

Data Sharing for Research Case Study: Khan Academy

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.

Data Sharing Type

Internal, Closed Trusted Partnerships

Organization and Partners

Company

Khan Academy is a US-based nonprofit organization founded in 2008 that operates a website and related applications providing online educational programming for students through instructional videos, online exercises, and instructional articles. Khan Academy has approximately 230 employees and reported revenues of over \$59 million for the fiscal year 2021.¹

Khan Academy and Formative Assessment Partners

Khan Academy representatives said they partnered with a third-party formative assessment provider to co-develop an educational tool based on student assessment scores. As part of their product-development partnership, they also developed a research partnership to understand how students' use of Khan Academy products affects student assessment scores across different demographics. The research project was co-developed by Khan Academy and its assessment partner and used a secure data warehouse to facilitate researcher access to shared, de-identified data.

Khan Academy and Standardized Test Partners

Another data-sharing model has been used, at a somewhat smaller scale, with standardized test provider partners. Khan Academy offers test preparation courses for different standardized tests. Providers of such tests wanted to understand the relationship between student use of Khan Academy test-prep materials and student

¹ Khan Academy. 2022. Khan Academy 2021-2022 Annual Report.

https://s3.amazonaws.com/KA-share/annualreport/Khan_AnnualReport-22_R5.pdf

scores on standardized tests. The sample for these studies is smaller than the potential population using the test-prep courses. This is because the testing partner only requested consent from test takers and not from all users; additionally, not all test takers provided consent. Once the testing body sent Khan Academy the user list and confirmation of consent, Khan Academy queried the relevant data and securely shared it with the testing body, which merged in test scores and de-identified the record to complete their analysis. For the publishing phase, Khan Academy staff reviewed the researcher's drafts and provided feedback.

Khan Academy and School District Partners

In 2017, Khan Academy started partnering with school districts to measure how students' use of Khan Academy affects their scores on a standardized state test. In this data-sharing model, the school district provides consent and direction for Khan Academy to collect student data on its platform and share the identified student usage data with the district securely. The district merges in student test scores and a subset of demographic data, de-identifying before securely sharing the full dataset with Khan Academy. Khan Academy then stores the data in a secure data warehouse, conducts the analyses, and shares the findings with the district. Khan Academy has worked with several school systems on research involving test scores from their districts and prefers long-term partnerships instead of one-time data requests. An example of published research from this type of partnership can be found in a company report titled [Use of Khan Academy and Mathematics Achievement](#).

Partnership Considerations

Company Data Sharing Team

Khan Academy's Efficacy & Research team comprises three full-time people supporting data-sharing partnerships as part of their overall responsibilities. This particular team's larger scope focuses on research into the efficacy of instructional techniques and often collaborates with external organizations or researchers. The team aims to develop

queries that can be reused across different partnerships' data requests, thus reducing repetitive work. They have created clear data dictionaries so that partners can accurately understand shared data, and they offer some consultative support for data-sharing partners.

Data Sharing

Khan Academy collects data for internal and external research and analysis through the operation and use of its platform. As part of its educational mission, Khan Academy is particularly interested in research to understand how the use of Khan Academy's learning platform affects mastery of the subject matter and student outcomes. In connection with providing its services, Khan Academy seeks out opportunities to partner with school districts and others to advance its research program. Districts that participate in research studies involving the use of test scores or other assessment data provide consent for assessment data to be shared with Khan Academy for efficacy research. Its research efforts generally focus on studies conducted in conjunction with its school district customers and other trusted partners. Khan Academy occasionally supports external research conducted by universities but generally declines third-party researcher requests, given the labor-intensive process required to curate fit-for-purpose data and negotiate data-sharing agreements (DSAs). Successful partnerships involve identifying a dedicated counterpart at the partner organization with whom to negotiate the DSA and associated expectations, requirements, and terms.

Risks and Benefits

Khan Academy's legal framework addresses data governance, including privacy and security. The company expects any external research using its data to be under the auspices of an ethical framework, such as an Institutional Review Board (IRB) authorization, and conducted using de-identified data sets. External research partners enter into DSAs with terms that vary depending on the use of data and type of study. DSAs typically address the research goals, roles of each party, secure processes for

preparing, transmitting, de-identifying, and storing the data, limits on the use of the data, and expectations regarding the publication or sharing of findings. In order to protect student data privacy, the company typically shares only de-identified data with research partners. An exception to this is data sharing with school districts. At the request and direction of the school district, Khan Academy will share the school district's own identified data, which the school district typically uses to merge with assessment data prior to de-identifying the research record. Moreover, Khan Academy applies data minimization principles and shares only a subset of all collected data. These partnerships publish research based on the shared data either jointly or independently. Regardless of who is publishing, disaggregated data is never shared in a publication.

Partnership Information

Khan Academy: <https://www.khanacademy.org/>

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.