

# Data Sharing for Research CASE STUDY



## GRAVY ANALYTICS

### Executive Summary

The [Future of Privacy Forum](#) (FPF) analyzed a diverse sample of data sharing partnerships between companies and academic researchers and produced a series of case studies distilling our findings. We learned that there is broad consensus regarding the potential benefits of industry/academic data sharing partnerships, including the acceleration of socially beneficial research, enhanced reproducibility of research breakthroughs, and broader access to valuable data sets. At the same time, companies and academic researchers understand and take steps to mitigate risks — particularly ethical and data protection risks. Increasingly, stakeholders are identifying risks arising from re-identification threats or data breaches while acting to mitigate those risks through the use of data sharing agreements (DSAs) and Privacy Enhancing Technologies (PETs).

FPF's analysis of corporate-academic data sharing partnerships provides practical, evidence-based recommendations for companies and researchers who want to share data in an ethical, privacy-protective way. These case studies demonstrate that corporate-academic data sharing partnerships offer compelling benefits to companies, research, and society. Risks exist, but effective mitigation strategies can reduce the likelihood of harm to individuals, communities, and society. For many organizations, data sharing partnerships are transitioning from being considered an experimental business activity to an expected business competency. This trend is most pronounced among established firms; it is an opportunity for researchers to access new data for scientific discovery.

[Read the full Case Study Report.](#)

### Data Sharing Type: Closed Trusted Partnerships

## ORGANIZATION AND PARTNERS

### Company

Founded in 2011, Gravy Analytics is a location technology company based in Virginia with about 60 employees and a reported annual revenue of \$16.9 million in 2022.<sup>1</sup> The company primarily provides location data and analytics to other companies but also maintains a [Data for Social Good](#) program, which offers reduced rates for research-based institutions.

### University of Florida

Dr. Xilei Zhao is an Assistant Professor in the Department of Civil & Coastal Engineering at the University of Florida. Dr. Zhao's research team maps the evacuation flows of people affected by natural disasters to inform public safety decisions. For example, the team explores how specific wildfires impact population movements, such as who chooses to evacuate, who doesn't, when, how they move, and why. Dr. Zhao partnered with Gravy Analytics on her team's research to map people's movement during the 2019 Kincade Fire in California and the 2021 Marshall Fire in Colorado. Their goal is to eventually be able to forecast and create real-time emergency management tools to inform responses to future disasters.

### Columbia University

Dr. Sandra Matz formed a data sharing partnership with Gravy Analytics through the Data Science Institute at Columbia. Dr. Matz is an associate professor in the Business School at Columbia University. Trained as a computational social scientist, Dr. Matz studies people's online and offline behavior and the relationship between the events and locations they frequent. Dr. Matz wanted to be able to match an individual's psychological profile with the content they interact with online and where they physically spent their time, such as attending a concert or a political gathering. Her theory was that these factors can contribute to people's sense of identity, political ideology, and social values. Dr. Matz combined traditional psychology methods in controlled settings with Gravy Analytics data, allowing her research to investigate questions at a larger scale in applied settings and potentially infer causality.

### Johns Hopkins University

Dr. Anton (Tony) Dahbura is the Co-Director of the Institute for Assured Autonomy and the Executive Director of the Information Security Institute at Johns Hopkins University. In 2020, as the COVID pandemic began, Dr. Dahbura and his team wanted to develop a community-level data center to

provide accurate epidemiological models for how respiratory infections like COVID spread throughout a community and what kinds of interventions are most effective. Traditional epidemiology often uses agent-based modeling, where researchers create simulated people and give them activities, such as the need to find food or go to sleep, and then track how the simulated agents move around. However, agent-based models may not behave like real people within their communities, so Dr. Dahbura wanted to be able to model his simulation using real location data from Gravy Analytics to increase its accuracy.

## PARTNERSHIP CONSIDERATIONS

### Data Access and Sharing

Researchers can apply to purchase data from Gravy Analytics at a reduced price through an online form on the Gravy Analytics' website. Company representatives said that applications are reviewed internally for feasibility, capacity, and mission alignment. Their application vetting process usually begins with a call with the requester to discuss their qualifications, institutional support, privacy considerations, the data requested, and research questions to be addressed. Gravy Analytics formalizes partnerships with a standard data services agreement (DSA), which it can adapt based on the partner's needs. Representatives at Gravy Analytics indicated that private universities have demonstrated more flexibility in negotiating DSAs than larger public universities. In either case, Gravy depends on the institution for compliance with the agreement terms and data governance.

Each data request requires work from multiple teams, and Gravy Analytics estimates a single request can take up to 50 hours of internal work before anyone receives data. Data is shared only with the granularity required to answer research questions, which varies by use case. Gravy Analytics distributes data under a non-disclosure agreement (NDA), but, in concert with its DSA, it allows the researchers' results and findings to be shared and published. Every request requires data cleaning, organization, and the application of privacy techniques, as well as additional researcher obligations to protect data privacy and security.

University of Florida researchers identified several factors that went into their decision to partner with Gravy Analytics, including data accuracy and data use time limits. Gravy Analytics offered a three-year contract that met the needs of the team's research and grant timeline. Dr. Zhao recommended that researchers ask for sample data sets from potential data sharing partners. After comparing different companies, Dr. Zhao selected Gravy Analytics and

signed an NDA for the data purchase. Dr. Zhao combined the purchased data from Gravy Analytics with other open-source data such as census, demographic, parcel, and land use data. All data purchased was funded by a NIST grant supporting the research project.

At Johns Hopkins University, Gravy Analytics provided Dr. Dahbura with about one month's worth of location data for the state of Oklahoma. The choice of state allowed the research team to demonstrate how COVID spreads in more isolated communities, as opposed to denser suburbs or large cities. Dr. Dahbura mentioned that the Institutional Review Board (IRB) process, data request, and transfer all went much faster than normal because the research was related to a public emergency (COVID), but generally, he felt that his research was well supported by the institution and legal team that worked to affect the data sharing partnership.

## Privacy, Publishing, and Teaching

The University of Florida research team was given data that had already undergone aggregation and some anonymization from Gravy Analytics, but which still allowed them to answer their research questions. Despite these precautions, the research team decided that the shared data was sensitive enough to warrant storage in a secure environment, access restrictions, and several privacy techniques to minimize reidentification before publishing.

Dr. Matz noted that, in her experience with corporate data sharing, most companies expect a faster turnaround on the research produced from data sharing than academic publishing allows. Managing expectations with Gravy Analytics early on helped avoid conflicting expectations for project and publication timelines. Dr. Matz mentioned that corporate data sharing is largely impractical for students to engage with independently, as just getting DSA and IRB approvals can take over a year to obtain. Students often can't wait that long to access data; however, if they work with a faculty member involved in a long-term data sharing project, students can meaningfully participate.

## RISKS AND BENEFITS

### Risks

Company representatives said that location data is often very sensitive, so all data sharing must go through internal privacy and security checks to minimize risks to end users. Sharing location data with external partners may create reputational risks for Gravy Analytics and research partners, even when the underlying research is widely perceived as meritorious.

Columbia University offered Dr. Matz institutional support for the partnership with Gravy Analytics, involving the IRB as well as its legal team, which addressed ethical risks and legal obligations. While Dr. Matz uploads descriptions of the shared data, the methods used, and her code in an Open Science Framework instance, she explained that research reproducibility could be difficult with data sharing partnerships.<sup>2</sup> Many DSAs mandate data erasure after a certain period as well as prohibiting data sharing outside the research team. Meanwhile, some academic publishers require that the

data be kept by researchers for five years. Other researchers who want to reproduce her research based on the data sharing partnership could potentially petition Gravy Analytics for the data, but the decision to share it and at what price would be up to the company.

At Johns Hopkins University, Dr. Dahbura was concerned about the risks of handling sensitive location information, so the team worked with experts in cryptography and edge computing. They also processed the shared location data using AI to create a synthetic version of the data to model mobility patterns. This way, the model never directly used actual location data, significantly reducing reidentification risks. This method provides greater confidence in the simulation outcomes and increases the potential scale and demographic accuracy.

### Benefits

Gravy Analytics concluded the benefits of partnering with research institutions outweigh the risks, including helping to legitimize socially beneficial uses of location data, showcasing what can be done, and pushing the industry forward. These partnerships have also been shown to generate positive public relations and marketing for the company.

Dr. Zhao indicated that Gravy Analytics data allowed her to answer questions using novel methods. Previous methods in disaster evacuation studies involved sending surveys to the people impacted by a disaster after the fact. This sampling method created many problems with accuracy and sampling bias. Using GPS data to map population movement during a natural disaster is more accurate and offers new insights for how to increase public safety and possibly reduce future injuries or fatalities caused by natural disasters.

Dr. Matz commented that attempting to collect or generate the data she received from Gravy Analytics on her own would have been nearly impossible and taken much longer. Dr. Matz flagged a benefit of using data controlled by companies: very few people will have conducted research on it. Publishing in peer-reviewed journals, a necessity for many academics, generally requires researchers to explore novel ideas or methods. In disciplines forced to rely on a limited set of open data, it's challenging to find an idea or method that someone else hasn't already published about. Until there is more open data for researchers to work with, corporate data sharing offers the best chance for those researchers to investigate new data, publish their results, and advance their discipline.<sup>3</sup>

Dr. Dahbura noted the importance of accurate epidemiological modeling, not just for initial public safety decisions in a pandemic but also because research suggests that the public quickly loses trust in epidemiology after a single inaccurate forecast. Once epidemiology loses public trust, people are less likely to adhere to public health precautions in the future. Dr. Dahbura maintains that his team's research is a great example of how sensitive information like location data is essential for public safety, notwithstanding the potential risks of using the same kind of data in other ways. Dr. Dahbura sees his work as attempting to develop precision public health efforts, a cousin of precision medicine, to produce community-specific mitigation protocols.

### PARTNERSHIP INFORMATION

Gravy Analytics: [gravyanalytics.com](https://gravyanalytics.com)

Gravy Analytics, Data for Social Good Program: [gravyanalytics.com/data-social-good](https://gravyanalytics.com/data-social-good)

Dr. Xilei Zhao, University of Florida: [essie.ufl.edu/people/name/xilei-zhao](https://essie.ufl.edu/people/name/xilei-zhao)

Dr. Sandra Matz, Columbia University: [sandramatz.com/](https://sandramatz.com/)

Data Science Institute at Columbia University: [datascience.columbia.edu](https://datascience.columbia.edu)

Johns Hopkins University: [jhu.edu](https://jhu.edu)

Dr. Anton Dahbura, Johns Hopkins University: [engineering.jhu.edu/faculty/anton-dahbura](https://engineering.jhu.edu/faculty/anton-dahbura)

### RESEARCH PAPERS FROM THE PARTNERSHIPS

- Estimating wildfire evacuation decision and departure timing using large-scale GPS data
- A highway vehicle routing dataset during the 2019 Kincaid Fire evacuation
- Wildfire evacuation decision modeling using GPS data

To learn more about data sharing partnerships, read [The Playbook: Data Sharing for Research](#) or join the [Ethics and Data in Research Working Group](#) for updates on legislative developments and monthly calls with experts.

*This project is supported by the Alfred P. Sloan Foundation, a not-for-profit grantmaking institution whose mission is to enhance the welfare of all through the advancement of scientific knowledge.*

<sup>1</sup> <https://www.cbinsights.com/company/gravy-analytics/financials>

<sup>2</sup> The Open Science Framework (OSF) is a platform for sharing the products of a research lifecycle such as data, protocols, study designs, reports, or publications. See <https://osf.io/>.

<sup>3</sup> Matz acknowledged the valuable support of the Data Science Institute (DSI) at Columbia. As outlined by DSI's Executive Director, Sharon Sputz, DSI has a lot of experience negotiating data sharing agreements and engaging the Columbia Sponsored Projects Office and lawyers. Sputz attends to the publication needs of researchers and their students, which often militate against non-disclosure agreements. She remarked on the tensions between research and privacy, a topic being addressed through an NSF grant.