held by other organisations besides the government that could provide a lot of benefit to society. Academia [sits on a wealth of research data](#) on every topic from astronomy to zoology, often making this information available through open access. Charities and philanthropics fund major projects around the world that [create valuable data, tools and insights for wider use](#). However, it is private sector companies that may be the holders of critical data for helping us combat crises such as climate, biodiversity, and the cost of living. As affiliate members of the [Industry Data for Society Partnership \(IDSP\)](#), we are working to turn this potential into a reality.

The value of data sharing in the private sector

The value of data can be an elusive concept. Some try to quantify the value of datasets based on econometric approaches, and speak of non-rivalry and externalities. Though that scholarship can be important, it can also be misleading, and it is important to remember that the value of data is really about how it can improve people's lives.

It is primarily from this standpoint that the ODI has been researching and advising to further unlock the data held by businesses and the consequent benefits. Our [reports](#) and [case studies](#) often demonstrate the potential value that sharing data can bring to the companies that hold it to encourage more open behaviour, which can ultimately lead to wider societal good. We know that increasing access to data held by the private sector can provide wider societal benefits by:

- **Increasing access to valuable services.** [The OpenActive initiative](#), led by the ODI and Sport England, helps publish data held by private companies in the fitness and leisure sector about the different products and services they offer. This has led to an increase in physical activity nationally, with [estimated health benefits](#) of up to 110 premature deaths and £3 million in health costs avoided per year.
- **Creating insights through data pooling.** [HiLo Maritime Risk Management](#) pools safety data shared by shipping companies from across the industry to provide safety recommendations based on insights from the data. This service supports shipping companies to operate more safely, having led to reduced lifeboat accidents by 72%, engine room fires by 65%, and bunker spills by 25%.
- **Addressing sector-wide challenges.** [National Grid ESO is working with Open Climate Fix](#) on its solar 'nowcasting' service for its national control room, which uses and publishes open data and code. This open data collaboration is leading to more accurate forecasts that could reduce the number of gas plants the National Grid relies on, aiding the shift to a net zero system by cutting out [100,000 tonnes of CO2 per year](#).



Perspectives from the Open Data Institute



ODI & IDSP

We are thrilled to see this collection of companies coming together to collaborate around new ways of leveraging open data infrastructure to drive benefits to society. To name a few of the highlights from the individual companies of the partnership we have seen that:

- UK Power Networks launched a world-class [open data portal](#), publishing data about the energy network infrastructure and usage, alongside examples of reuses and case studies to show researchers, decision makers, and potential innovators the existing value of these open datasets.
- Microsoft's [Planetary Computer](#) combines a multi-petabyte catalogue of global environmental data with open APIs, a development environment full of open-source tools, and numerous applications enabling users to power environmental sustainability and earth science programmes.
- Northumbrian Water Group published [the first open data strategy in the water sector](#) and is a leading force for the industry through both its annual [Innovation Festival](#) and especially as the lead organisation of the [Stream](#) initiative seeking to build a data institution focused on publishing data held by water companies openly.

But the sum of this partnership is bigger than its parts. As noted in the Data Challenge section, this was demonstrated when the IDSP and the ODI launched the [Data for Local Environments Challenge](#) to highlight the role publicly available data can play in better understanding how local governments can improve local environments, such as reaching their net zero goals.

In June 2023, a global judging panel decided upon the three winners of the Challenge:

- *Estimated Energy Attribution Certificates*, linked to Councils by Ayrton Bourn which aids UK local authorities in addressing challenges toward achieving net-zero by leveraging energy data from UKPN and Microsoft's data for renewable generation modelling, to estimate local authorities' total renewable generation and the consequential carbon emissions reduction.
- *InfraTracker, Open Data for Local Governments to Reduce Infrastructure Development-Co2 Emissions (Kenya)* by the *Demography Project*, which integrates community-centred public engagement, open data from Motor Vehicle Registration, GIS, and GHG emissions, supporting local and county governments in curbing carbon emissions and minimising environmental impact in road construction.
- *Council Climate Action Scorecards (UK)* by *Climate Emergency UK* which pioneers citizen-driven evaluation of UK councils' journey to net zero by consolidating comprehensive data on all feasible and implemented measures, uniting the local climate action movement and addressing the existing fragmented, sector-specific data issue.

Perspectives from the Open Data Institute

The Road Ahead

The benefits and opportunities are clear and within reach, and the IDSP is an important vehicle to help us, and others like us, achieve our ambitions to use data for societal impact. Looking forward, we have identified three high impact opportunities to take the next steps.

It is an exciting time to be working with like-minded organisations in the private sector. The opportunities to improve people's lives are in front of us, and we have the tools to unlock them together. The ODI is thrilled for the next year of collaboration with the IDSP.

1

An expanded focus on the environment and net zero

The Data for Local Environments Challenge was a great test of the value and demand for data about this topic, already held by IDSP members. Making more data available and bringing in new members with complementary datasets will only increase impact in a critical area. Bringing a public health lens into scope, as it relates to environmental health and the climate crisis, as well as including organisations in the Global South can greatly improve the impact of data collaboration in this area.

2

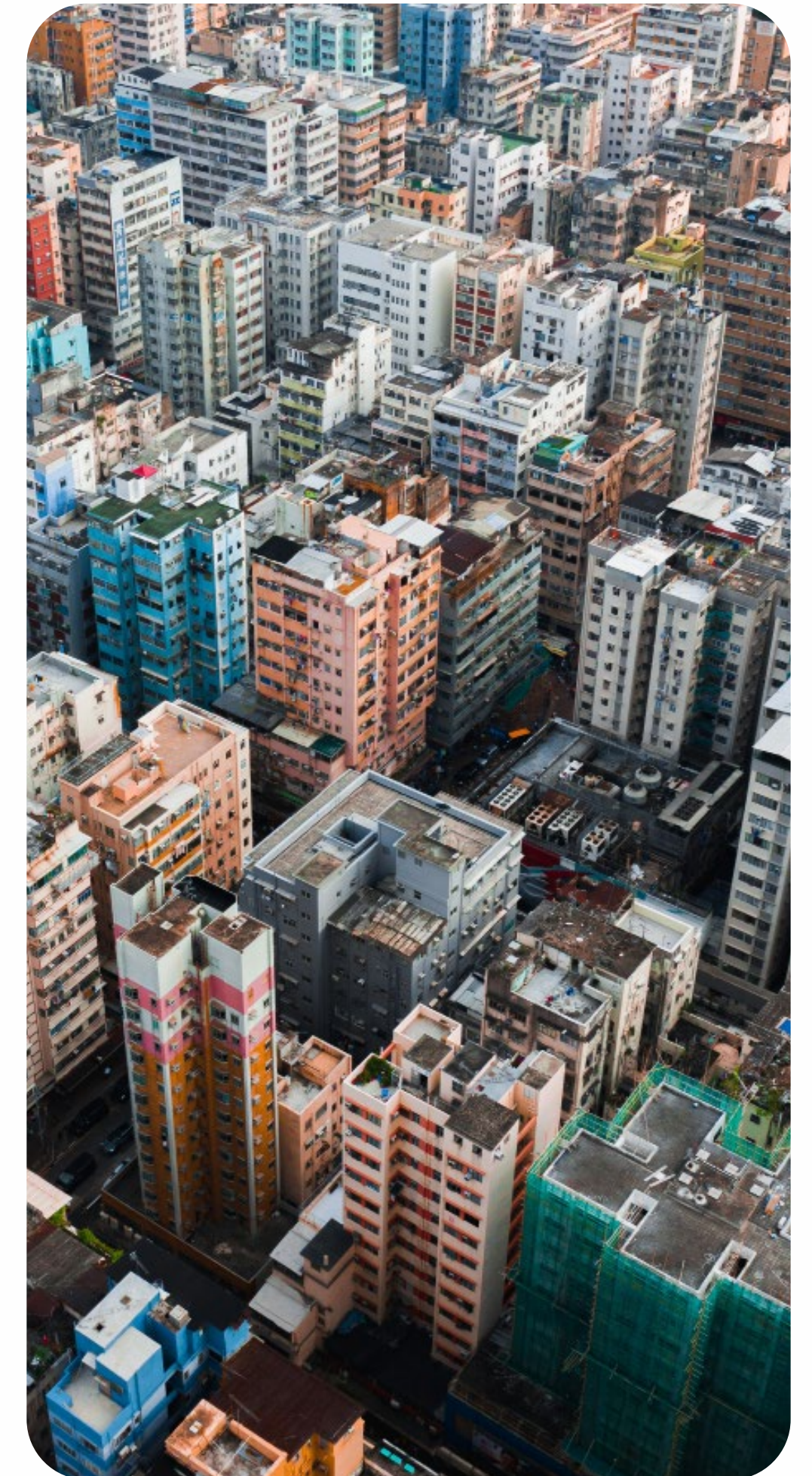
Showcase the impact of open and shared data

IDSP members and affiliates should create and collate case studies of how open and shared data can create benefits for consumers, communities, the environment and even data holders themselves. This can help to encourage other organisations beyond the IDSP to follow their lead, or even seek to join the partnership in order to collaborate with them.

3

Review novel ways that IDSP members can increase access to the data they hold

This could be done by researching the ecosystem of [privacy-enhancing technologies](#) that would allow for more data to be published or shared with a reduced risk of exposing sensitive and/or personal data through different anonymisation and aggregation techniques. This could also include exploring the world of [Smart Data](#). Countries such as the UK are [currently experimenting with 'Smart Data'](#) — secure, consented data sharing between organisations, enabling them to offer innovative services to consumers and small businesses. IDSP members are well placed to support this effort.





FUTURE DIRECTIONS

Future Directions and Opportunities

As the IDSP looks ahead to its next phase, our primary focus remains on providing greater access to private sector data to help develop solutions to societal challenges. We do this by mapping to our commitments to make data more open and accessible, share knowledge, and invest in skills development while protecting individuals' privacy.

The IDSP was launched in 2022 with the vision of creating a network of industry leaders who are committed to advancing the public good through the responsible use and sharing of private sector data. The partnership is excited about the progress and contributions made in its inaugural year, demonstrating the benefit of opening up private sector data to develop solutions for today's challenges. Highlights include:

- Launching the Data for Local Environments Challenge, which focused on harnessing publicly available data to empower local governments in enhancing their environments.
- Partnering with the National Academies to plan the U.S. Research Data Summit, which brought together leaders and experts from various sectors to discuss the opportunities and challenges of research data.
- Engaging with the global research community through events and publications, showcasing the IDSP's work and vision for the future of open data in the private sector.
- Showcasing industry insights and impact from the partnership's members, such as GitHub's Innovation Graph, Northumbrian Water's Open Data Strategy, and the Open Energy Data Maturity Framework hosted by UKPN.

The IDSP aims to build on the progress achieved in its inaugural year, with its primary objective to facilitate private sector data sharing between the private sector and data practitioners, policymakers, civil society organizations, and academia to tackle societal challenges. To support the primary objective, the partnership will develop best practices, showcase impactful initiatives, and expand its network and outreach.

Additionally, recognizing the importance of collaboration and knowledge exchange, the IDSP will continue to partner with its affiliates, The GovLab and the Open Data Institute (ODI), as well as work to engage new members.

Guided by the partnership's commitments of opening, or expanding access to private sector data, disseminating knowledge and insights, investing in the development of a diverse group of professionals, and safeguarding individuals' privacy throughout all of these activities, we are pursuing the following activities in 2024.

Create Practical Resources for Guiding the Private Sector:

The IDSP is committed to creating practical resources tailored for the private sector, offering guidance to like-minded companies that are exploring ways to make their data more accessible for societal benefit. The partnership aims to facilitate and streamline the process of data sharing, promoting a collective effort in addressing pressing societal issues.

Engage the Global Research Community: The IDSP is committed to active dialogue and collaboration with the broader research community, and will continue to build international connections to share and learn about the newest advances in making data open or more accessible to solve societal challenges. The partnership will continue its engagement with the National Academies and the U.S. Research Data Summit, including disseminating the Proceedings of a Workshop slated for publication in early 2024. Additionally, the IDSP will explore potential collaborations emerging from the summit while actively participating in global research conferences such as the Data for Policy Conference 2024.

Provide case studies and best practices for the application of private sector data: The IDSP aims to highlight how private sector data can be applied to societal challenges, as demonstrated through our Data for Local Environments Challenge. We will continue to showcase how data supply meets demand in areas such as sustainability, economic opportunity, and skilling, and how the hurdles to data access can be addressed through governance and technical measures.



Join Us in Shaping a Data-Driven Future

In wrapping up our Year in Review, the partnership invites private sector organizations to contact the IDSP if you are interested in learning more or becoming involved. The IDSP has made progress against our commitments in the past year, but there is more to do, and we know we can learn more from others.

The partnership provides a network of companies committed to harnessing private sector data for the betterment of society. Member organizations play a vital role in amplifying the impact of collective efforts and advancing the essential role of private sector data in addressing complex societal challenges.

To learn more, contact us at info@industrydataforsociety.com or explore our Industry Data for Society Partnership [website](#).



2023 YEAR IN REVIEW