

Module 7: Data Collaboratives

Welcome back!

In earlier modules, we talked about the value of using data to define problems and why and how to use open government data as one source of information.

Today, we look at another source of data, namely private sector data, and how to gain access to it for use in defining your problem. We call such exchanges of data and talent between the public and private sectors a data collaborative.

By the end of this module, you will:

1. Understand what a data collaborative is and its value for solving public problems;
2. Be able to describe the incentives for private companies to collaborate with other sectors in the public interest;
3. Know the process for establishing data collaborative and the key considerations involved.

Let's start our discussion of data collaboratives with an example.

In 2017, the Category 5 Hurricane Irma tore through the Atlantic, eventually striking the state of Florida. In times of crisis, when thousands of people are forced to evacuate their homes on short notice, information about fuel is critical. In the face of Irma, like with so many disaster situations, the demand for fuel spiked while supply dropped due to the sudden increase in long-distance travel.

Without up-to-date information, people seeking to relocate to a safe place might find themselves looking for fuel in the wrong place, diminishing their chance for a successful evacuation and risking their lives.

GasBuddy--a tech company that provides real-time fuel prices from 140,000 gas stations in the United States, Canada, and Australia--is in a unique position to supply information related to fuel during a crisis situation. GasBuddy provides a mobile and web application that shares data on fuel prices, fuel types, and other information about gas stations in their area. Much of the data is crowdsourced from GasBuddy's user base.

In response to Hurricane Irma, GasBuddy and the Florida Division of Emergency Management (DEM) launched a Data Collaborative to ensure that government agencies and the public had access to the most accurate and up-to-date information about the fuel they needed to evacuate safely. GasBuddy analysts travelled to Tallahassee and volunteered to facilitate real-time data sharing between the company and response agencies in the State Government. The analysts also shared their expertise by helping to analyze the fuel data as it arrived to inform government decision-making and public communications. Rick Scott, Florida's governor at the time, took to television and social media to encourage people to use GasBuddy to check fuel availability and to share information through the app for the benefit of others.

The GasBuddy-Florida DEM Data Collaborative enabled evidence-based refueling efforts based on analysis of various public and private data streams.

The public-private partnership also resulted in Florida evacuees accessing - and contributing - up-to-date fuel information in massive numbers. During Hurricane Irma, GasBuddy jumped from being the 57th to the second-most downloaded iPhone app in the US, in just 24 hours. It was downloaded 350,000 times on September 7, 2017 – when the storm was quickly moving through the Caribbean and toward Florida – compared with 30,000 downloads on a typical day.

So - what are data collaboratives? Data collaboratives are a new form of collaboration, in which participants from different sectors—and in particular, private companies—exchange their data to create public value.

Data Collaboratives are an emerging field building on existing movements – like open data, public-private partnerships, and corporate social responsibility – to help actors from across sectors work together to solve public problems.

Around the world, there is an increasing recognition that corporations can build new partnerships around data sharing (or donation) to help governments, international organizations and NGOs solve problems more effectively.

Thus, when looking for data for your own problem-solving practices, consider private sector data as a potentially viable option.

Generally speaking, there are six types of Data Collaborative.

There are **Data Cooperatives or Pooling Data Collaboratives**, in which corporations and other important dataholders group together to create “data pools” with shared data resources.

For example: The National Institutes of Health, the U.S. Food and Drug Administration, 10 biopharmaceutical companies and a number of non-profit organizations are sharing data to create new, more effective diagnostics and therapies for medical patients based on an improved understanding of drugs’ effects on certain biomarkers.

Prize and Challenge Data Collaboratives are those in which corporations make data available to qualified applicants who compete to develop new apps or discover innovative uses for the data.

The GBDX for Sustainability Challenge gave innovative problem-solvers access to DigitalGlobe’s useful geo-spatial and satellite imagery data as they tried to address problems across topics like disaster response, food security, and poverty elimination.

Corporations may also share data with universities and other academic organizations in **Research Partnership Data Collaboratives** that give researchers access to consumer datasets and other sources of data to analyze social trends.

For example, in partnership with the mobile provider Safaricom, the Harvard School of Public Health project analyzed aggregated cell phone data from 15 million people in Kenya together with the country’s Malaria Atlas to better understand the spread of the disease and support eradication efforts.

A fourth kind of collaborative is an **Intelligence Product Data Collaboratives**. These use shared, and often aggregated, corporate data to build a tool, dashboard, report, app or another technical device to support a public or humanitarian objective.

For example, The LinkedIn Economic Graph relies on user data to help spot trends in talent migration as well as in-demand skills. Organizations like the World Bank, Australian Treasury, and National League of Cities use the data to connect people to job opportunities.

“Application Programming Interfaces” - also known as APIs are collaboratives that allow software developers and others to access data for testing, product development, and data analytics.

For example, Ashoka Trust for Research in Ecology and the Environment, for example, uses the Google Earth Outreach API to provide forest information to policymakers to help them create strategies for protecting tigers and elephants in Indian forest reservations.

The final type of collaborative is the **Trusted Intermediary Data Collaboratives** in which corporations share data with a limited number of known partners. Companies generally share data with these entities for data analysis and modelling, as well as other value chain activities.

For example, Social Science One is run by a commission of senior academics who act as a trusted intermediary between Facebook and academic researchers interested in using the social network’s data to study disinformation. The commission judges which research proposals have the most merit and brokers access to relevant Facebook data.

So, why do data collaboratives matter?

Recent years have seen exponential growth in the amount of data being generated and there is increasing recognition that this data can play a key role in solving some of the most difficult public problems we face.

However, much of the potentially useful data is currently privately held and not available for public insights. Data such as online purchasing, web searches on a search engine, mobile phone records and satellite location data are typically held private sector companies.

Companies also generate an ever-growing stream of information from our proliferating sensors and devices.

Data Collaboratives offer a way to access this data to enable its use for better understanding and addressing public challenges.

Data collaboratives can be useful for solving public problems but the natural question is “what’s in it for companies providing their data?”

The current practice of Data Collaboratives around the world suggests there are six primary motivations - the 6 “R’s” - that can be leveraged to get private-sector data holders on board with a data collaborative.

First, **Reciprocity**.

Corporations may share their data with others for mutual benefit, especially when gaining access to other data sources that may be important to their own business decisions.

Second, **Research & Insights**.

Opening up corporate data may generate new answers to particular questions, providing companies insights that may not have been extracted otherwise and help to identify new niches for activity or inform new business models.

Data collaboratives can also help corporations tap into data analytical skills outside of their own company.

Third, **Reputation and Talent Retention**.

Sharing data for public good may enhance a firm's corporate image and reputation, potentially attracting new users, customers, and investors who value socially conscious corporate actors. Data collaboration in the public interest may also offer an opportunity to gain (free) media attention.

The opportunity to work on projects that could positively impact people's lives can also be valuable in attracting and retaining data science talent.

Fourth, **Revenue Generation**.

Opening up corporate data does not always have to be for free. Under some conditions, corporate data may be offered for sale, generating extra revenue for firms in business-to-business and business-to-government arrangements.

Fifth, **Regulatory Compliance**.

Sharing data can also help corporations comply with sectoral regulations and become more transparent and trusted.

In addition, many corporations generate specific datasets for the sole purpose of regulatory compliance. Sharing and using that data in a responsible manner for public and private benefit may increase the return on investment for that original data collection.

Our sixth and final R is **Responsibility and Corporate Philanthropy**.

Sharing corporate data achieves many of the goals sought by traditional corporate social responsibility or philanthropy. Companies can derive value from socially responsible behavior not just because of the positive image such an activity produces, but because opening up data can also improve the competitive business environment within which the business operates.

So there's good reasons for both public and private sector entities to consider a data collaborative.

Now, let's explore the process for establishing one.

We can organise this process into 8 overarching phases, and break these 8 phases down into 18 more granular steps that help you think about the key tasks that need to be attended to.

The first phase in establishing an effective data collaborative requires us to consider the demand side of the initiative: that is, an in-depth understanding of the problem to be addressed and the opportunity provided by cross-sector data sharing.

Here, the key steps are to ensure we have defined the problem well and that we are clear about the value proposition of the data collaborative, such as improved situational awareness, prediction or forecasting.

- Step 1: Define The Problem To Be Solved
- Step 2: Define The Value Proposition Of The Data Collaborative

Once the demand side of the equation is well understood and articulated, the supply of data and expertise should be explored to determine how it can address the demand.

This involves identifying what's available in-house and where external entities could fill any gaps in data or data science skills.

A data audit can help you to understand how useful the data you already have access to is and will inform your search for external data sources.

At this point, it's crucial to check that there's a match between supply and demand - if there's not, a data collaborative is unlikely to be effective for the issue at hand.

- Step 3: Identify data science expertise and organizational competency – internal competencies and external entities that could fill any gaps
- Step 4: Identify data supply
 - Internal: Conduct due diligence research and a data audit
 - External: Map the potential supply side based on identified gaps

Data collaboratives - by definition - are collaborations.

Choosing the right supply-side partner and adequately incentivizing them to collaborate are key steps for making the data collaborative a reality. The key task here is to identify the best data providers and identify specific incentives that will encourage their participation.

- Step 5: Select the most promising potential supply-side data providers and identify specific incentives for them to participate

The design phase involves determining the specifics of how the data will be shared and used, an assessment of the most salient risks and likely harms, the creation of a targeted strategy for mitigating those risks, and work to determine the ongoing governance framework for the data collaborative.

Steps six through ten are directed at helping you to do this.

- Step 6: Define the ideal type of data collaborative based on the supply and demand
- Step 7: Assess major risks, ethical concerns and potential challenges
- Step 8: Develop a multi-faceted risk mitigation strategy
- Step 9: Agree upon terms and conditions for arrangement
- Step 10: Establish a governance structure – how will decisions be made?

By phase five, all parties should have a clear understanding of how the arrangement will work, the key risks and strategies for addressing them, and the processes that will ensure the data collaborative runs as expected.

At the implementation stage, the data collaborative is launched, but only after expectations, roles, timeline and cost questions are answered. As with any product launch, the creation of a data collaborative requires thinking about what will happen, when, by whom and in what order.

To cross the chasm from idea to implementation and to persuade others to take the steps necessary to embark on a new plan of action, you must be able to draft an implementation memo that lays out the steps of the project from data sharing to use to analysis to evaluation.

The memo also needs to address the resources that need to be secured for the data collaborative to become sustainable. This goes beyond mere financial budgeting to include physical, human, data assets and cultural conditions needed to be successful.

- Step 11: Agree upon expectations, roles, responsibilities, timeline and operational specifics of data-sharing process
- Step 12: Determine resources: cost and funding models

In phase 6, we need to think about how we will engage the intended beneficiaries for the data collaborative and the communities that could help to promote the effort, such as the media.

This involves creating a multi-faceted communications strategy and an approach for sharing information on an ongoing basis.

- Step 13: Develop a communications strategy
- Step 14: Determine audience and information sharing approach

To maintain the data collaborative over time, including maintaining funding or accessing new funding, it's important to build in learning about the impact it is having.

This means measuring success against an agreed-upon baseline and metrics, and checking that the current mix of variables and decisions are yielding the optimal approach for the data collaborative.

- Step 15: Define a common baseline against which to measure progress
- Step 16: Measure progress against defined, agreed-upon metrics of success
- Step 17: Impact assessment

Our final step is a commitment to iteration.

Our starting assumption should be that our work can always be improved. This means explicitly creating opportunities to identify and make changes to the data collaborative that will enable this improvement.

- Step 18: Iterate as needed

That concludes our discussion of data collaboratives. Next, we look at a form of collaboration that seeks to tap into the knowledge and expertise outside an organization that is even more widely distributed and focus it on our problem. We call this form of collaboration open innovation.