UNLOCKING DATA: IDENTIFYING NEEDS & COLLABORATIVE APPROACHES

Data Series: Session #3

Data Collaboration and Governance

Stefaan G. Verhulst

Monday, November 6, 2023

THEGOVLAB innovate(us)

ABOUT THE GOVLAB

GOVLAB

The Governance Lab (The GovLab) is an action-oriented research center that seeks to improve people's lives by changing how we govern using new technologies.

Learn more at: govlab.org.

The **Open Data Policy Lab** is a resource hub supporting decision-makers as they work toward accelerating the responsible reuse and sharing of open data for the benefit of society and the equitable spread of economic opportunity.

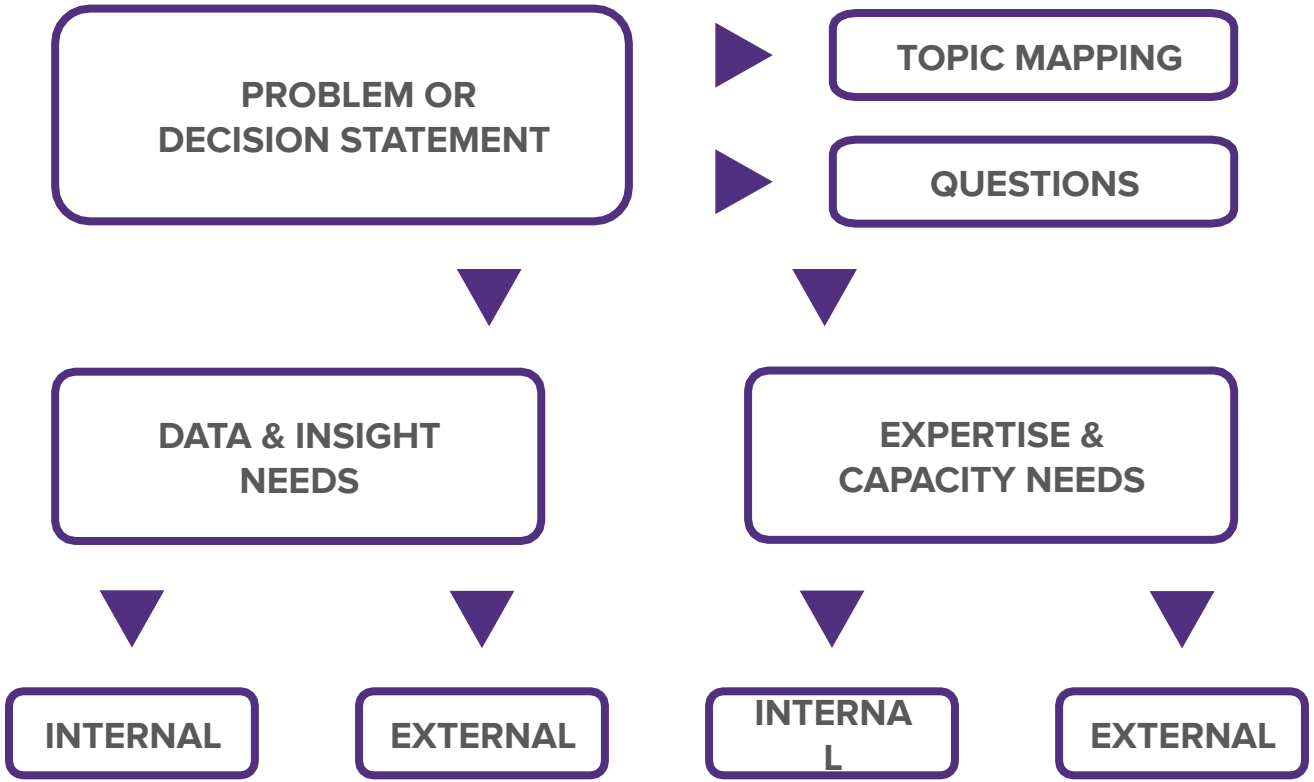
Learn more at: opendatapolicylab.org.

DATA CAN HELP TO INNOVATE HOW WE SOLVE PUBLIC PROBLEMS

DATA & DATAFICATION

1. Datafication has transformed the data landscape.
2. Data is not a thing, it is a process.
3. Data in itself has no value.
4. Data is never “raw” or “born” – it is created.
5. Data re-use provides the real opportunity.
6. Data likely resides elsewhere.
7. Metadata is the new data.

GOING FROM PROBLEM TO INSIGHT



GOING FROM PROBLEM TO INSIGHT

1. Start with the problem, not the data.
2. A participatory process can enable a new kind of “question science”.
3. Different questions facilitate different types of insights.
4. Focus on the minimum viable data.
5. Without the right expertise, the data’s impact will be limited.

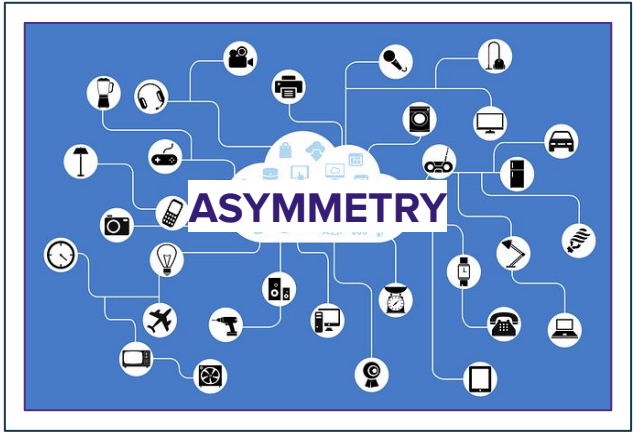
TODAY'S CHALLENGE: THE EMERGENCE OF DATA ASYMMETRIES

Volume

MIS-USE

Veracity

DEMAND



SUPPLY

Velocity

RE-USE

Variety

Premise 1:
Data collaboratives can help bridge data asymmetries.

DATA COLLABORATIVES: MATCHING DEMAND & SUPPLY

DATA COLLABORATIVES: MATCHING DEMAND & SUPPLY

"The term data collaborative refers to a new form of collaboration, beyond the public–private partnership model, in which participants from different sectors—in particular companies—exchange their data to create public value."

— Stefaan Verhulst, "Data Collaboratives: Exchanging Data to Improve People's Lives"

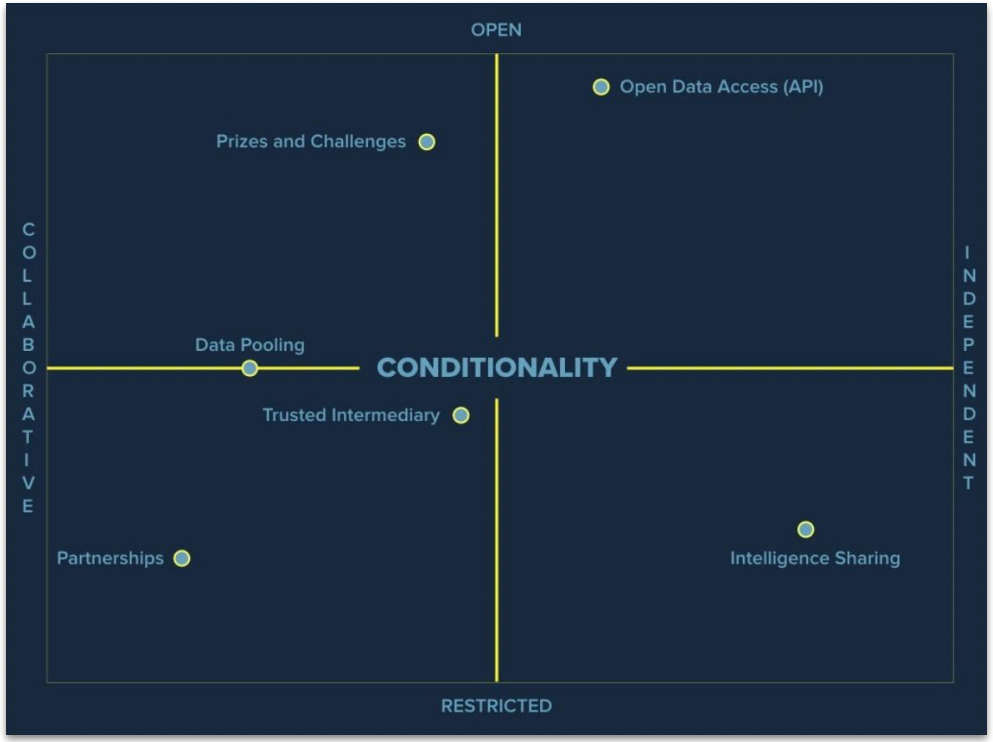
Premise 2:
When establishing data collaboratives, consider what is fit for purpose.

THE MATRIX OF CONDITIONALITY

Y-axis: Level of Openness
Open vs Restricted

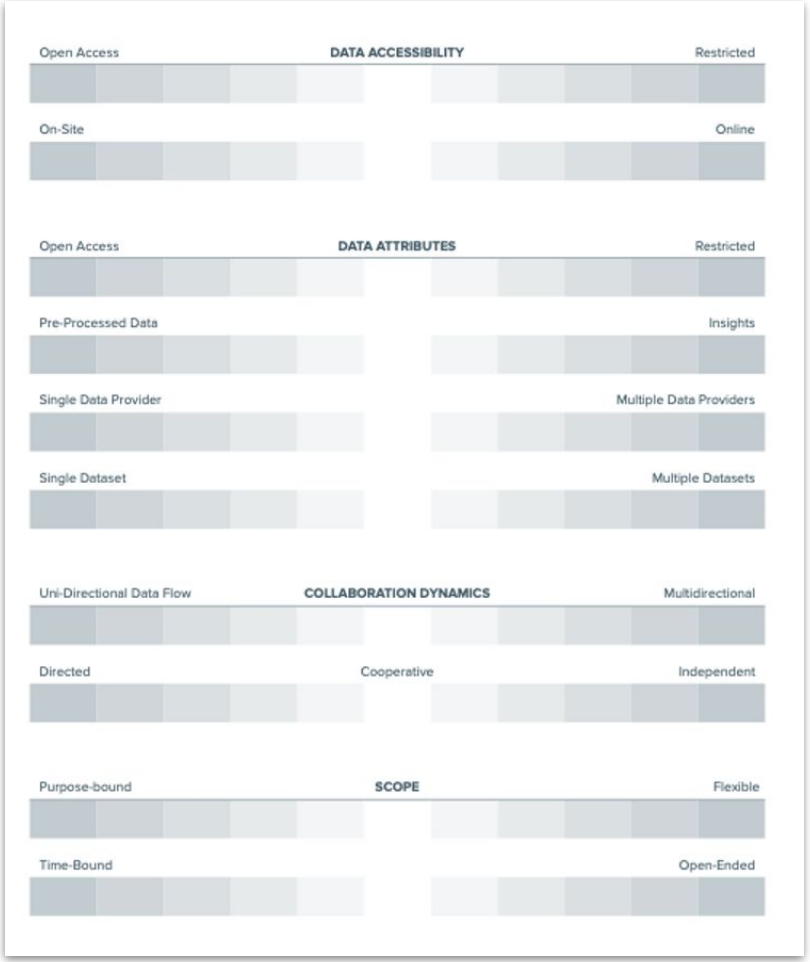
X-axis: Level of Collaboration
Collaborative vs Independent

	Independent Use	Cooperative Use	Directed Use
Open Access	Public Interfaces	Data Pooling	Prizes & Challenges
Restricted Access	Trusted Intermediary	Research & Analysis Partnership	Intelligence Generation



DETERMINING WHAT IS FIT FOR PURPOSE

- **Data Accessibility:**
  - Level of Openness
  - On-Off Line
- **Data Attributes:**
  - Raw/Pre Processed vs Insights
  - Single vs Multiple Data Providers
  - Single vs Multiple Data Sets
- **Collaboration Dynamics:**
  - Uni- vs Multi Directional
  - Directed vs Independent
- **Scope:**
  - Purpose-bound vs Flexible
  - Time-bound vs Open Ended
  - Budget



Premise 3:
Different data collaborative models are effective in different contexts.

DATA COLLABORATIVES: OPERATIONAL MODELS

- Public Interfaces
- Data Pooling
- Prizes and Challenges
- Trusted Intermediary
- Intelligence Generation
- Research and Analysis Partnerships

PUBLIC INTERFACES

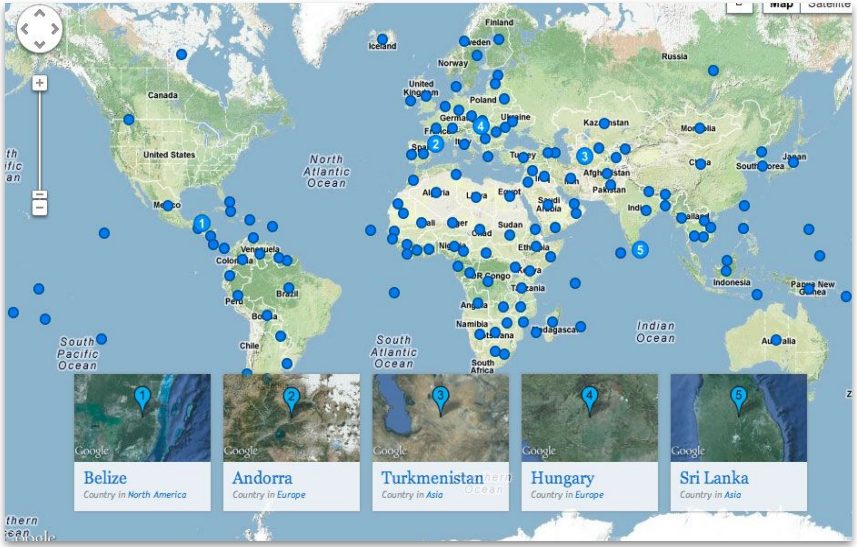
A single data holder provides access to certain types of pre-processed and/or data-driven tools for public use.

Data Holders: One data holder
Data Users: Many data users

Main Types:

- Application Programming Interfaces (APIs)
- Data Platforms

Google Earth Outreach



Public Interfaces:
Google Earth Outreach

API: GOOGLE EARTH OUTREACH

The Ashoka Trust for Research in Ecology and the Environment (ATREE) in India needed data to better understand how to protect tigers and elephants living in the Kalakkad Mundanthurai Tiger Reserve (KMTR).

Planning

Collecting

ATREE sought site specific information on their conservation efforts in the KMTR, including climate data, field data and socioeconomic information.

Processing

Google Earth Outreach provides a range of data and mapping tools to enable users to run analyses around specific research questions.

Sharing

ATREE uses Google Earth Outreach's public data catalogue to combine satellite images with maps of the KMTR.

Analyzing

Researchers stitch together satellite images from clear to cloud-covered days, and then use a masking algorithm to remove the cloud cover. This enabled them to build accurate maps of the KMTR.

Using

ATREE's maps inform policy making to protect tigers and elephants and are used to measure the impact of sustainable practices.

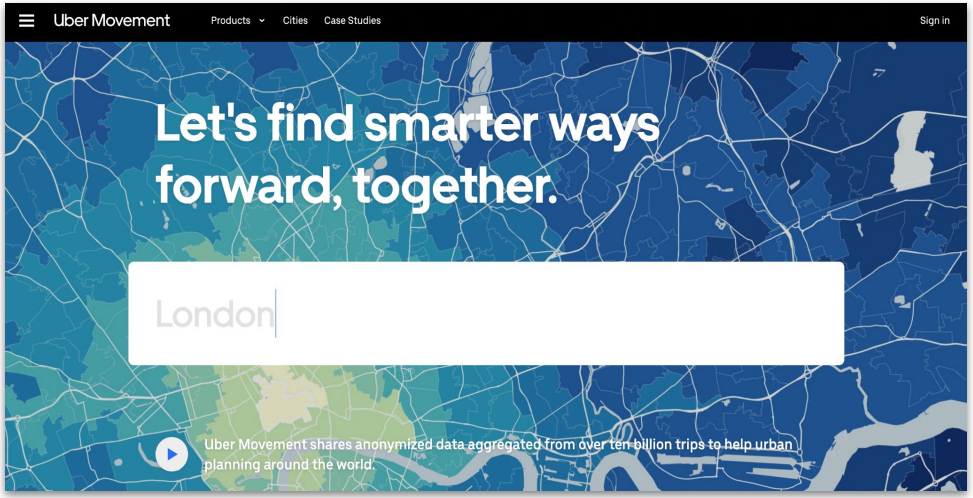
Kalakad Mundanthurai Tiger Reserve

Google Earth Outreach

ASHOKA TRUST FOR RESEARCH IN ECOLOGY & THE ENVIRONMENT

DATA PLATFORMS

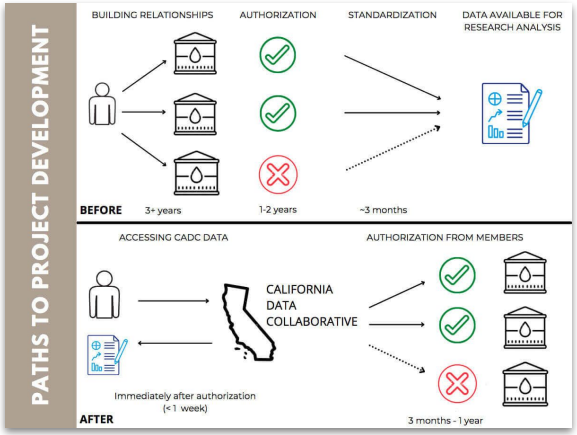
Data platforms make private-sector data assets and tools accessible to the general public through web or mobile applications. Often, companies develop these platforms with certain users in mind (e.g. humanitarian actors or city planners), but data assets are generally made accessible to any user.



Data Platforms:
Uber Movement

"We've gotten consistent feedback from cities we partner with that access to our aggregated data will inform decisions about how to adapt existing infrastructure and invest in future solutions to make our cities more efficient. We hope Uber Movement can play a role in helping cities grow in a way that works for everyone."
– Uber Movement

CALIFORNIA DATA COLLABORATIVE



Data Pooling:
California Data Collaborative

DATA POOLING

Data holders pool datasets as a collection designed to be accessible by multiple parties.

Data Holders: Some data holders
Data Users: Some data users

Main Types:

- Public Data Pools
- Private Data Pools

PRIVATE DATA POOL: CALIFORNIA DATA COLLABORATIVE

Water supply in California is strained but utilities lack clean, reliable data outside their jurisdiction that can help them make decisions and follow regulation.

Planning

Collecting

Different water utilities agree to form a nonprofit that hosts a private pool of their different water-meter data collections. Only collaborative members have access to the raw collection.

Members to the collaborative maintain a central staff who maintain and clean data uploaded for analysis and visualization.

Processing

Sharing

New water utilities can apply to join, after paying a fee. If they are accepted by existing members, they can, in turn, opt to share data with verified research partners.

Water managers use the data compilation to answer specific governance questions or seek support from CaDC staff.

Analyzing

Using

Water managers act to realize new water efficiencies based on the analysis provided.

City of Thousand Oaks

Calleguas Municipal Water District

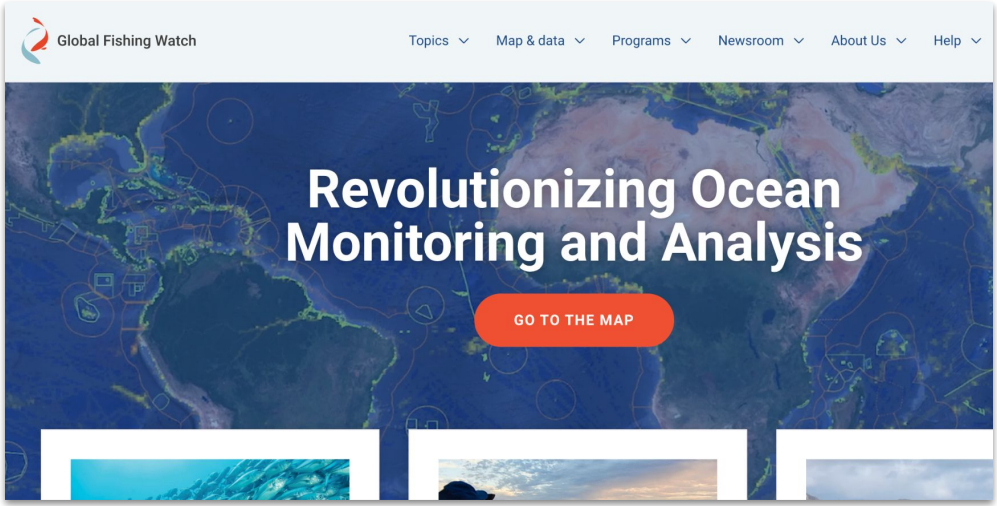
Water District

emwd Eastern Municipal Water District

City of Sacramento

PUBLIC DATA POOLS

Public data pools commingle data assets from multiple data holders and make those shared assets available on the web. Pools often limit contributions to approved partners, but access to the shared assets is open, enabling independent uses.



Public Data Pools:
Global Fishing Watch

"The Global Fishing Watch map is the first open-access online platform for visualization and analysis of vessel-based human activity at sea."
– Global Fishing Watch

PRIZES & CHALLENGES

Data holders make data available to participants who compete to solve problems or pioneer innovative uses of data for the public interest.

Data Holders: One or more
Data Users: Many data users

Main Types:

- Open Innovation Challenges
- Selective Innovation Challenges

Türk Telekom
Değerli Hissettirir

Prizes & Challenges:
Türk Telekom's Data 4 Refugees Challenge

SELECTIVE INNOVATION CHALLENGES: DATA FOR REFUGEES CHALLENGES

Syrian refugees in Turkey face a number of challenges from health and education to unemployment, safety and social integration that require data-driven interventions.

Planning

Collecting

The Data for Refugees (D4R) Challenge collected a database of anonymized mobile Call Detail Records of phone calls and SMS messages of Türk Telekom customers.

Processing

The data was cleaned and anonymized to ensure the privacy and safety of the refugees whose data was being shared.

Sharing

166 contestants, mainly data scientists and researchers, were given access to the dataset to solve urgent problems facing Syrian refugees.

Analyzing

Contestants combined computational analysis skills with domain expertise in social impact, especially work with children and refugees.

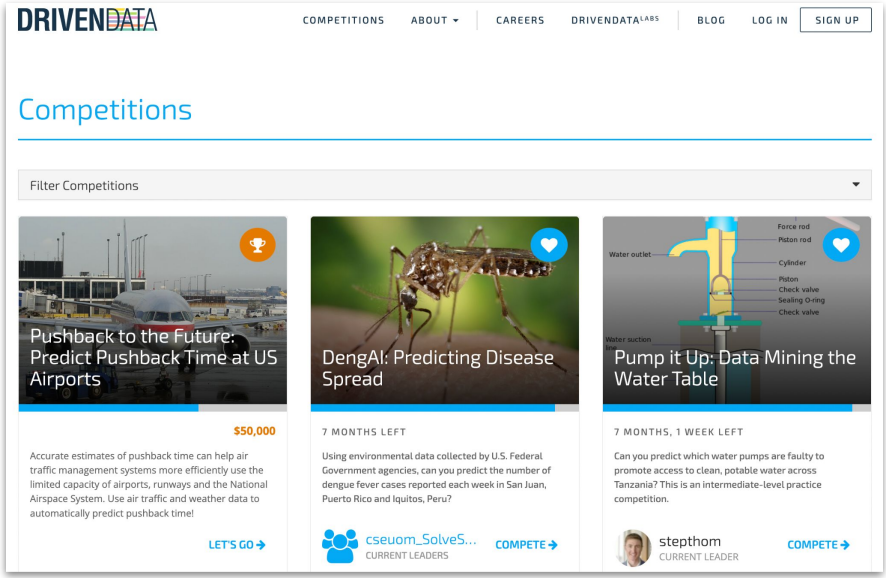
Using

The resulting solutions of the competition aimed to improve the well-being of Syrian refugees in Turkey.

IOM UN MIGRATION unicef UNHCR The UN Refugee Agency BOĞAZİÇİ ÜNİVERSİTESİ 1863 TÜBİTAK Türk Telekom Değerli Hissettirir

OPEN INNOVATION CHALLENGES

In **open innovation challenges**, companies provide open access to datasets to attract self-selecting participants to develop data-driven solutions, premises, or insights to public challenges.



Open Innovation Challenges:
DrivenData Challenge Platform

"We host online challenges, usually lasting 2-3 months, where a global community of data scientists competes to come up with the best statistical model for difficult predictive problems that make a difference."
– DrivenData

TRUSTED INTERMEDIARIES

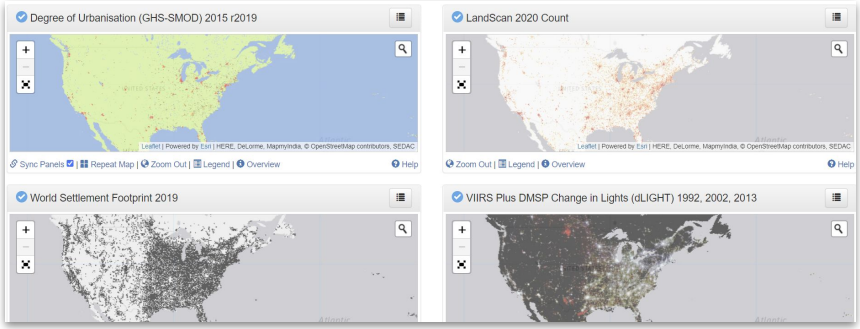
Third-party actor mediates collaboration between (private sector) data providers and data users from the public sector, civil society, or academia.

Data Holders: One or more
Data Users: One or more

Main Types:

- Data Brokerage
- Third-Party Analytics

POPGRID DATA COLLABORATIVE



Trusted Intermediaries:
The POPGRID Data Collaborative

DATA BROKERAGES: POPGRID DATA COLLABORATIVE

Columbia University noted a lack of consolidated georeferenced data on population, human settlements, and infrastructure for research.

Planning

Collecting

It creates POPGRID, which provides a platform for 10 research centres and government institutions to exchange georeferenced data for their benefit and the public's benefit.

Individual institutions clean the data based on their own methodologies in ways to allow side-by-side comparison. POPSTAT ensures inputs are high quality.

Processing

Sharing

The data is exchanged and made viewable through the POPGRID Viewer, a dashboard that allows visitors to examine and compare different inputs.

Researchers from various institutions use this collection to develop new research on human mobility and settlements.

Analyzing

Using

Organizations like the US Census Bureau use the insights to inform their policies and estimates.

Columbia Climate School
Center for International Earth Science Information Network

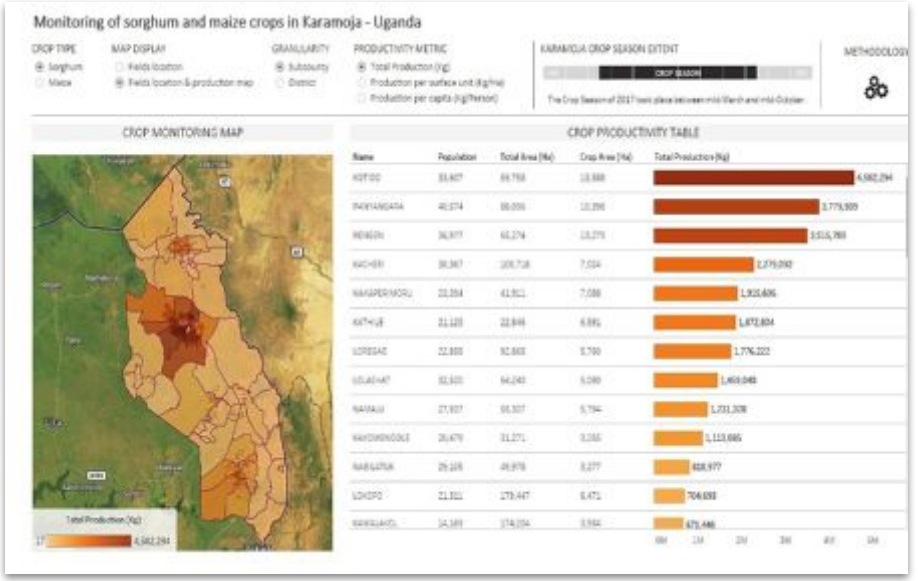
CUNY Institute for Demographic Research

internet.org by facebook

esri
THE SCIENCE OF WHERE

THIRD PARTY ANALYTICS

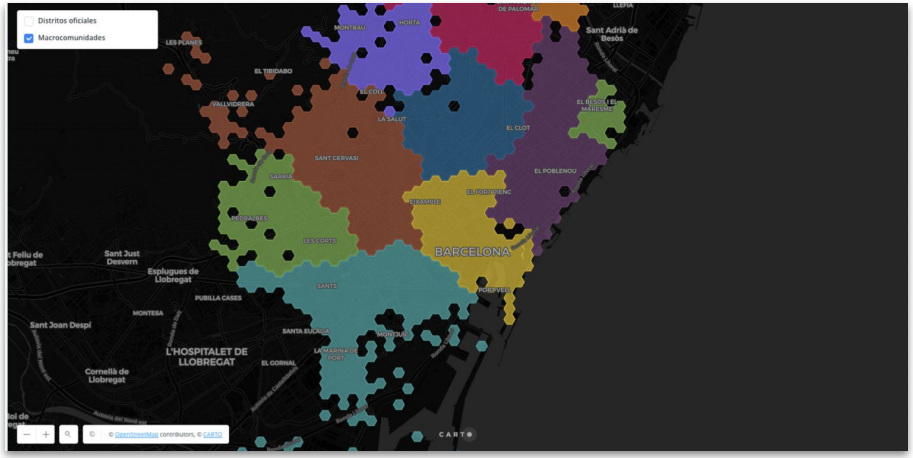
Third-party analytics projects see trusted intermediaries access private-sector data, conduct targeted analysis, and share insights, but not the underlying data, with public or civil sector partners. This approach enables use of data in the public interest while retaining strict access controls.



Third Party Analytics:
Dalberg Food Security Manager

"We built an interactive tool that tracks mobile phone top-up spending as a non-food household expenditure, and market prices of crop staples across the Greater Kampala Metropolitan Area."
– Dalberg Analytics

BBVA CARTO



Intelligence Generation:
BBVA Urban Discovery

INTELLIGENCE GENERATION

No data is shared with external parties, instead the results of analysis within the data holder's organization are shared with external actors.

Data Holders: One data holder
Data Users: No external data users

INTELLIGENCE GENERATION: BBVA URBAN DISCOVERY

The way a city is mapped plays an important role in its development. BBVA sought to re-imagine how we visualize urban settings to better reflect the ways residents live in the cities.

Planning

Collecting

BBVA Data & Analytics collected over 413 million card transactions from Madrid, Barcelona and Mexico City.

BBVA and CARTO used algorithms to develop a new urban configuration based on consumers' most common movements.

Processing

Sharing

BBVA and CARTO used the results of their analysis to develop an interactive data visualization.

The Urban Discovery tool allows users to re-imagine maps of the three cities based on factors relevant to their lens.

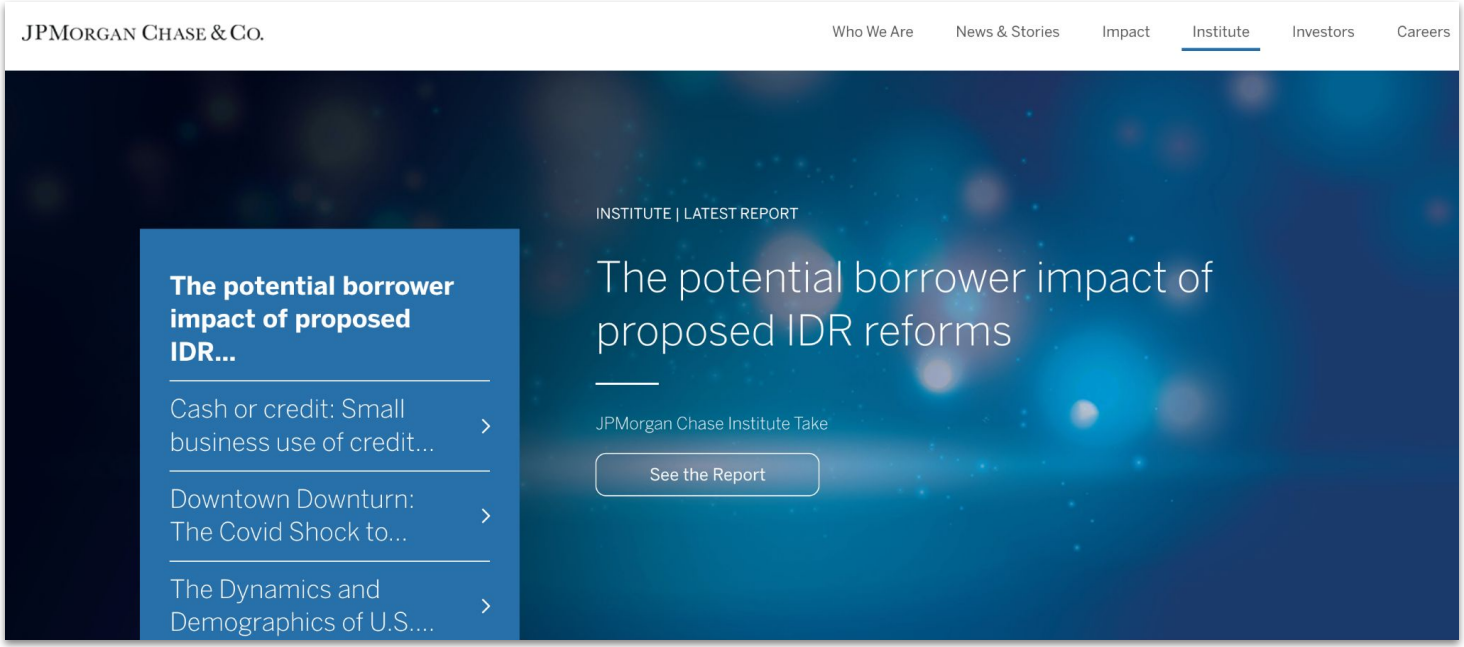
Analyzing

Using

Urban Discovery enables stakeholders to gain an understanding of the on the ground reality of a situation, and the impact of policy interventions.

BBVA CARTO GOBIERNO DE LA CIUDAD DE MÉXICO Ajuntament de Barcelona Comunidad de Madrid

INTELLIGENCE GENERATION



"Our data allow us to answer important questions about the financial health of US consumers and businesses at the national and local levels, as well as study labor and financial markets."

– The JP Morgan Chase Institute's Research Agenda

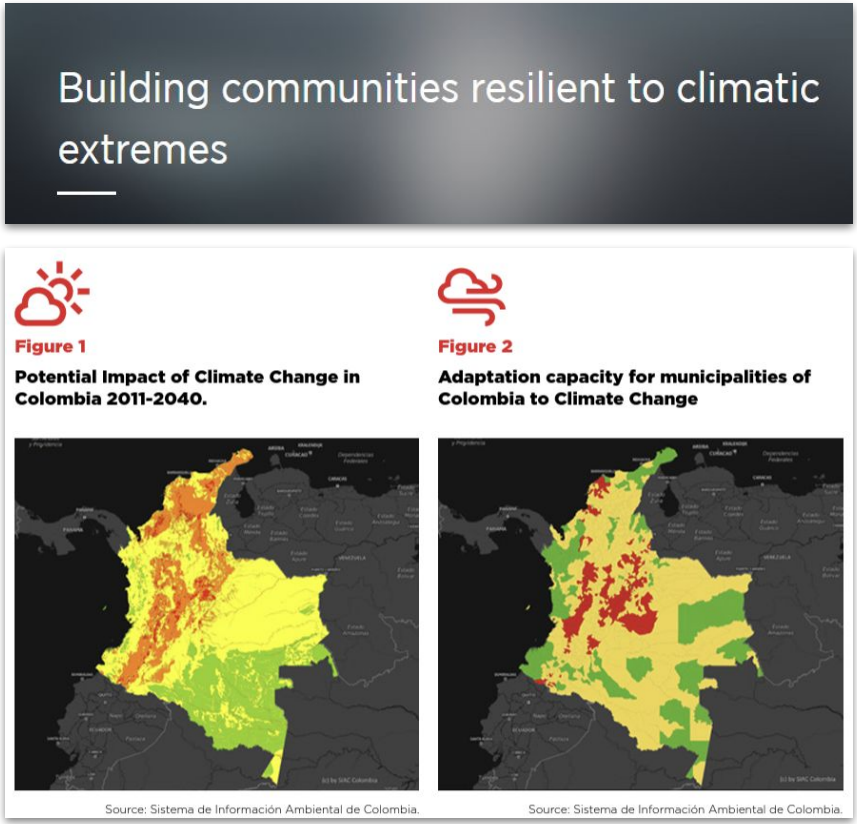
RESEARCH & ANALYSIS PARTNERSHIPS

A pairing between (private sector) data providers and data analysts or data users from the public sector, civil society or academia.

Data Holders: One or more
Data Users: One or more

Main Types:

- Data Transfer
- Data Fellowship



Research & Analysis Partnerships:
UN FAO's Building Communities Resilient to Climatic Extremes

DATA TRANSFER: UN FAO'S BUILDING COMMUNITIES RESILIENT TO CLIMATIC EXTREMES

UN FAO needed data to better estimate climate displacement in Latin America, particularly Columbia.

Planning

Collecting

It contacted Telefonica, a major telecom provider with Call Detail Records from 11+ million customers in Columbia.

Telefonica aggregated and anonymized its CDR data and combined it with relevant open and government data for FAO.

Processing

Sharing

Telefonica provided a platform for FAO and the Government of Columbia to monitor the data regularly.

Using a mobile data analysis platform, researchers found 12,000 people left the La Guajira region due to drought.

Analyzing

Using

FAO will use this insight to develop support for IDPs and better manage natural resources

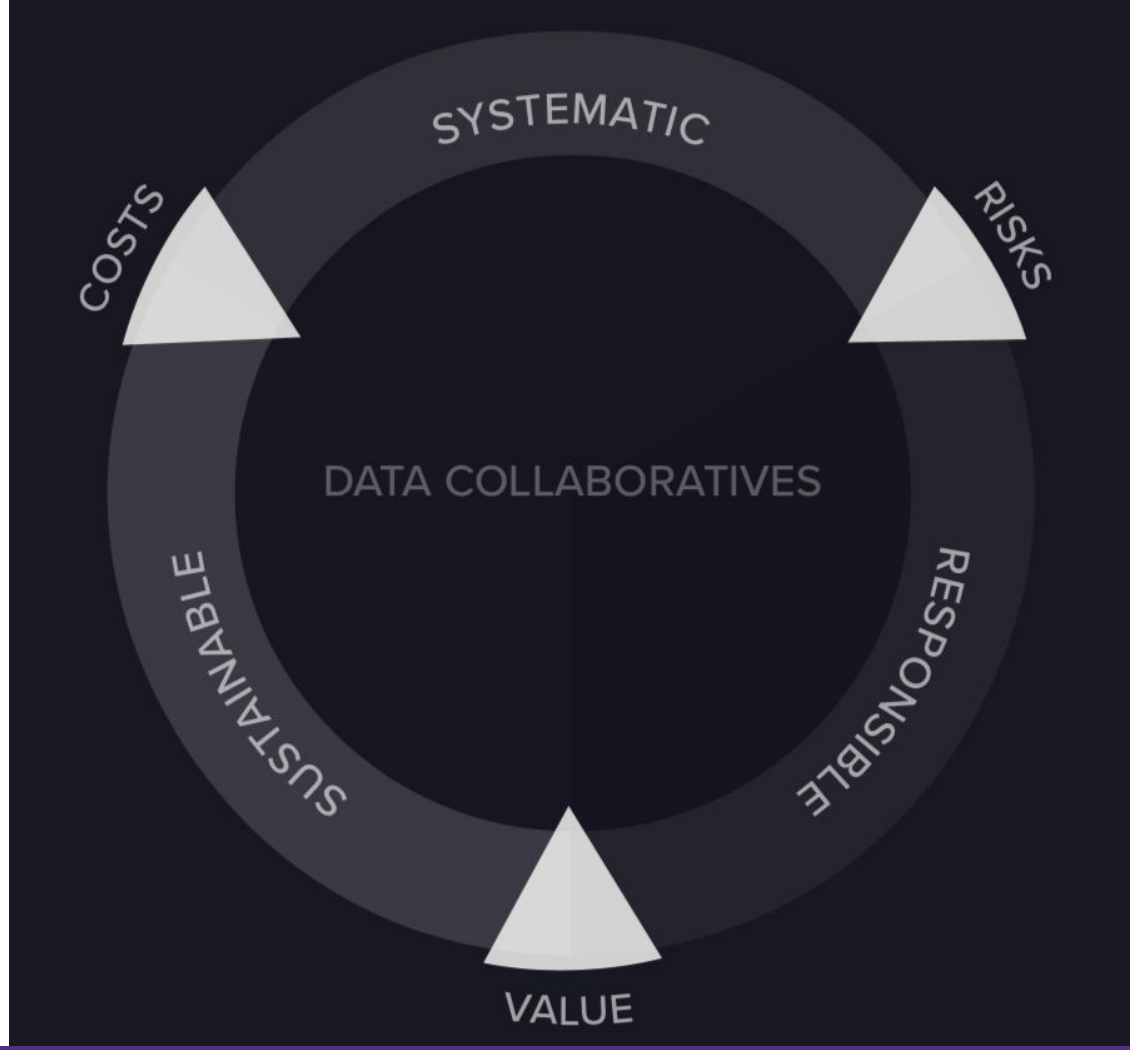
DATA FELLOWSHIP

With **data fellowships**, companies provide opportunities for their staff to support another organization. These opportunities often involve one organization's researchers being embedded in another organization to analyze their data assets or otherwise support their work.

Data Fellowship:
Google.org's Impact Challenges

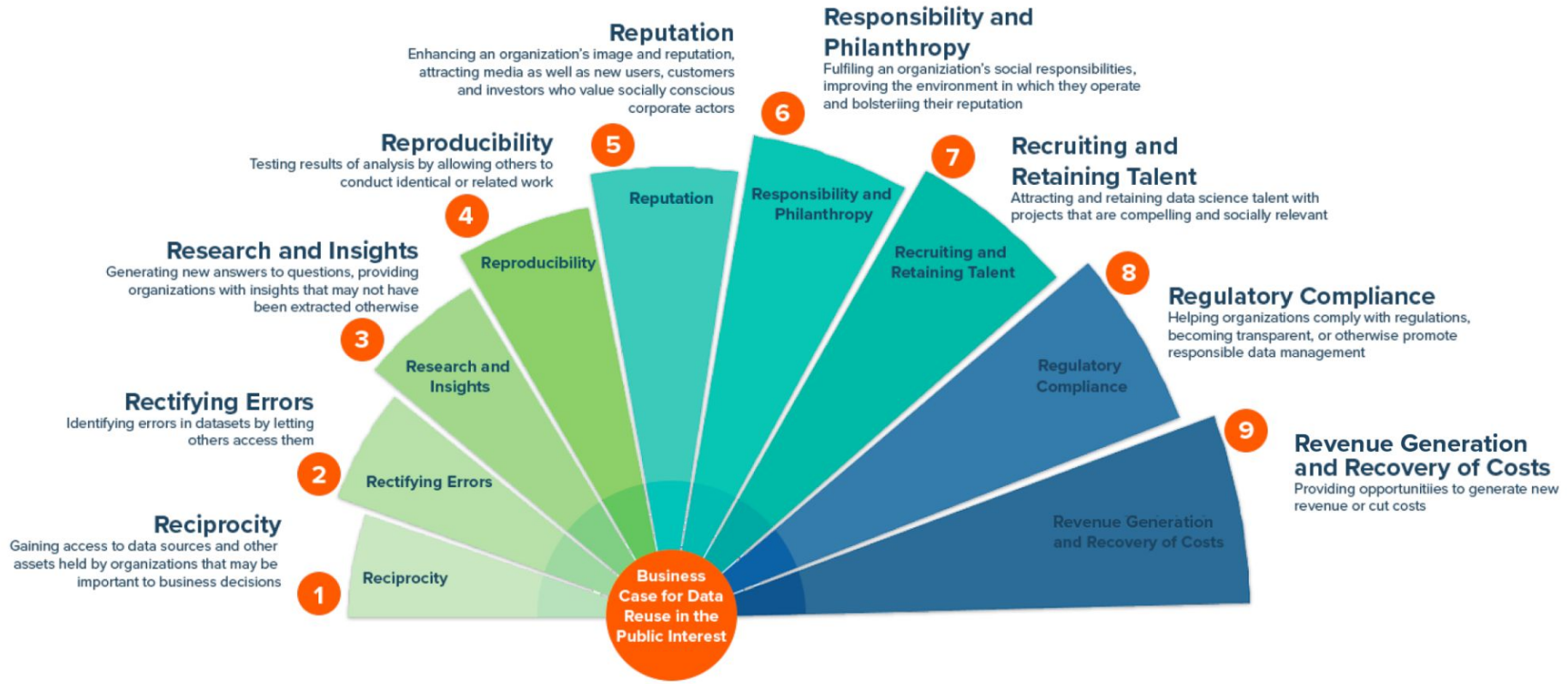
"Through Google.org Impact Challenges, we award nonprofits and social enterprises with support to help bring their ideas to life."
– Google.org

DATA COLLABORATIVES: THE CHALLENGE



Premise 4:
Determining the value proposition behind a data collaborative makes it more sustainable.

THE BUSINESS CASE FOR DATA RE-USE



RESEARCH & INSIGHTS

- Increased access to data can lead to new R&D and new methods that would not have otherwise been possible and at low cost.
- Development of new IPR/Models from the data that is shared
- Identify novel use cases and potentially new markets

Ask Yourself: By providing more access to its data, can your organization answer questions it might not have otherwise conceived or extract insights it might have otherwise lacked the resources, capabilities, or time to do so itself?

R3. Research and Insights

Definition: Increased access to data can lead to new research and insights that would not have otherwise been possible. By providing access to data, organizations can help advance the state of knowledge in their field. This, in turn, can help generate new questions, provide answers to old questions and lead to the development of innovative products and services.

Considerations: A major challenge for open data and data collaboration is ensuring they have sufficient data to inform meaningful innovation work. As one survey of open data projects notes, many initiatives are too technically focused and are "little more than websites linked to miscellaneous data files, with no attention to the usability, quality of the content, or consequences of its use" (Helbig, et al. 2014). Many organizations rely on "low-hanging fruit" that are easy to release but of little utility (Kitchin, 2013). The result is that data repositories become "data dumps" that cannot result in meaningful insights or action. This problem is closely related to the research–policy gap in which significant research effort is expended to generate insights but these insights are not acted upon or adopted (Martin, et al. 2019). It is often beneficial for organizations to have their work directed toward specific, tangible action. This problem is closely related to the research–policy gap in which significant research effort is expended to generate insights but these insights are not acted upon or adopted (Martin, et al. 2019). It is often beneficial for organizations to have their work directed toward specific, tangible action.

Ask yourself: By providing more access to its data, can your organization answer questions it might not have otherwise conceived or extract insights it might have otherwise lacked the resources, capabilities, or time to do so itself?

○ Very Irrelevant ○ Somewhat Irrelevant ○ Somewhat Relevant ○ Very Relevant ○ Most Relevant of All

Capture Notes Here:

Inspiration:

A desire for new insights that can yield solutions has long been a motivating factor behind open data portals. Swisscom, for example, used its reach across Switzerland to create an open data portal (Swisscom Open Data, 2022). The 44 anonymized datasets stretch across communication, mobility, network (mobile, railway and Swisscom hotspots) and sites (Swisscom shops). Created through the belief that innovation can be ignited through high-quality data, Swisscom shares data so that fresh, new thinking can be applied across Switzerland.

Additional Relevant Resources:

- #Data4COVID19 Review: Report from The GovLab, with support from the Knight Foundation, reviewing lessons from the use of non-traditional data during COVID-19
- The YODA Project Partnership with Johnson & Johnson: Short case study by The GovLab on YODA

RESEARCH & INSIGHTS

Why Solution Blog Team We're hiring Login Request demo

Announcing: The Circular City

by Team Numina / March 7, 2019

RECIPROCITY

- Reciprocity refers to what occurs when organizations that open their data gain access to:
  - Other data sources, and
  - Assets and expertise held by other institutions that could be important to business decisions.

Ask Yourself: By providing more access to its data, can your organization answer questions it might not have otherwise conceived or extract insights it might have otherwise lacked the resources, capabilities, or time to do so itself?

ROI Case #1: Knowledge and Insights: The ways that data reuse can create a return on investment by generating new insights that an organization would otherwise be unable to extract from its information if it kept its dataset closed and inaccessible to outside parties.

R1: Reciprocity

Definition: Reciprocity refers to what occurs when organizations that open their data gain access to other data sources and assets (e.g. expertise, technology) held by other institutions that could be important to business decisions. Huston, Edge and Bernier (2019) document how the real value of open data is not just about making datasets available for use. Often, opening datasets provides impetus for others to do so as well, creating a reciprocal data ecosystem and providing a multiplier effect that benefits all parties (Apheris, 2022). It can also lead to opportunities for data fusion in which one holder's dataset is enhanced by another, either by improving its accuracy or applicability for different use cases (Chatzichristos, et al. 2022).

Considerations: Organizations might want to be cognizant of the broader context in which their data is being used and how "reciprocity" might allow others to make unwanted insights. For example, they might think critically about whether reciprocal exchanges could contribute to the "mosaic effect" in which compilations of disparate datasets can be combined to reveal new sensitive insights that might undermine the privacy and wellbeing of individuals and communities (McInerney, 2020).

Ask yourself: By providing more access to its data, can your organization answer questions it might not have otherwise conceived or extract insights it might have otherwise lacked the resources, capabilities, or time to do so itself?

○ Very Irrelevant ○ Somewhat Irrelevant ○ Somewhat Relevant ○ Very Relevant ○ Most Relevant of All

Capture Notes Here:

Inspiration:

Consider an example from the environmental sector: Trusted data exchanges are crucial to monitoring the changes brought about by the climate emergency and to inspire innovative solutions to mitigate impact. Toward that end, The National Biodiversity Network Trust is a collaborative partnership created to exchange biodiversity information (National Biodiversity Network, 2022). With over 200 members including commercial organizations, supporters and sponsors, data is collected and shared within the network. The exchange works due to reciprocity: organizations know that by participating and providing data, they expand the data available to all, including themselves.

Additional Relevant Resources:

- Collaborative Data Ecosystems: Article discussing how organizations that collaborate on data, machine learning and data science could be the key to solving complex problems.
- Global Fishing Watch: GovLab case study on a data collaboration among SkyTruth, Oceana, and Google to map and measure fishing activity worldwide using data.

RECIPROCITY

THE WORLD BANK
IBRD • IDA

WHO WE ARE | WHAT WE DO | WHERE WE WORK | UNDERSTANDI

Who We Are / News / Events

EVENTS

Development Data Partnership Day 2023

May 10, 2023

Washington, D.C.

Add to calendar

R2: Rectifying Errors and Improving Data Quality

Definition: A benefit of data openness is that it can improve the quality of data. When data can be accessed through open channels, it is typically reviewed and re-used by people other than those who collected and originally processed it who might be able to provide a fresh perspective or a different domain expertise. This review process can help identify errors and inaccuracies in the data that might not have been otherwise caught. The reviews and feedback from others can also help to improve the quality of future data collections.

Data quality can have a tangible impact on organizations and how they operate. Gartner has found that poor quality data is responsible for an estimated USD 12.9 million annually in losses (Moore, 2018). Writing for the EQUATOR Network, an international initiative to promote transparency in health research, Professor Dorothy Bishop explains that providing access to data "exposes researchers to the risk of being found out to be sloppy or inaccurate." As with peer review, there is a strength in numbers to improve a dataset (Bishop, 2014).

Considerations: To help others review efficiently, organizations may wish to publish data collection methods, privacy statements, frequently asked questions, policies, and other information that provides context to those looking to re-use the data on offer. Providing a contact form so that people can ask questions or provide feedback may also prove helpful.

Ask yourself: By providing more access to its data or models, can your organization identify and rectify errors that might otherwise go unnoticed or unaddressed?

○ Very Irrelevant ○ Somewhat Irrelevant ○ Somewhat Relevant ○ Very Relevant ○ Most Relevant of All

Capture Notes Here:

Inspiration:

Open data's value in improving quality can be seen in the response to COVID-19. As the public and private sector both entered crisis over the disease, companies looked to data collaborations that could increase the quality of data available to public health authorities. Telecommunications data allowed leaders in the European Union to supplement and increase the granularity and quality of traditional health surveys that might predict where and how COVID-19 spreads. As The Atlantic magazine reports, COVID-19 has shown the need to "check multiple data sources and [...] 'triangulate' between them" while bringing an end to the "fragmented data status quo" that might lead to low-quality estimates (Ladyzhets, 2022).

Additional Relevant Resources:

- Why We Should Care About Bad Data: A blog post by The GovLab that looks at the impact bad quality data has on decision-making
- What makes quality open data: An online learning resource that covers data usability, standards and marques of qualityas a researcher.

RECTIFYING ERRORS & IMPROVING DATA QUALITY

- Data re-use can provide fresh perspectives and diverse domain expertise to review the data.
- This helps identify errors and inaccuracies that might not have been caught by the original data collectors or processors.

Ask Yourself: By providing more access to its data or models, can your organization identify and rectify errors that might otherwise go unnoticed or unaddressed?

R4. Reproducibility

Definition: When data is shared, it becomes possible for others to check and verify the results of the analysis. This ability to test results helps build trust and advance innovation agendas. For example, by making data available to others, organizations allow others to conduct identical or related work by providing an opportunity to confirm insights or revisit and improve upon previous analyses. In addition, by being open about data policies and processes, organizations can receive input and feedback into these and improve their data science capabilities as a whole.

Considerations: Amid the ongoing replication crisis, whereby shoddy findings get circulated because of a lack of external validation or journal due diligence, replication has proven especially persuasive for academia and a variety of other sectors (Piper, 2020). When trying to promote reproducibility, organizations may want to ensure their work focuses on real, quantifiable metrics, that they promote transparency in their methods as well as their data, and that they regularly verify their own processes internally to identify and counteract issues before they arise.

Ask yourself: By providing more access to its data, can your organization verify and validate its analysis, practices, and overall capabilities?

○ Very Irrelevant ○ Somewhat Irrelevant ○ Somewhat Relevant ○ Very Relevant ○ Most Relevant of All

Capture Notes Here:

Inspiration:

Consider the example of Microsoft Research's Open Data Repository. The repository, in Microsoft's own words, "makes available datasets that researchers at Microsoft have created and published in conjunction with their research." Run on its Azure platform, this repository provides a simple platform for Microsoft researchers and collaborators to share the basis of their work explicitly so that it can be validated by peers both within Microsoft's research community and those outside it. The platform gives internal and external researchers easy access to data, encouraging researchers to test one another's work and findings.

Additional Relevant Resources:

- Identifying and addressing data asymmetries so as to enable (better) science: Article which seeks to start filling the analytical gap regarding global data asymmetries.
- What the drive for open science data can learn from the evolving history of open government data: Article that looks at the history of the open government data movement and how the open science movement can identify ways to move forward and learn from it.
- A manifesto for reproducible science: Paper, published in Nature, that argues for the adoption of measures to optimize key elements of the scientific process.
- Should you share your data?: Blog post discussing the opportunities and challenges of sharing data as a researcher.

REPRODUCIBILITY

- When data is shared, it becomes possible for others to check and verify the results of the analysis. This ability to test results helps build trust and advance innovation agendas.
- When trying to promote reproducibility, organizations may want to ensure that they promote transparency in their methods as well as their data.

Ask Yourself: By providing more access to its data, can your organization verify and validate its analysis, practices, and overall capabilities?

REPRODUCIBILITY

Reproducibility with Microsoft Research Open Data

Vani Mandava
February 2020
Invited talk at The AAAI 2020 Workshop on Reproducible AI - RAI2020, NYC
Related File

REPUTATION

- One way in which providing access to data can help build brand equity is by enhancing an organization's image and reputation. When organizations open their data, it demonstrates their commitment to transparency and social responsibility.
- Organizations can use data collaboration for:
    - Strategic messaging that can help generate positive narratives AND leads for commercial business
    - Work with journalists and media outlets to feature content from the data collaboration.

Ask Yourself: By providing more access to its data, can your organization improve its image and reputation to media, customers and/or investors who value socially conscious corporate actors?

ROI Case #2: Brand Equity: The ways that data reuse can create a return on investment by promoting an organization's image to internal and external stakeholders.

R5. Reputation

Definition: One way in which providing access to data can help build brand equity is by enhancing an organization's image and reputation. When organizations open their data, it demonstrates their commitment to transparency and social responsibility. This, in turn, can improve the public's perception of the organization and build trust, which can attract new users and investors who value socially responsible organizations. The Mastercard Center for Inclusive Growth, for example, noted that it was important that they "show up as more than just a payments company" and demonstrate to potential partners that their company data "can be used for good."

Considerations: To turn an open data initiative into a brand-building exercise, organizations may want to create a compelling narrative that can be packaged into a variety of media formats and shared. Ideas could include building bespoke open data portal or creating branded content that can be shared across marketing channels.

Ask yourself: By providing more access to its data, can your organization improve its image and reputation to media, customers and/or investors who value socially conscious corporate actors?

○ Very Irrelevant ○ Somewhat Irrelevant ○ Somewhat Relevant ○ Very Relevant ○ Most Relevant of All

Capture Notes Here:

Inspiration:

The reputational benefits of data openness are especially important for organizations in the business sector, where data sharing can establish trust and help attract more customers at a lower cost and increase customer retention. The Open Data Institute quotes Sarah Hitchcock, partner at the location-analytics company GeoLytix, as stating that "open data products allow us to build a reputation in our sector and expand our customer base, but also contribute back to the ecosystem and help improve the quality of open data" (Shadbolt, 2015). Similarly, as Smita Jain, of the Mastercard Center for Inclusive Growth noted in an interview about her company's work with data sharing: "We want to show that we're committed to communities, committed to their geographies we operate in and it's not just a purely business transaction. We are there to be a supportive player in the community, too."

Additional Relevant Resources:

- Data Responsibility: Using corporate data to save lives: TEDxMidAtlantic by The GovLab co-founder Stefaan Verhulst
- Uber Movement: Explore how Uber is playing a role in helping cities grow in a way that works for everyone.
- Mastercard's Data Responsibility Principles: Read more about their open data initiative, Inclusive Growth Score, here.

REPUTATION


Print subscriptions
Sign in
Search
Support the Guardian
Fund independent journalism with €5 per month
Support us
News
Opinion
Sport
Culture
Lifestyle
More
World
Europe
US
Americas
Asia
Australia
Middle East
Africa
Inequality
Global development
Coronavirus
This article is more than 3 years old
Phone location data could be used to help UK coronavirus effort
Exclusive: BT in talks with government about mapping phone users' movements

R6. Responsibility and Philanthropy

Definition: Data openness can also help bolster an organization's reputation as a responsible corporate actor. When data is shared, it can demonstrate an organization's commitment to generating social value. This can help attract new users, customers, and investors who value socially conscious corporate actors. Importantly, philanthropy does not need to be something that comes at the expense of other activities. Civic technologist Matt Stempeck, writing in the Harvard Business Review, notes that data philanthropy "allows companies to give back in a way that produces meaningful impact, and reflects the businesses' core competencies while preserving or expanding value for shareholders" (Stempeck, 2014).

Considerations: Organizations might want to think about what data releases can actually accomplish in what context. A well-defined goal of how data can be used for good—and by which audiences—can minimize the risk of publishing data for just publishing sake. Superficial data exchanges can be seen as "open washing"—superficial efforts to publish data without full integration with transparency commitments (Verhulst and Zahuranec, 2022).

Ask yourself: By providing more open access to its data, can your organization fulfill its social responsibilities, improving where they operate or achieving some social good?

○ Very Irrelevant ○ Somewhat Irrelevant ○ Somewhat Relevant ○ Very Relevant ○ Most Relevant of All

Capture Notes Here:

Inspiration:

Many companies report pursuing data openness out of a sense of obligation or social responsibility. Matt Webb, Head of Enterprise Data Management at UK Power Networks cites his organization's reasons for publishing portions of its data:

> *"The main driver is the recognition of the key role that data plays in the smart grid of the future and supporting the whole system interoperability, contributing towards Net Zero targets. We have a role to play as an organization; certainly from the perspective of corporate responsibility, but equally, if not more so in terms of our social responsibility."*

This attitude matches with opinions in the United Kingdom, where 75% of adults are concerned about climate change (UK Office of National Statistics, 2021).

Additional Relevant Resources:

- The Open Data for Social impact Framework: A tool, developed by Microsoft, that organizational leaders can use to further understand how best to put data to work to solve important societal challenges.
- Corporate Social Responsibility for the Data Age: Article that defines data responsibility, its three pillars and the cultural shift that is needed.
- UK Power Networks' Open Data Portal: Visit one of the UK's biggest sets of data about the electricity network.
- Sharing Data is a Form of Corporate Philanthropy: Article in the Harvard Business Review that argues that providing access to data should be considered a corporate social responsibility initiative.

RESPONSIBILITY & PHILANTHROPY

- When data is shared, it can demonstrate an organization's commitment to generating social value. This can help attract new users, customers, and investors who value socially conscious corporate actors.
- Organizations might want to think about what data releases can actually accomplish in what context. A well-defined goal of how data can be used for good can minimize the risk of publishing data for just publishing sake.

Ask Yourself: By providing more open access to its data, can your organization fulfill its social responsibilities, improving where they operate or achieving some social good?

RESPONSIBILITY & PHILANTHROPY

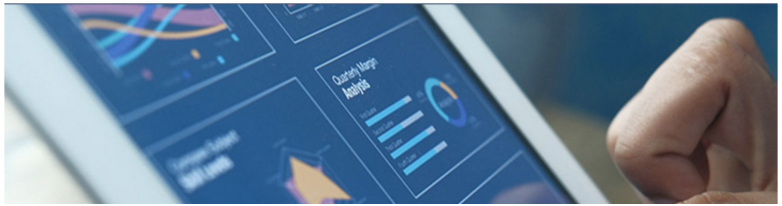
Center for Inclusive Growth

About Insights Impact Events Newsroom

DATA FOR GOOD

Data Philanthropy Offers New Avenues for Solving Old Problems

Leveraging private-sector data-driven insights for the public's benefit is rising.



Stanford SOCIAL INNOVATION Review

Informing and inspiring leaders of social change

SIGN I

SOCIAL ISSUES SECTORS SOLUTIONS MAGAZINE GLOBAL EDITIC

Business

Corporate Social Responsibility for a Data Age

Proprietary data can help improve and save lives, but fully harnessing its potential will require a cultural transformation in the way companies, governments, and other organizations treat and act on data.

CITE SHARE COMMENT PRINT ORDER REPRINTS

By Stefaan G. Verhulst | Feb. 15, 2017

/s & Quotes

BARRON'S

Topics Stock Picks Lists & Rankings Magazine Data Advisor Penta

The Case for Including Data Stewardship in ESG

COMMENTARY By Stefaan G. Verhulst March 31, 2023 3:43 pm ET Order Reprints

RETAINING & RECRUITING TALENT

- Data collaboration can help companies:
  - Identify and attract individuals with the right skills and values, by creating a corporate culture that is attractive to top talent.
  - Increase the morale across the whole organization

Ask Yourself: By providing more open access to its data, can your organization keep or attract diverse talent with projects that are compelling and socially relevant?

R7. Retaining and Recruiting Talent

Definition: One of the most important ways in which data sharing can help companies is by allowing them to identify and attract individuals with the right skills and values. Data sharing also helps create a corporate culture that is attractive to top talent. When data is shared, it can help foster an environment of transparency and trust. This environment can be appealing to individuals who are looking for organizations that are responsible and accountable. As noted by Smita Jain, Manager, Impact Data Science of the Mastercard Center for Inclusive Growth, "Data for good really resonates with our employee base. It's a very key point for those folks interested in working for Mastercard, and it's good for the business because our talent is excited about working here and getting involved."

Considerations: In their paper reviewing corporate social responsibility and organizational attractiveness, authors Story, Castanheira and Hartig (2016) find that "CSR can play an effective role in attracting potential employees." The researchers note that "practices that improve the quality of the natural environment and the well-being of the society" are most convincing when combined with practices that "protect and develop" employees. Organizations might want to center these ideas if they pursue their own open data efforts.

Ask yourself: By providing more open access to its data, can your organization keep or attract diverse talent with projects that are compelling and socially relevant?

○ Very Irrelevant ○ Somewhat Irrelevant ○ Somewhat Relevant ○ Very Relevant ○ Most Relevant of All

Capture Notes Here:

Inspiration:

Google.org demonstrates these ideas can look in practice. The charitable arm of the Google company, Google.org harnesses the funding, tools, staff, data, and other resources of Google to support nonprofits, social enterprises, and civic entities around the world. While the fellowship has clear reputational benefits, it is also critical in keeping staff interested and engaged in data work, providing opportunities for employees to "make a difference" through data collaborations. In testimonies available on Google.org's website, staff highlight how they were "motivated by the opportunity to combine technical skills" with their passion for other sectors and domains. Individuals also highlight their appreciation for working on topics that made "immediate and tangible difference in the community." It encouraged them to remain with Google while allowing them to tackle issues they cared about.

Additional Relevant Resources:

- Experimental Statistics: Article describing how the Office for National Statistics created an indicator of job vacancies with Adzuna job board data.
- Google.org: Explore some of Google's open data sets here.

RETAINING & RECRUITING TALENT

There's a "morale crisis" at Meta

Emily Peck, author of Axios Markets

Juan M. Lavista Ferres • 1st
Chief Scientist and Lab director at Microsoft AI for Good Res...
1d •

Meet the team behind our Microsoft AI for Good Lab, representing a global reach, and in Redmond last week to talk more about how we use technology to make the world a better place.... ...see more

Gretchen Deo and 294 others 10 comments • 9 reposts

rice

REGULATORY COMPLIANCE

- As data has become a larger part of daily life and individuals have grown concerned about the concentration of data ownership, there has been growing demands for organizations to be more transparent about their operations.
- Organizations who operate in multiple jurisdictions might want to look at the regulations and laws related to data collaboration where they operate.

Ask Yourself: By providing more open access to its data, can your organization better comply with regulations or improve its social license by demonstrating responsible data use and management?

ROI Case #3: License to Operate: The ways data reuse can create a return on investment by strengthening the mission or permission for an organization to operate.

R8. Regulatory Compliance

Definition: As data has become a larger part of daily life and individuals have grown concerned about the concentration of data ownership, there has been growing demands for organizations to be more transparent about their operations. This in turn has been accompanied by increased attention to the need for organizations to open their datasets.

Many laws today require organizations to be more transparent and open. Government agencies in a number of jurisdictions are required to launch open data portals or publish data on their operations (City of New York, NYC Open Data, n.d; Kim, 2019). Likewise, companies in the United States and other countries are increasingly expected to report on their financials, the management of their workforce, and other information (Probst, 2021). These efforts intend to boost transparency and the legitimacy of certain data efforts.

Considerations: Organizations who operate in multiple jurisdictions might want to look at the regulations and laws related to data collaboration where they operate. Rules, procedures, and expectations may vary from country to country and even within countries (Klosowski, 2021).

Ask yourself: By providing more open access to its data, can your organization better comply with regulations or improve its social license by demonstrating responsible data use and management?

○ Very Irrelevant ○ Somewhat Irrelevant ○ Somewhat Relevant ○ Very Relevant ○ Most Relevant of All

Capture Notes Here:

Inspiration:

For example, in April 2017, new legislation in the United Kingdom came into force that requires all organizations with 250 employees or more to annually report on their gender pay gap. This includes public authorities, private and voluntary employers. The gender pay gap is the difference in average earnings of men and women in the labor market. Created and published by the Government Equalities Office (GEO), gender pay gap reporting guidance outlines the data that organizations must collect, the methodology they must use to perform calculations and what kind of supporting narrative–an explanation of internal and external factors resulting in the gap–is expected. Organizations may also choose to publish strategies that describe how they intend to take action to narrow, and ultimately, close the gap.

Additional Relevant Resources:

- Gender Pay Gap Service: Use the online service to explore data submitted about the gender pay gap.
- State of Open Data Policy repository: Repository by the Open Data Policy Lab to assess policy developments on open data and data reuse.

REGULATORY COMPLIANCE

Council of the EU Press release 24 March 2023 19:40

Data act: member states agree common position on fair access to and use of data

With a view to making the EU a leader in our data-driven society, member states' representatives (Coreper) reached a **common position** ("negotiating mandate") allowing the Council to enter negotiations with the European Parliament on the proposed legislation regarding harmonised rules on **fair access to and use of data** (data act).

> Today's agreement will facilitate the digital transformation of our societies and economies. The data act will unlock the economic and societal potential of data and technologies in line with EU rules and values. It will contribute to creating a single market to allow data to flow freely within the EU and across sectors for the benefit of businesses, researchers, public administrations, and society at large.
>
> — Erik Slottner, Swedish minister for public administration

R9. Revenue Generation

Definition: Private-sector organizations are incentivized to generate new revenue and cut costs. Providing access to data can help organizations accomplish both of these goals by providing opportunities to develop new revenue streams, and by reducing the costs through improved operational efficiencies. The Open Data Institute has published a guide on potential business models for open data—suggesting organizations look at freemium models, subsidize openness through the value it can provide other sectors, or rely on networks to support datasets (Open Data Institute, 2013).

Considerations: Data openness can present opportunities but organizations often struggle to scale up efforts due to inadequate funding among other issues (Deiglmeier and Greco, 2018). For efforts to succeed, institutions need to be willing to supply the start-up costs associated with them. They also need to be able to build broad-based support across their organization instead of limiting it to one department or team.

Ask yourself: By providing more open access to its data, can your organization find new opportunities to generate revenue or cut costs?

○ Very Irrelevant ○ Somewhat Irrelevant ○ Somewhat Relevant ○ Very Relevant ○ Most Relevant of All

Capture Notes Here:

Inspiration:

Cuebiq is a location-based measurement company that provides its customers with anonymized information on the locations of smartphone app users. One major branch of Cuebiq was its Data for Good initiative, which sought to provide "positive social impact through ethical and responsible use of location-based data" (Cuebiq Marketing Team, 2019). This program (now under the umbrella of Spectus.ai) had various motivations but necessarily thinks about financial sustainability. In the words of Cuebiq's Brennan Lake:

> "For shared value initiatives such as Cuebiq Data for Good to succeed in creating long-term social impact, it is important to integrate a financial sustainability component into the program's overall strategy. We do not view Data for Good as a primary driver of profit within Cuebiq, but we do pursue a blended distribution model which includes both data philanthropy and double bottom line partnerships."

Additional Relevant Resources:

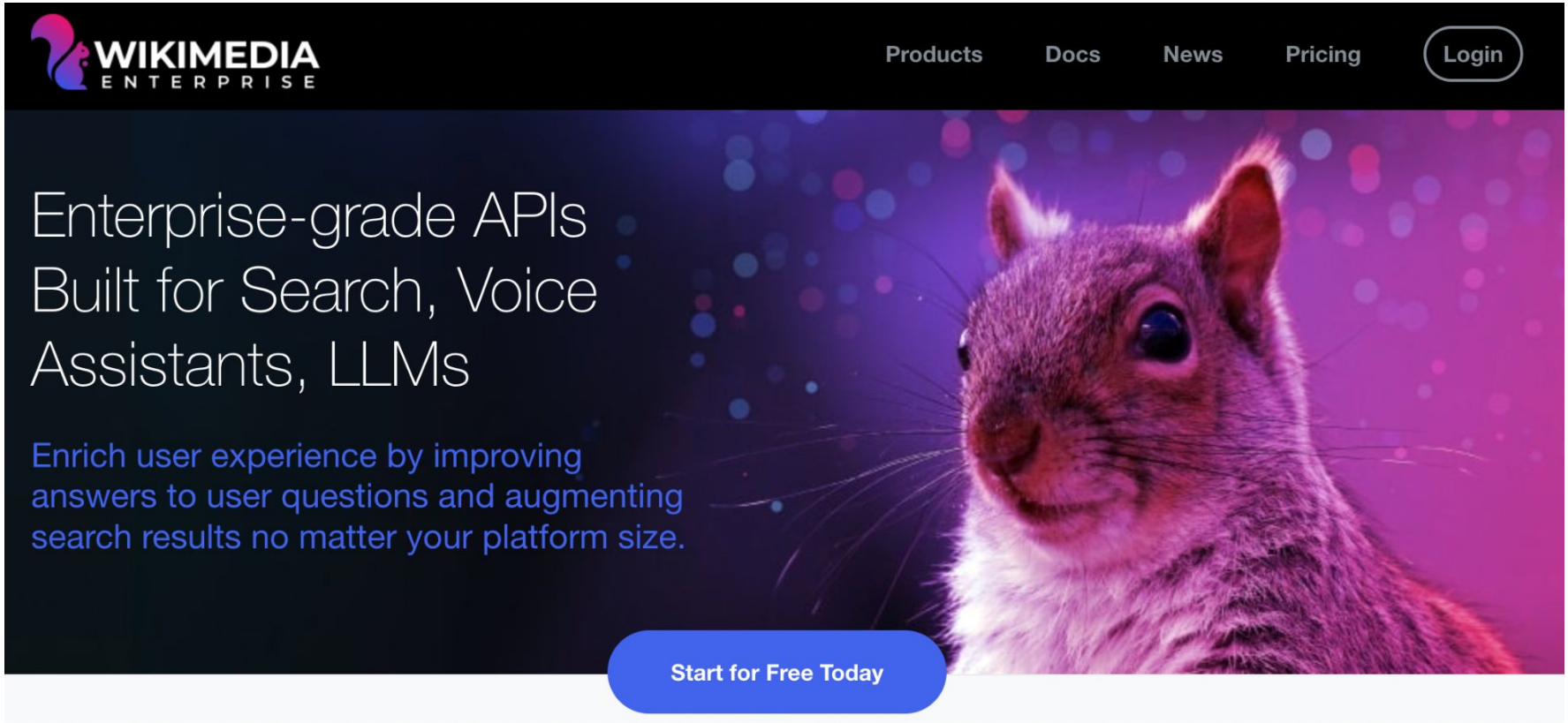
- Evaluating the Economic and Social Returns of Open Data Standards: A white paper from Frontier Economics evaluating the benefits of "openness" in different contexts.
- Case study: The value of sharing data for improving market reach: A case study from the Open Data Institute on how a UK leisure operator shared data to better reach new customers.

REVENUE GENERATION

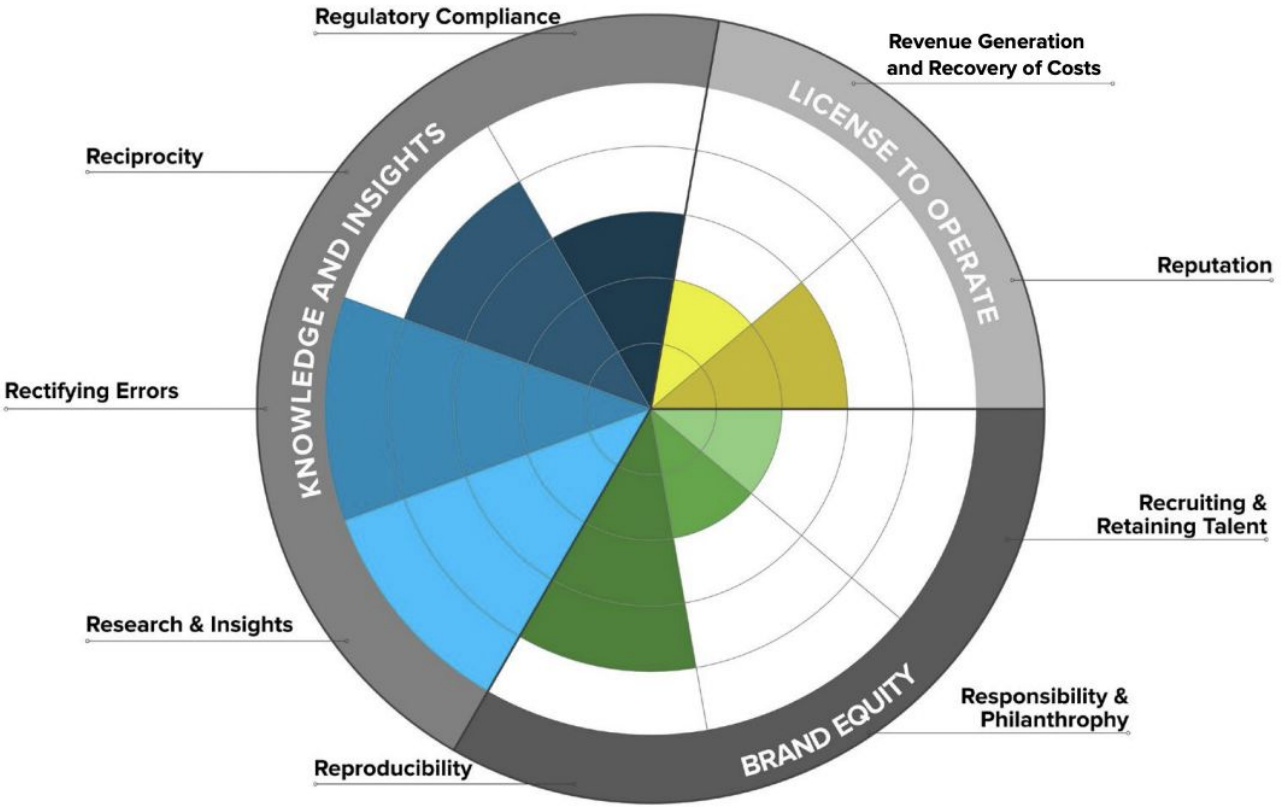
- Providing access to data can help organizations accomplish both of these goals by providing opportunities to develop:
  - New revenue streams or monetization opportunities,
  - Identify new commercial leads and by
  - Reducing the costs through improved operational efficiencies.
- For efforts to succeed, institutions need to be willing to supply the start-up costs associated with them.

Ask Yourself: By providing more open access to its data, can your organization find new opportunities to generate revenue or cut costs?

REVENUE GENERATION



THE BUSINESS CASE FOR DATA RE-USE



MEASURING THE COST & BENEFIT OF DATA (RE-)USE

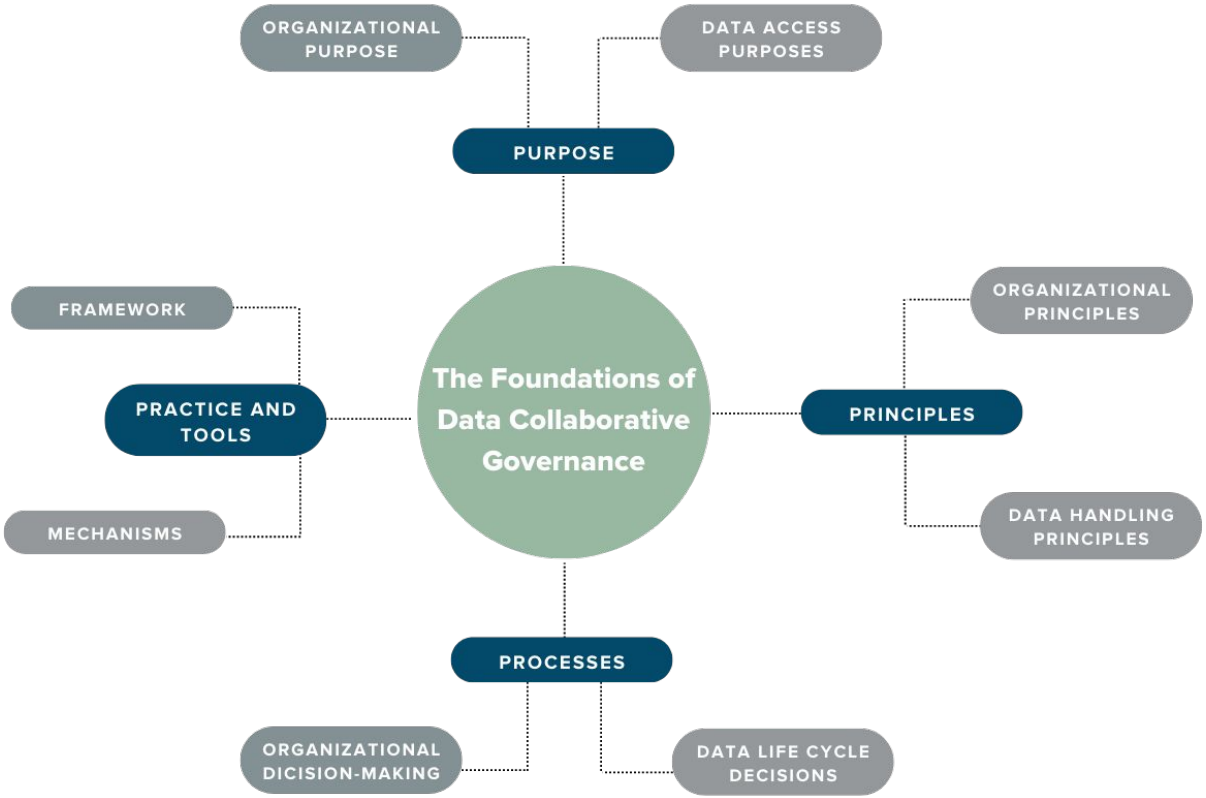
Stakeholders	Direct Benefits		Direct Costs	
	Recurrent	One-Off	Recurrent	One-Off
Data Holders (private and public sector)				

MEASURING THE COST & BENEFIT OF DATA (RE-)USE

Stakeholders	Direct Benefits		Direct Costs	
	Recurrent	One-Off	Recurrent	One-Off
Data Holders (private and public sector)	**VALUE** Improved value of own data sets and access to research, analysis methods and models previously not available, adhering to the principle of embracing expertise and efficiency **RISKS** Removal of legal and reputational risks (access model by public entities agreed and validated)		**PROCESSING** Costs of normalization and making datasets available for reuse (including in absence of compensation depending on the use-case). **PARTNERSHIP** Cost associated with formalising partnership including contractual arrangements	**INVENTORY** Costs of cataloguing and identifying data that can be valuable for public interest purposes **SETUP GOVERNANCE** new Data Steward position, developing internal data governance approach to comply with relevant laws and regulations

Premise 5:
Establishing legitimate governance enhances responsibility.

THE 4 Ps OF GOVERNING A DATA COLLABORATIVE



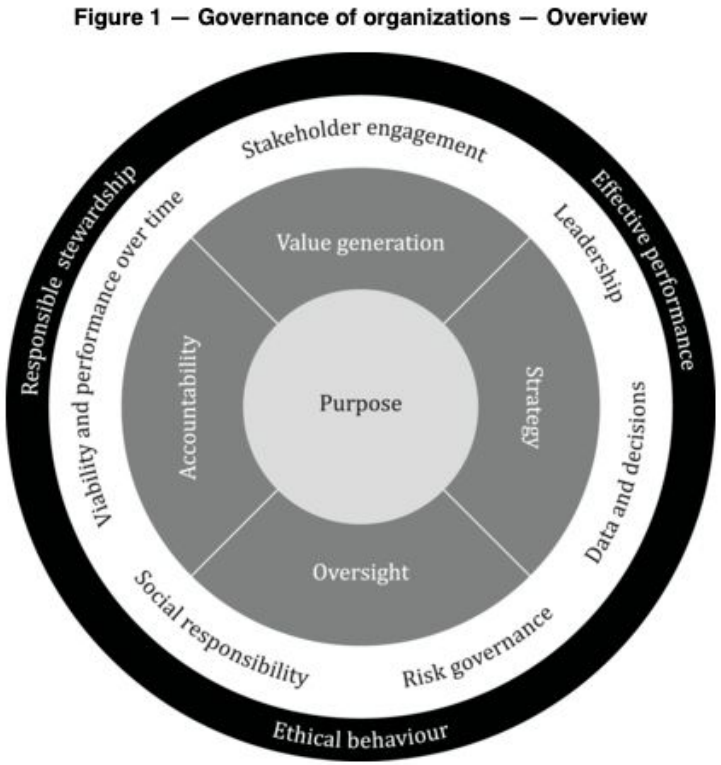
PURPOSES

Organizational Purposes:

- What is the purpose/mission of the data collaborative (beyond providing access to data)?
- Who does the data collaborative seek to inform/serve?
- What impact or action will be enabled by the data collaborative that could not be pursued by other means?

Data Purposes:

- What's the purpose/topic for which access to data is needed?
- Who are the various stakeholders that have or need data?
- What topics and/or questions need to be prioritized?



PRINCIPLES

Organizational Principles:

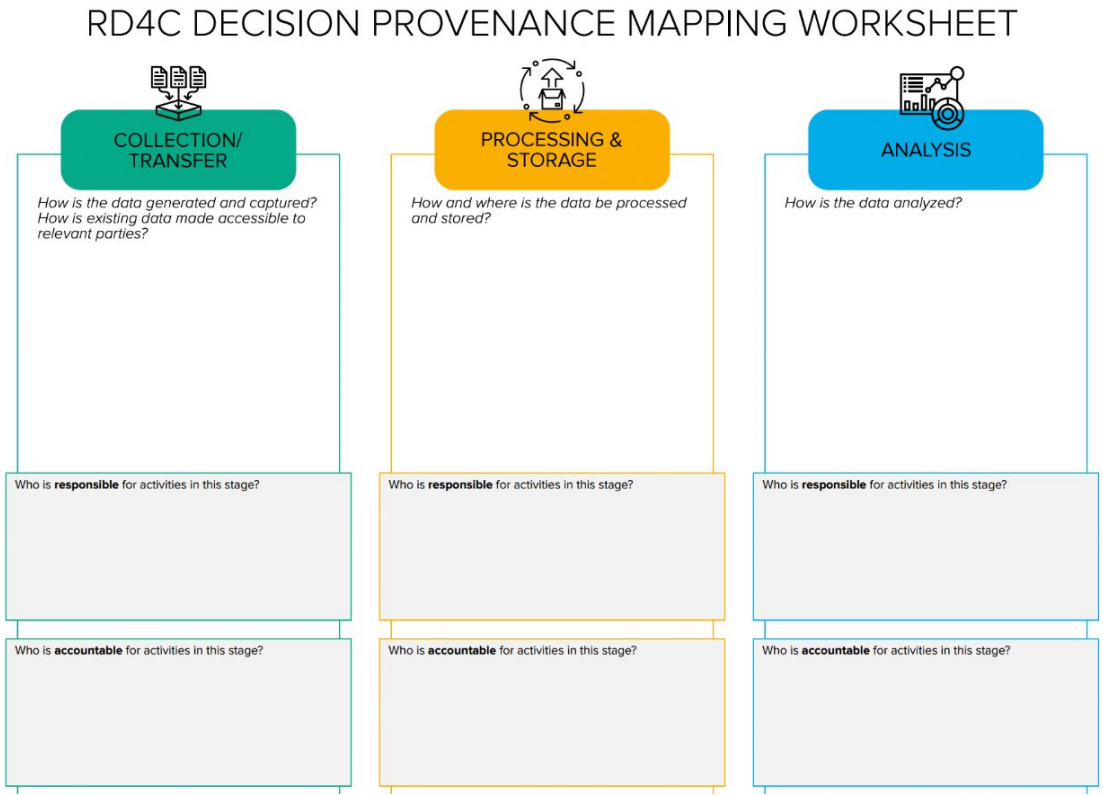
- What are the principles that will inform decisions at the organizational level (to meet the purpose)?
- How can we establish trust and legitimacy?

Data Principles:

- What are the principles that will inform how data is handled?
- What will help this project be effective?

	Design Principle	Conditions for Success	Implementation Recommendation(s)
LEGITIMATE			
1	**TRANSPARENT:** Openness and transparency, including procedural and financial transparency	Transparency should be considered an institutional priority when creating PLACE Trust. It will therefore be critical to adopt and use best practices for public access to key financial and long-term planning documents, as well as the development of records that document both the input and the analysis that leads to the decisions made.	Governance of PLACE Trust should include procedures and processes that foster collective agreement in a timely and transparent fashion including involvement from PLACE Trust Staff, Members, and external stakeholders. Decision-making processes should also consider active community and external input, which can be facilitated by an Engagement Director.
2	**PARTICIPATORY**: Strong participation by all affected stakeholders on recommending criteria and policies that support	Participatory decision-making is a core element of legitimate governance. Perhaps equally important, the perceived legitimacy of participatory processes can have significant ramifications on	PLACE Trust should develop agile, adequate and inclusive decision making processes including, for instance, a set of Committees composed of members, partners and affected constituencies. PLACE Trust should also establish processes

PROCESSES



Data Lifecycle Decisions:

- What are the key decision processes and makers across the data life cycle?
- Who is responsible for data provenance?
- How will we maintain accountability?

DATA ASSEMBLIES

- A collaboration among and between citizens, civil rights organizations, key data holders and policymakers.
- Example: 2020 COVID-19 pandemic data response in New York City
  - Identified varying concerns, expectations, and opportunities surrounding data re-use;
  - Co-designed innovative solutions and design principles
  - Created the Responsible Data Re-Use Framework

THE DATA ASSEMBLY

PRACTICES

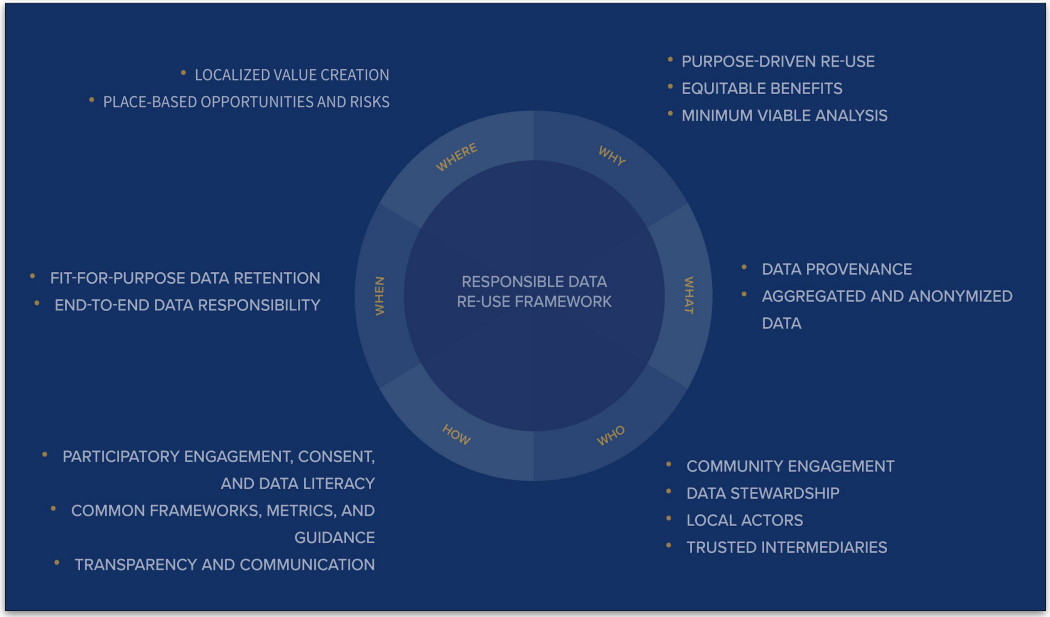
C4DC CONTRACTS FOR DATA COLLABORATION

Frameworks:

- Responsible or ethical use frameworks

Mechanisms:

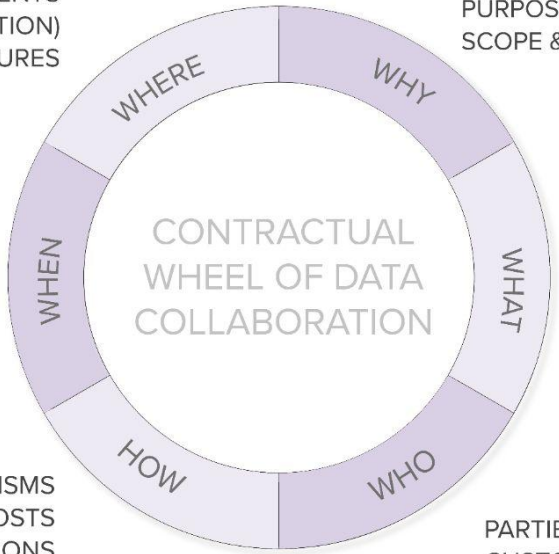
- Contracts or Data Sharing Agreements
- Codes of Conduct
- Terms and Conditions
- Technologies



THE CONTRACTUAL WHEEL OF DATA COLLABORATION

PUBLICATION AND (PRE)DISSEMINATION REQUIREMENTS
JURISDICTION IMPLICATIONS (INCLUDING INDEMNIFICATION)
ENFORCEMENT PROCEDURES

PURPOSE OF THE DATA COLLABORATIVE
SCOPE & LIMITATIONS



DURATION (PERIOD OF AGREEMENT)
FREQUENCY OF UPDATES
DATA RETENTION
TERMINATION & MODIFICATION

DATA ASSETS & (RE)SOURCES
FORMATS STANDARDS AND TECHNICAL
REQUIREMENTS
PROVENANCE

OPERATIONAL MODELS & SHARING MECHANISMS
RESOURCES & COSTS
PERMISSIONS
LEGAL & PROFESSIONAL REQUIREMENTS
GOVERNANCE & AUDIT
DISPUTE RESOLUTION & RISK MITIGATION STRATEGIES

PARTIES (PROVIDERS & USERS)
CUSTODIAL DUTIES
ACCESS CRITERIA
RIGHTS & RESPONSIBILITIES

Premise 6:
A technical infrastructure is needed to strengthen a data collaborative.

ACCESS CONTROLS

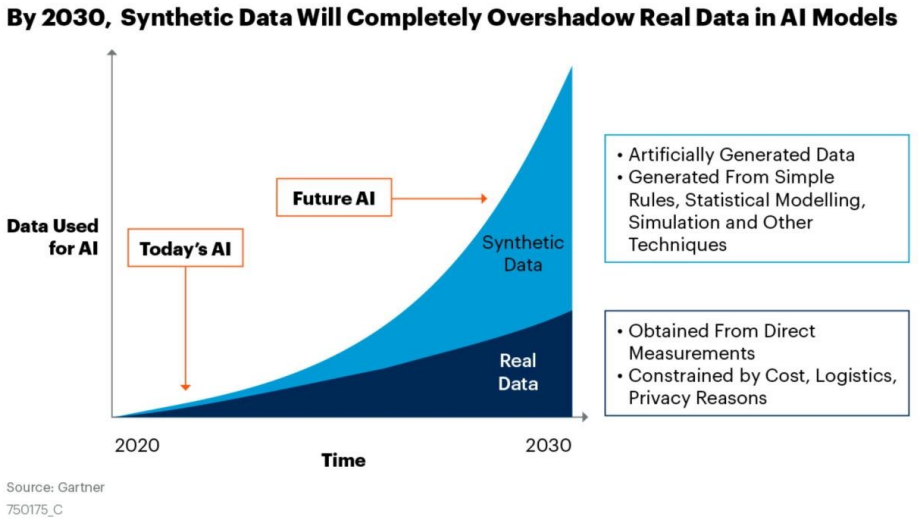
Process by which users are granted access and certain privileges to systems, resources or data assets.

Technical Component	Purpose
Homomorphic Encryption	Allows computation to be performed on encrypted data with the result of computation remaining encrypted; can allow collaborative engagement in computations.
Multi-Factor Authentication (MFA)	Requires authorized parties to provide at least two identifying credentials before accessing data in order to protect data from unauthorized access.
Multilevel Secure Database Management System (DBMS)	Seeks to provide users with different access levels to a shared database in order to protect data with varying levels of sensitivity.

DATA PREPARATION & SECURITY

Processing data to ensure anonymization, confidentiality, and/or protection of individual privacy.

- De-identification
- Hashing
- Data Protection Audit
- Differential Privacy
- K-Anonymization
- Synthetic Data
- Pre-Computed Indicators



Source: Gartner, "Maverick Research: Forget About Your Real Data – Synthetic Data Is the Future of AI," Leinar Ramos, Jitendra Subramanyam, 24 June 2021.

DATA STORAGE & TRANSFER

Holding data within a single database; copying or moving data between multiple databases.

Technical Component	Purpose
Data Enclave / Safe Sandboxes	A data storage and experimentation technique that provides access to data on a remote server while restricting users from downloading the data or otherwise moving it outside of the secure environment.
Distributed Ledger / Blockchain	Introduces immutable properties to databases, proving data has not been tampered with.
Edge Processing	Allows data to be processed closed to where it is produced (e.g. on a sensor) instead of transmitting it long distances to data centers or the cloud for processing.
Secure Cloud	Stores data securely in the cloud by encoding user data with a specific encryption key and requiring password authentication to access the key.

Premise 7:
A data steward plays a key role towards making data collaboratives more systematic.

MATCHING DEMAND & SUPPLY

DEMAND SIDE

Demand side actors seek data to **understand the situation, identify cause and effect, make predictions,** and **solve public problems.**

SUPPLY SIDE

Supply side actors often have access to vast stores of **siloed data.** These sources can, when used responsibly, answer critical questions.

SCOPING THE DEMAND/SUPPLY

DATA AUDIT AND ASSESSMENT OF VALUE AND RISK
Monitoring and assessing the value, potential, and risk of all data held within an organization.

What's the **question** we seek to answer?

DEMAND SIDE

What's the minimum **viable data** needed to answer the question? And what is available>

Who needs to know what?

SUPPLY SIDE

What **data** do we have that can answer the question?

What is the **value proposition** for us sharing our data?

What's the **quality** and **interoperability** of our data?

ASSESSING THE RISKS

DATA AUDIT AND ASSESSMENT OF VALUE AND RISK
Monitoring and assessing the value, potential, and risk of all data held within an organization.

DEMAND SIDE

SUPPLY SIDE

What's the **quality** of the data (bias)?

What's the **reputation** of the data holder?

Was the data acquired through **appropriate and ethical** means?

What may happen following a **breach**? Who may be harmed?

Are there **confidentiality** concerns?

Is it **legal** to share our data? Do we need an open data license?

How will **insights** be used?

Is our data **sensitive**? How do we anonymize our data?

SCOPING OPERATIONAL MODELS

DATA AUDIT AND ASSESSMENT OF VALUE AND RISK
Monitoring and assessing the value, potential, and risk of all data held within an organization.

What are our **requirements** for data access?

Data **methodology**?

API? Data pool/lake? **Secure Sandbox**? Bi-lateral/multilateral?

What is **fit for purpose?**

DEMAND SIDE

How can we provide **access** to our data?

Protocols for **secure transfer** of data insights?

Is our data **reusable** and how to share?

How Open? Off-site? On-line? One-time? On-going? Security?

SUPPLY SIDE

SCOPING OPERATIONAL MODELS

DATA AUDIT AND ASSESSMENT OF VALUE AND RISK
Monitoring and assessing the value, potential, and risk of all data held within an organization.

What are the **conditions** for accessing the data?

Is there any **metadata**?

DEMAND SIDE

How **fresh** is the data?

Are there **data standards**?

SUPPLY SIDE

How to **format** our data?

How to **document** what the data is all about?

How do we **define** terms used?

How to update **different versions** over time?

DEVELOPING PARTNERSHIPS

PARTNERSHIP AND COMMUNITY ENGAGEMENT
Proactively and responsively reaching out to and vetting potential partners.

How to **engage communities**?

DEMAND SIDE

SUPPLY SIDE

How to develop a **data sharing agreement**? NDA? Licensing?

Other **beneficiaries** that need to be considered?

Which **organizations** might have valuable data?

Who else should be part of the data collaborative?

Who will **touch the data**? Vetting of analytics team?

Are the partners **trustworthy**?

INTERNAL COORDINATION

INTERNAL COORDINATION AND STAFF ENGAGEMENT
Securing internal coordination and sign-off from various company actors.

Who do I need to **sign off**?

DEMAND SIDE

SUPPLY SIDE

Who do I need to **sign off**?

Do we have any **internal expertise**?

Who **else is doing what in the data space**?

Who can **lead** our partnership efforts?

Any **internal expertise** needed to make data re-useable?

ASSESSING FINANCIAL SUSTAINABILITY

NURTURE DATA COLLABORATIVES TO SUSTAINABILITY
Gathering the needed resources and support so as to ensure broad and long-term impact.

How to raise **funds** to secure access and cover the costs?

DEMAND SIDE

SUPPLY SIDE

What's the **business case**?

How can we **sustain access** for the long term?

Do we have **adequate resources** to support this project?

How much will this **cost**?

What are the long term **benefits** of this project?

MAKING IMPACT

DISSEMINATION AND COMMUNICATION OF FINDINGS
Raising awareness, disseminating findings and communicating outcomes from data collaboratives

Can these **findings** inform other projects or future collaboratives?

DEMAND SIDE

Who needs to be informed and **act** upon the insights?

How can we effectively **communicate** our findings to maximize impact?

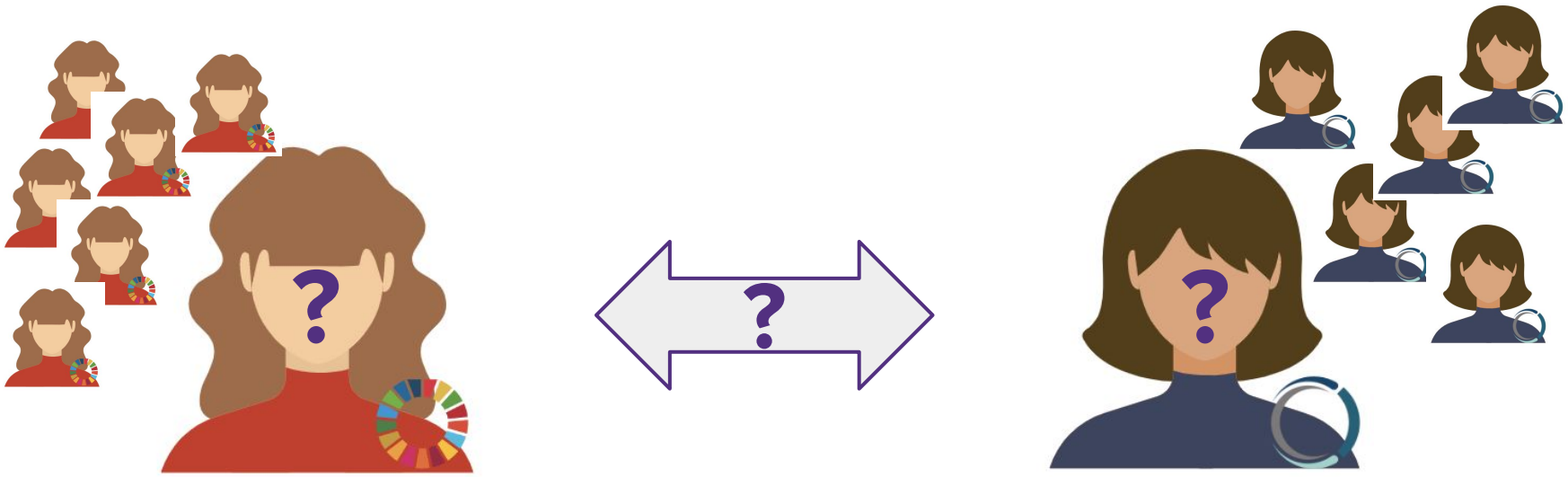
SUPPLY SIDE

Can these **outcomes** help or harm our business practices? (Vetting of insights)

Should we **sunset** this initiative?

How can we effectively **communicate** our findings to maximize impact?

TODAY'S SITUATION



Most data collaborative efforts fall flat because of the lack of a designated data stewardship function.

WANTED: DATA STEWARDSHIP

DATA STEWARDSHIP: the functions and roles that enable the re-use of data for public benefit in a systematic, sustainable and responsible way through data collaboration.

RE-IMAGINING DATA STEWARDSHIP

Data Stewardship within a Scientific and Library Context

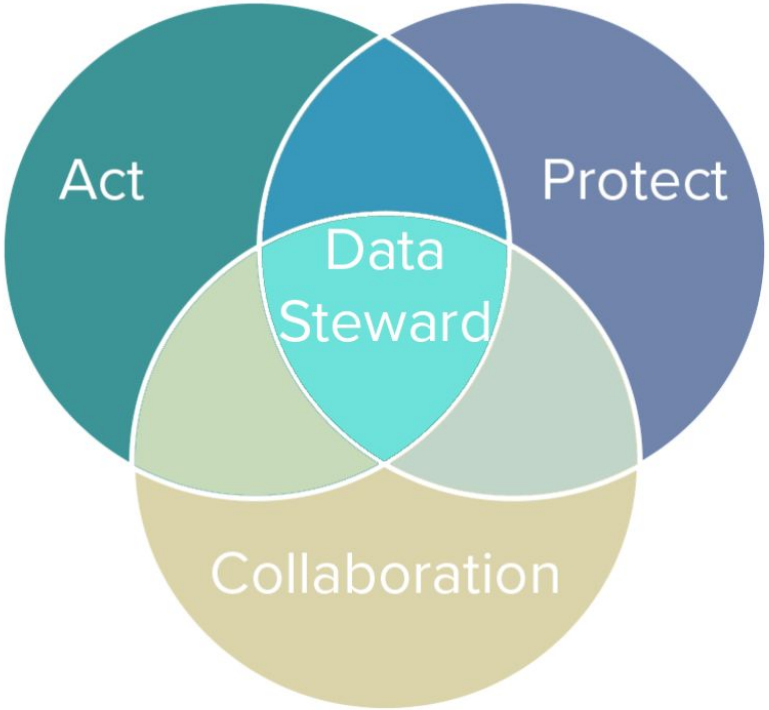
Data Stewardship within a Corporate Data Governance Context

Integrity of the Data

Security of the Data

Re-Use of the Data?

RE-IMAGINING DATA STEWARDSHIP



Collaborate:
Working with others to unlock the inherent value of data when it serves the public good.

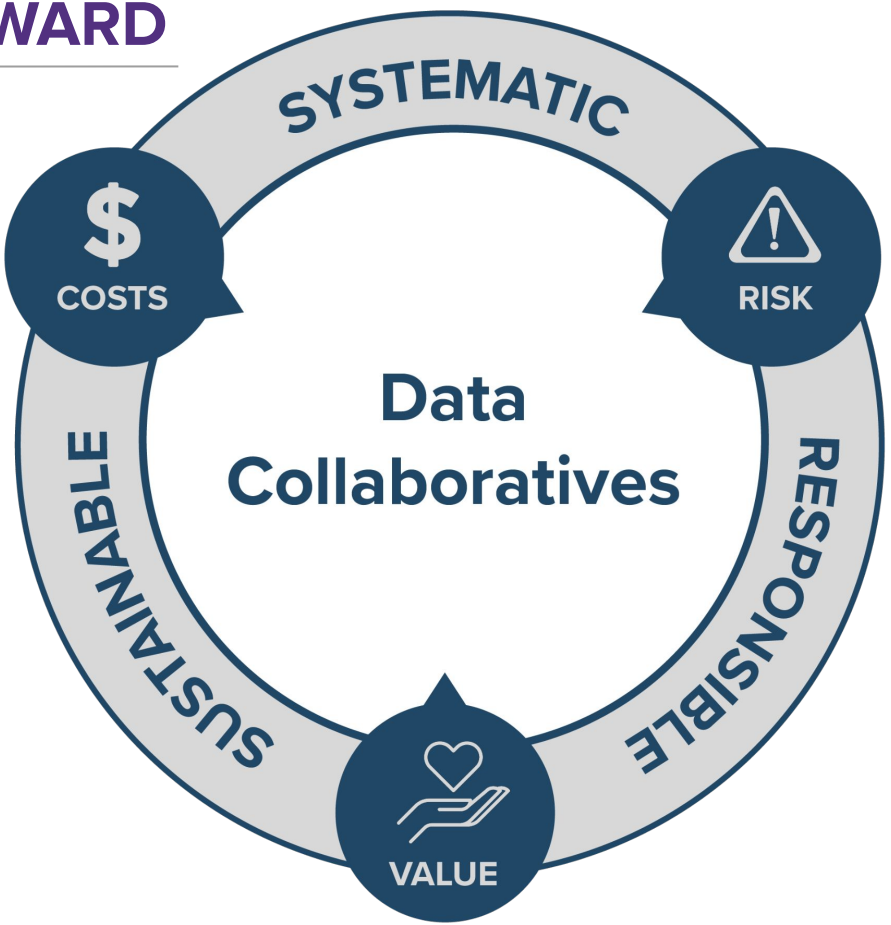
Protect:
Managing data ethically and preventing harm to all whose data may be shared.

Act:
Proactively identifying partners who can unlock value and insights.

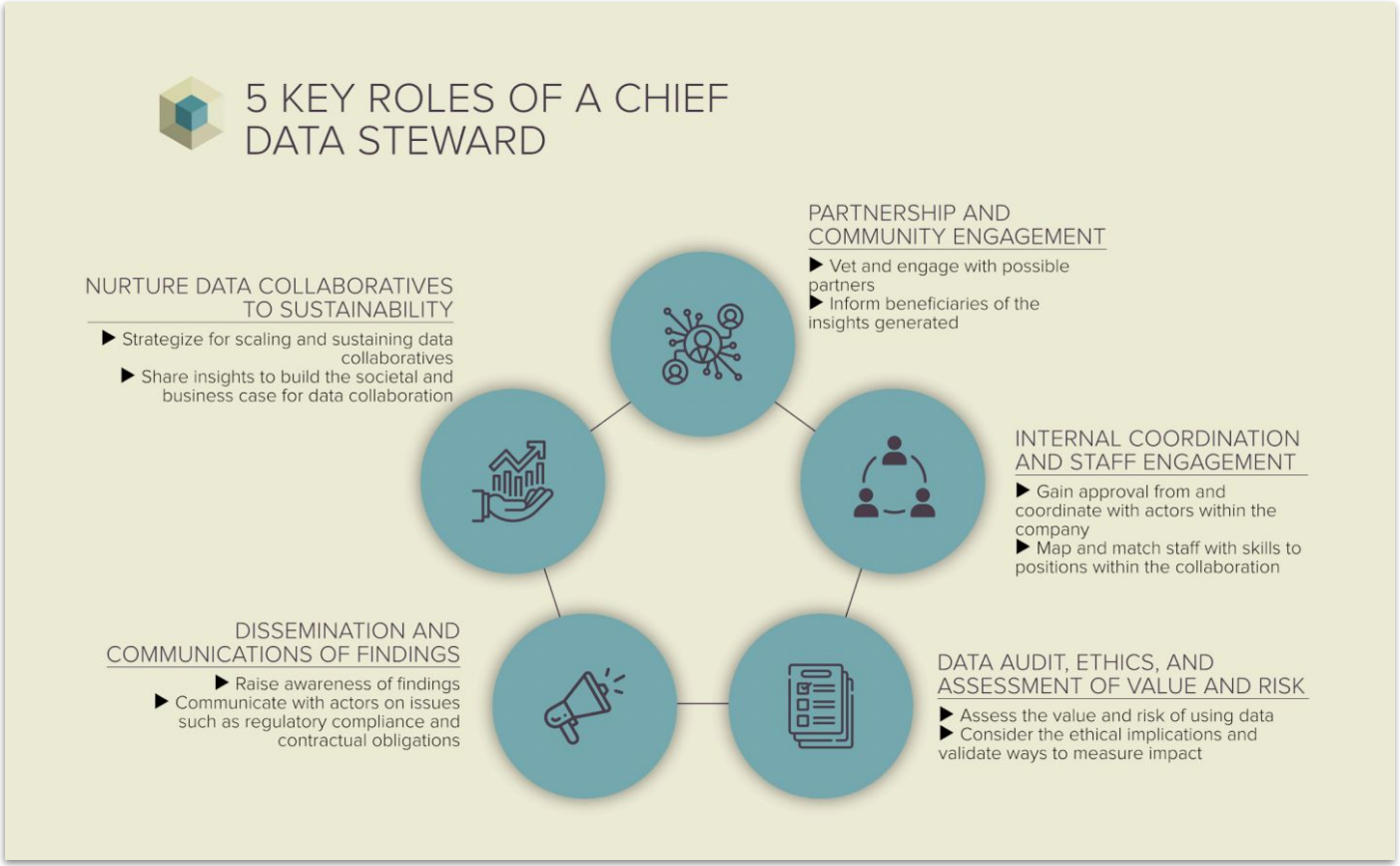
THE ROLE OF A DATA STEWARD

Data Stewardship is about making data collaboratives more:

- Systematic
- Sustainable
- Responsible



THE ROLE OF A DATA STEWARD



DATA AUDIT, ASSESSMENT & GOVERNANCE

Determining and assessing the value, potential, and risk of data held and needed within an organization

“Stewarding Data Assets for and in the Public Interest”

- Help Formulate and Determine **priority questions** (vis-à-vis Value Proposition or Problem Definition)
- **SCOPING and ITERATING:** Assess “minimum viable” data needed vis-a-vis the questions at hand
- Identify and document data assets
- Consider the ethical and fundamental rights implications and other **risks** of using (or not using) data
- Help establish operational, technical and governance models that are **“fit for purpose”**
- Validate ways to measure impact

Photo by Chris Liverani on Unsplash

Photo by Hannah Busing on Unsplash

PARTNERSHIP AND COMMUNITY ENGAGEMENT

Proactively and responsively reaching out to and vetting potential partners or users.

“Stewarding Relationships”

- **EXTERNAL RELATIONS:** Be the point of contact regarding re-use of data
- Identify, map, vet and engage with possible relations, partners and other stakeholders
- **USER-DRIVEN DESIGN:** engage users of data products and insights
- Help establish the **“SOCIAL LICENSE”** of re-using data through deliberation and community engagement
- Establish data agreements and other contractual relationships

INTERNAL COORDINATION AND DATA OPS

Securing internal coordination and establish data operations

“Stewarding Internal Resources, Expertise and Authorities”

- **INTERNAL RELATIONS:** Gain approval from and coordinate with actors within the organization;
- Ensure all internal stakeholders and organizational leadership are informed and aligned.
- **DATA OPS:** Map and match Internal resources, expertise and skills needed to enable data collaboration

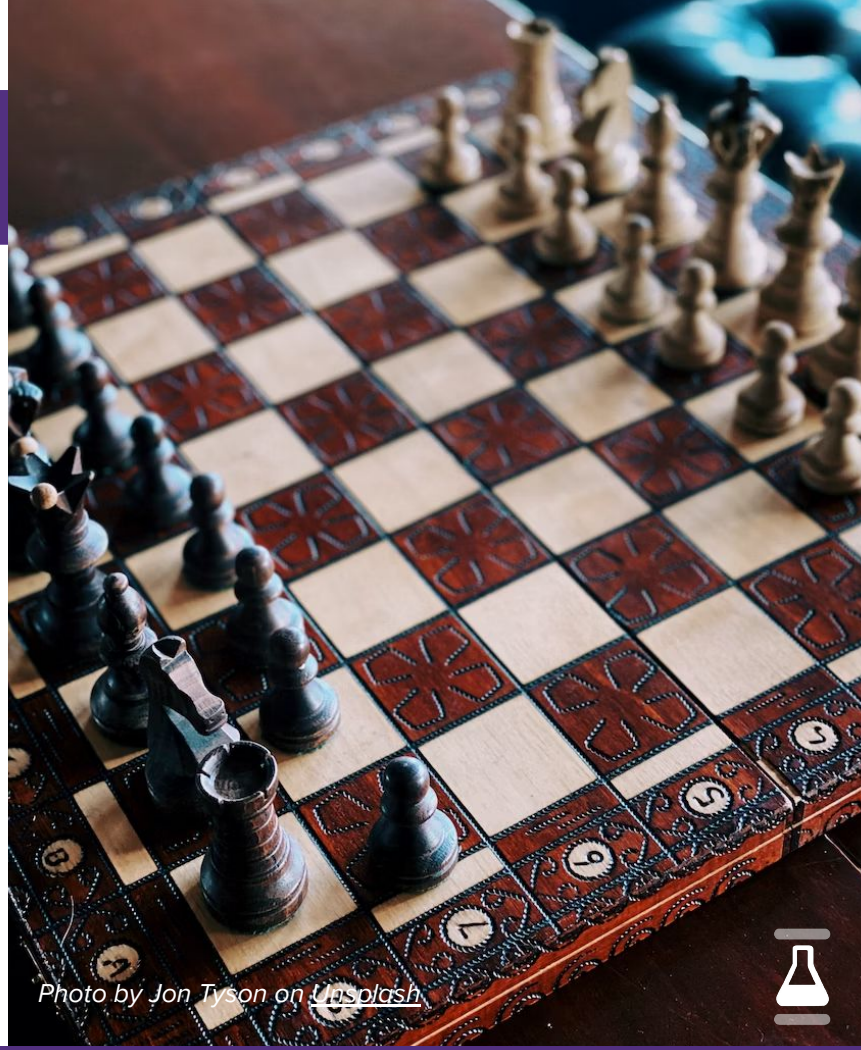


Photo by Jon Tyson on Unsplash

Photo by Annie Spratt on Unsplash

NURTURE DATA COLLABORATIVES TO SUSTAINABILITY

Gathering the needed resources and support so as to ensure broad and long-term impact.

“Stewarding Sustainability”

- **INSTITUTIONALIZE DATA INNOVATION:** Make re-use of data systematic (and institutional)
- **DEVELOP THE BUSINESS CASE:** Strategize for scaling and sustaining of data innovation.
- **EVALUATION:** Measure impact and share insights to build the societal and business case for data collaboration

DISSEMINATION AND COMMUNICATION OF FINDINGS

Raising awareness, disseminating findings and communicating outcomes from data collaboratives

“Stewarding Insights”

- **COMMUNICATIONS:** Raise awareness of insights with users, partners, government and other stakeholders
- Enable the translation of data intelligence into decision intelligence
- Communicate with actors on issues such as regulatory compliance and contractual obligations

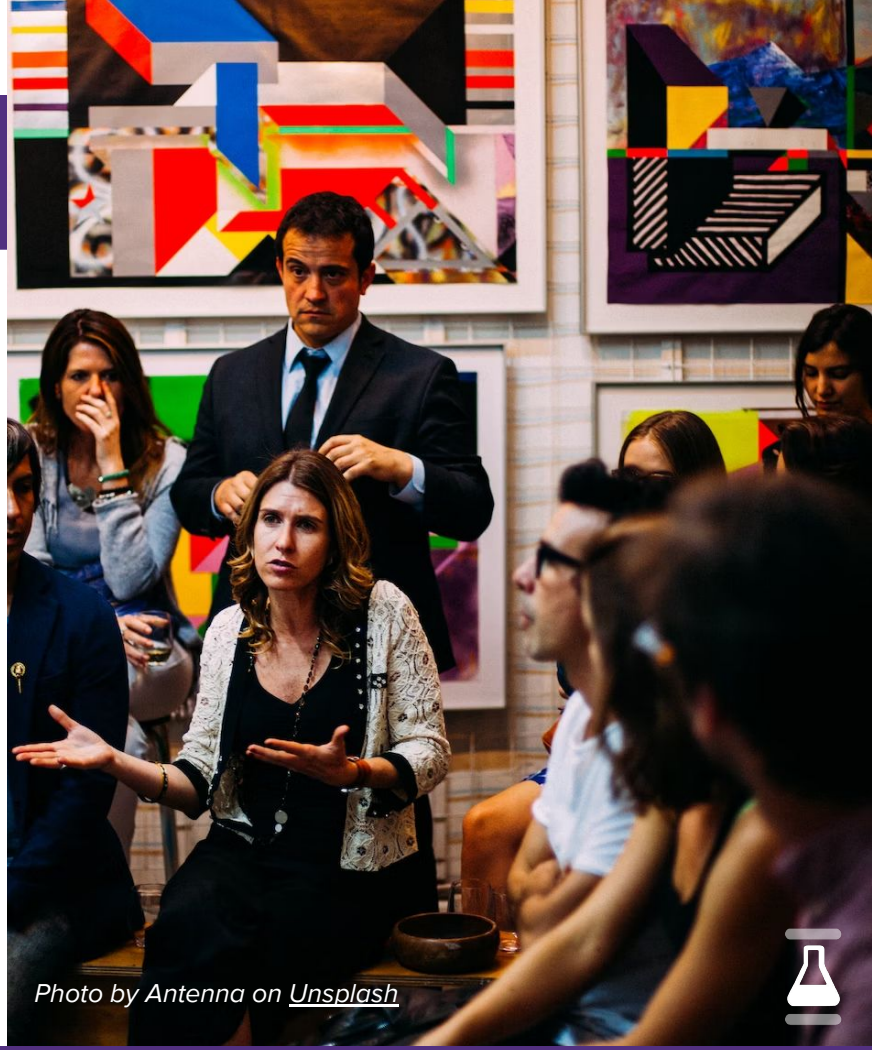
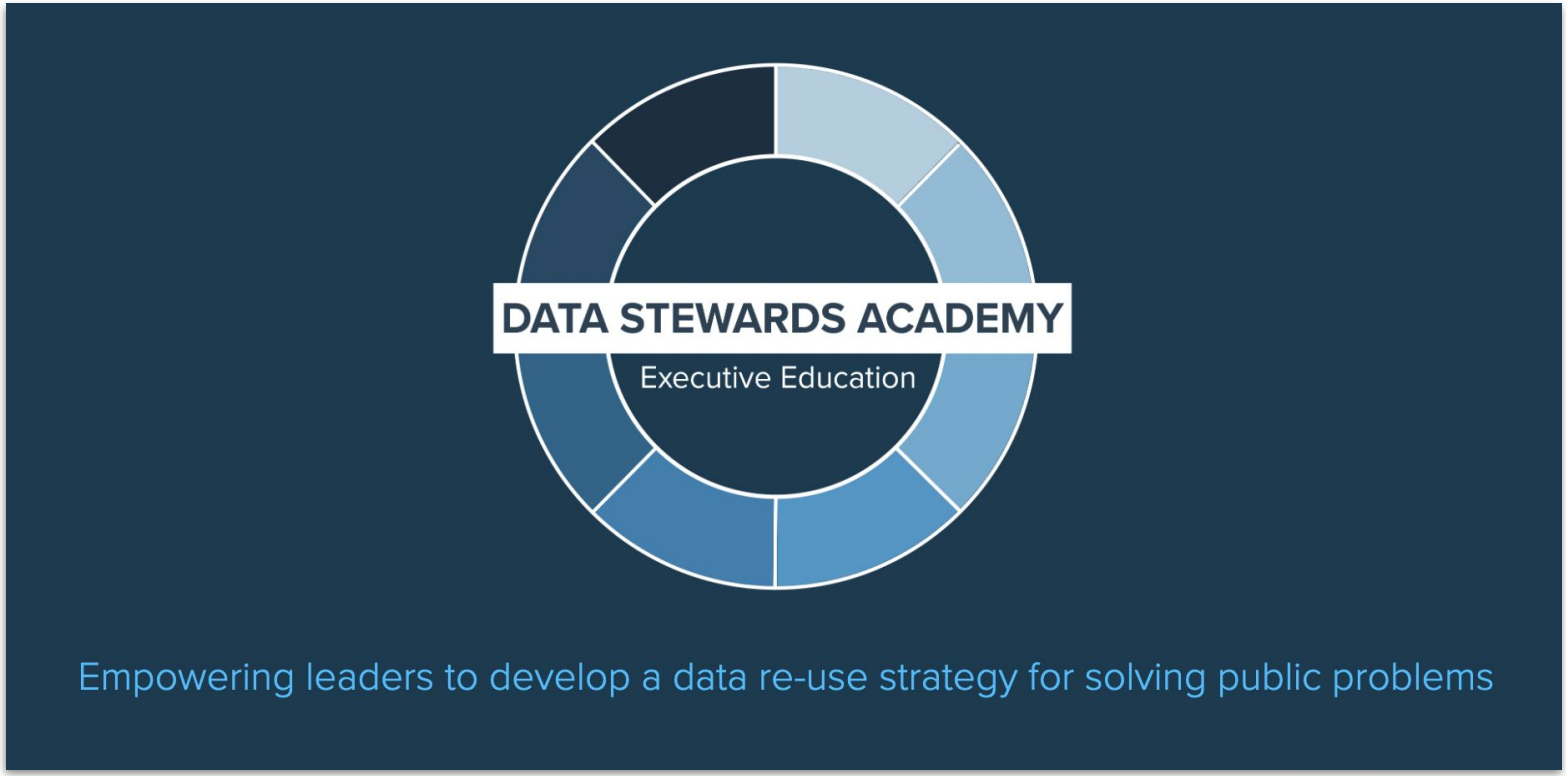
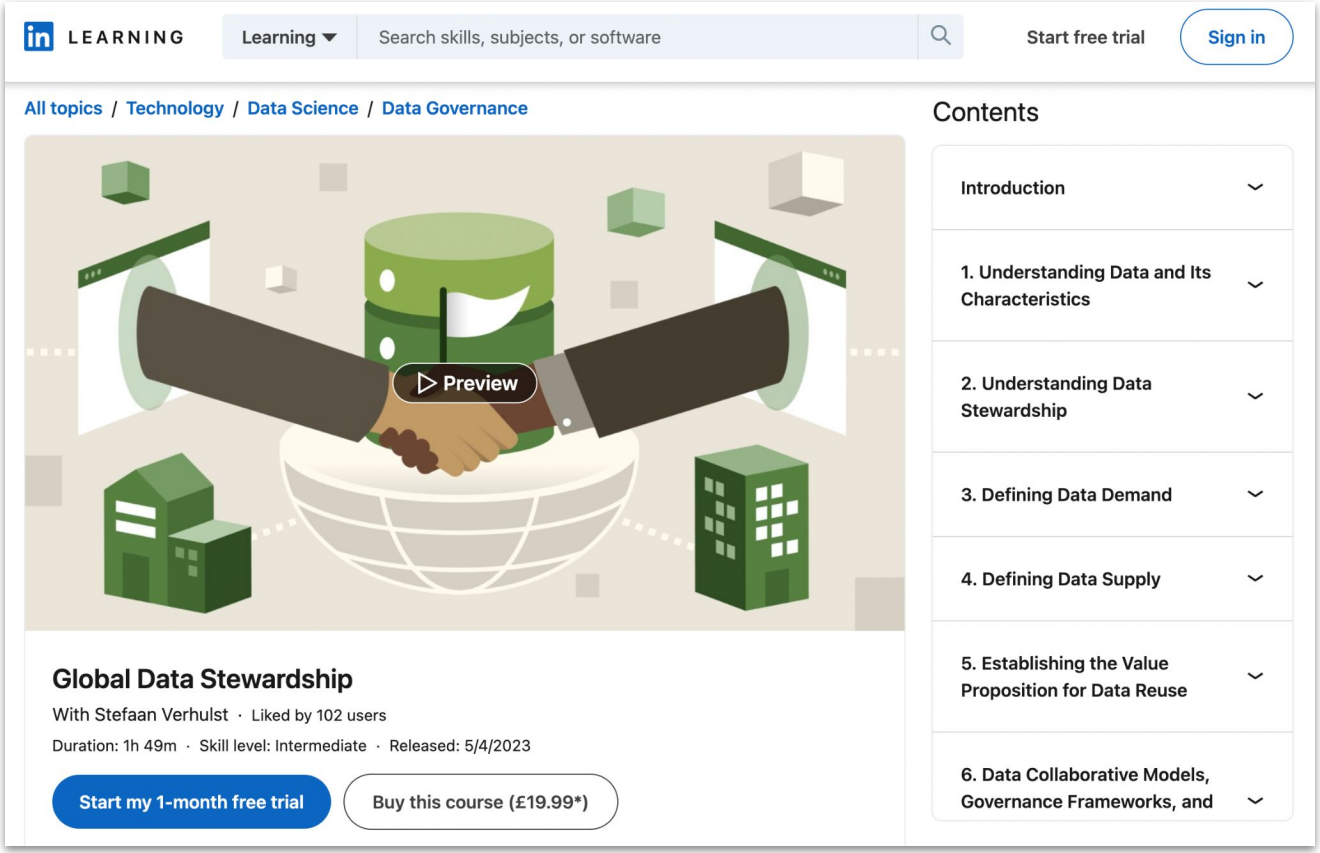


Photo by Antenna on Unsplash

THE DATA STEWARDS ACADEMY



GLOBAL DATA STEWARDSHIP



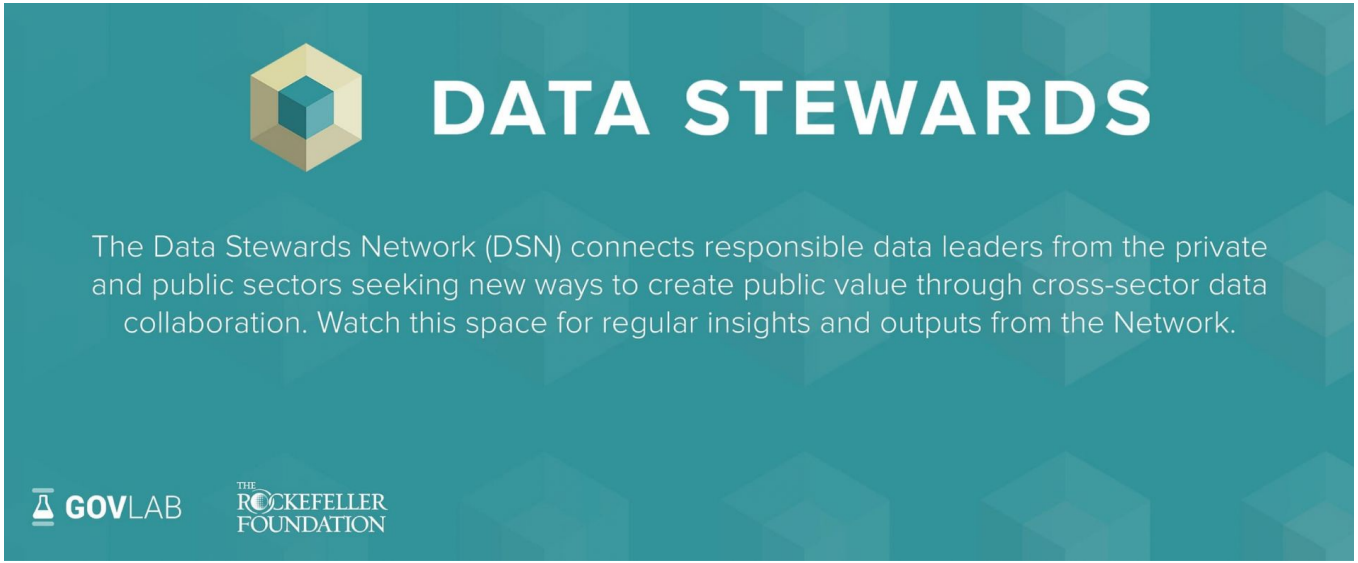
BUILDING DATA COLLABORATIVES

1. Data collaboratives can help bridge data asymmetries.
2. When establishing data collaboratives, consider what is fit for purpose.
3. Different data collaborative models are effective in different contexts.
4. Determining the value proposition behind a data collaborative makes it more sustainable.

BUILDING DATA COLLABORATIVES

5. Establishing legitimate governance enhances responsibility.
6. A technical infrastructure is needed to strengthen a data collaborative.
7. A data steward plays a key role towards making data collaboratives more systematic.

STAY IN TOUCH & RECEIVE UPDATES



Data Stewards Network

https://medium.com/data-stewards-network

@TheGovLab

datastewards@thegovlab.org

@The Governance Lab

www.thegovlab.org