# Data Sharing for Research
## CASE STUDY

### LINKEDIN

**FUTURE OF PRIVACY FORUM**

## 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 Partnership

## ORGANIZATION AND PARTNERS

### Company

Founded in 2003, LinkedIn is an employment-centered social media platform with over 900 million registered users, reported an annual revenue of $13.8 billion in 2022, and has more than 21,000 full-time employees. In 2016, LinkedIn became a subsidiary of Microsoft.

## PARTNERSHIP CONSIDERATIONS

### Data Sharing Partnerships

LinkedIn has a specialized team of data scientists and public policy managers who administer its Data for Impact program, which is the primary mechanism LinkedIn uses to share aggregated, anonymized datasets with external partners at no cost. According to company representatives, there are generally three forms of data sharing partnerships: project-based data sharing (often in the context of long-term institutional relationships), observatory sharing, and collaborative research. The first category is typically one-time data requests, and their interactions with the project teams are relatively quick, even if their relationships with the institutions are long. The second category is when LinkedIn provides regular delivery of new data to an external partner, with the value being the consistency and recency of data across time. The third category, by contrast, requires many in-depth consultations because they engage with institutional staff on building a new indicator or co-authoring a report. The research questions, goals, internal costs, technical requirements, and privacy protections influence whether LinkedIn agrees to share data and, if so, what that data sharing looks like. A typical partnership launch involves a meeting to understand the researcher's request and how those align with data and privacy considerations. Often, partners want more

granular data than LinkedIn is willing to give. Next, LinkedIn conducts an internal assessment of the privacy risks and methods needed to execute the request. If approved, LinkedIn delivers the data to the external partner and then works with them to ensure the requestor's intended analysis aligns with the shared data to ensure methodological quality.

LinkedIn representatives stated they receive project-based data requests through the Development Data Partnership and the Industry Data for Society Partnership. These are usually used when researchers intend to produce a research product such as a report, organizational strategy document, or peer-reviewed publication. Most of the communication between partners and LinkedIn is concentrated between the initial request and the data shipment and then at key points along the process toward publication. Data observatories and embedded data products operate differently. An example of a data observatory is LinkedIn's partnership with the Inter-American Development Bank's Labor Market Observatories or the German Federal Statistical Office (Destatis). LinkedIn has explored long-term research relationships with entities like the Institute for Employment Research (IAB). These partnerships require more coordination and investment but also provide potentially higher-impact research outcomes.

### Data License Agreements

LinkedIn has a standard data license agreement (DLA) and will only modify it slightly, if necessary, based on the institution they are partnering with. The company's representatives commented that it would be difficult to develop unique DLAs for each institution, thus their use of a standard DLA as much as possible. Their DLA focuses on guaranteeing LinkedIn member privacy, meaning no personally identifiable information or data is ever shared with external partners, and it also requires LinkedIn to review research partners' drafts prior to publication to ensure data is being interpreted and used correctly.

## Data Menu

LinkedIn offers a public-facing Data Menu that displays a list of datasets offered to external partners. Company representatives emphasized that datasets listed on the menu undergo several layers of review for quality and are continually supported. The current categories of data include 1. LinkedIn hiring rate (their most popular), 2. Career Transitions, 3. Skills Genome, 4. Skills Similarity, and 5. Skills Penetration. All data in the menu is aggregated, anonymized, and can speak to labor market dynamics in 80 countries. LinkedIn offers various scheduled refreshes on data depending on the indicator, and data availability changes over time. Recently added datasets include indicators for gender and the green economy. To support a service like the Data Menu, LinkedIn has a continual updating, aggregation, and review process for quality and privacy. Representatives claimed that the data menu integration process gives partners more confidence in the data, as these data sets have gone through additional internal vetting and have been used by other partners. They also communicated that they'd like to move towards more automation for data sharing and are exploring using differential privacy and synthetic data to help assure that no individual's data can be re-identified.

## Data Sharing Capacity

LinkedIn representatives shared that they might be interested in data sharing more frequently, but only if doing so was tied to positive social or economic impact, and the company could maintain user privacy. They said they want the team to be the right size for the requests they receive. Getting additional investment in data sharing would also require justification. Several factors decide if a data sharing project is justified, including formal criteria such as the project's feasibility, potential impact, additive effect (asking if existing data could accomplish the same thing), and thematic relevance (asking if data sharing contributes to equity, sustainability, or resiliency). Several things can inform how impact is measured, such as the number of downloads, views, or citations a research product using LinkedIn data receives, better-informed decisions regarding global economic development policy, or influencing the future of employer training and skill development. LinkedIn is also exploring using an API for data sharing so researchers and policymakers can pull aggregated indicators without needing direct staff support.

## Data Sharing and Privacy

There are several people in the company that help calibrate the right level of privacy safeguards and data granularity for sharing, including data scientists, economists, and policy experts. The data sharing team said they err on the side of caution, observing that there are generally two reasons they choose not to share data. First, if there are any privacy concerns within a data request that cannot be mitigated, and second, if the external partner is ill-equipped to understand the statistical limitations of the available data. Sometimes LinkedIn data can be complex, incomplete, or unsuited for the statistical methods the partner wants to use, and in those cases, LinkedIn does not share data with them.

## Costs

Data storage, computation, IT infrastructure, and legal support were all listed as ongoing costs for data sharing, but the biggest cost for the company is the staff time to manage the program. LinkedIn reiterated that the cost of staff time is why they have strong DLA policies that reduce the negotiation period when onboarding new partners. The high cost of bespoke data sharing requests motivates their focus on developing long-term solutions and automated data sharing techniques.

# RISKS AND BENEFITS

## Risks

LinkedIn identified legal, reputational, intellectual property, and privacy violations as potential risk areas when sharing data. The company minimizes risks by only sharing aggregated, anonymized datasets with trusted public benefit partners who have signed DLAs. It manages remaining risks as effectively as possible by being transparent with members about the risks and mitigation techniques when members consent to share their data. They noted that partners occasionally mischaracterized or misinterpreted LinkedIn data in draft research outputs. When this happens, LinkedIn has to go back and meet with the researchers to rectify the error before publication. Lastly, there can be perception risks related to data sharing for public benefit. It can be challenging to effectively convey the benefits of data sharing with the public when there is reasonable public mistrust of data-collecting institutions.

## Benefits

LinkedIn representatives said that the company's vision is to create economic opportunity for every professional in the world and that data sharing with external researchers and policy partners helps LinkedIn achieve that vision. Additionally, data sharing partnerships complement the data analysis done by its internal researchers. The representatives also conveyed that data sharing has led to unexpected ideas, creativity, and learning opportunities. For example, the IMF provided feedback about LinkedIn's skills data that gave the company insight into using their data in new ways.