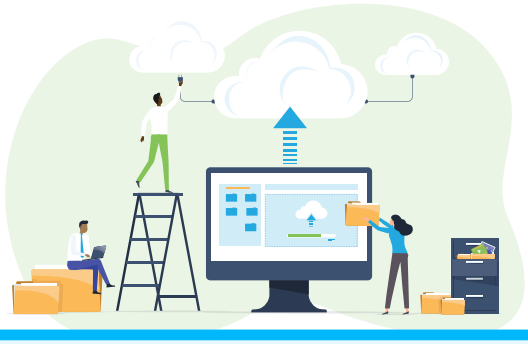
Data Sharing for Research
CASE STUDY
AIMS Collaboratory



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 Partnerships

ORGANIZATION AND PARTNERS

Organization

The AIMS Collaboratory is a "community of practice with the goal of accelerating research and development in learning strategies for algebra education, especially for Black/Latinx students and students in poverty."[1] The community is composed of fourteen partnerships, and each partnership — called a trio — has three distinct members: 1) researchers; 2) schools or school districts; and 3) curriculum developers, some of which are educational-technology providers. For example, one trio called "Rice Algebra Initiative for Success and Equity (RAISE)" is a "trio" partnership between: 1) Rice University (researchers); 2) the Houston Independent School District, and 3) OpenStax (curriculum developer). Where common issues arise among multiple trios, cross-organizational efforts are organized to address them. Funded by the Bill and Melinda Gates Foundation,[2] AIMS has a facilitation team staffed by representatives from menloEDU, WestEd, and the National Network of Education Research-Practice Partnerships that organizes and supports the trios.

Data Access

Data sharing among the trio members is an essential feature for the partnerships to be successful at improving math education, according to AIMS representatives, who added that while using a standard data sharing agreement for all partnerships would be desirable for efficiency, it has proved challenging. Most trios start with a standard data sharing agreement (DSA) template, and then individual members' legal teams usually add on specific terms and conditions to make the agreement appropriate for the trio. AIMS continues to adapt its standard DSA to make it useful for future trios with the goal of streamlining data sharing while protecting student data privacy and ensuring legal compliance.

Within each trio, data sharing can take different forms. Some trio members only share limited data with researchers, while others have fewer restrictions. In the trio between 1) the University of Toronto and the Abdul Latif Jameel Poverty Action Lab (primary researchers); 2) the Puerto Rico Department of Education (school district); and 3) Khan Academy (curriculum developer), some researchers are also employees of Khan Academy, which allows them greater access to data and more freedom to share within the trio. Increasingly, school districts are designating a single individual to manage external partnerships like these and support data sharing functions, with titles such as 'Director of Research' or 'Coordinator for Partnerships.'

While AIMS' mission is to "support education research," and data sharing is seen as essential to accomplish that under their model, the team cautioned that more data sharing isn't always better unless researchers employ principled and critical methods to avoid unanticipated and harmful consequences. For example, AIMS staff observed that some members of the data science community assume that if something shows up in data, it must be the truth, which is often not the case. They suggested that being able to validate data with members of the community can be an important quality check on data. They also remarked on a common tendency to collect all student data that *can* be collected rather than asking if it *should* be collected. These two behaviors, seeing data as objective truth and the proclivity to collect everything, are behaviors that AIMS has encountered across educational sectors, industries, and disciplines. AIMS staff reiterated that data should only be collected if it is directly tied to a research question and if there are adequate privacy and security protocols to keep data safe.

Ownership and Consent

According to experts from AIMS, questions about data ownership come up regularly in trios, sometimes with conflicting views among members that need to be addressed before the trio can move forward. For example, some members of a trio might argue that school districts legally own student data, but there are others who argue that the data belongs to the students and that the district is only a data steward.[3] AIMS representatives did not express formal positions on what constitutes appropriate data ownership for trios but stated that competing data ownership approaches from trio members were a common obstacle. AIMS experts also identified consent as a regular issue that needs to be addressed when sharing student data. A common practice in AIMS is for school districts to get consent from parents and assent from students who are under 18. However, if a researcher or a school district is seeking data about students from a third-party platform, then they might not get parental consent if the data is generated through general classroom practices, such as a teacher assigning students to use the platform as part of instruction. According to AIMS, If third-party platform-generated student data is not collected as part of directed class time, then their process around legal ownership and consent may be different.

Costs

Data sharing partnerships in AIMS have a mix of start-up, ongoing, and ad hoc costs. Most trio members have legal teams in-house or on retainer that work on DSAs. The more standardized the DSA, the more efficient and cost-effective it may be for AIMS. A common cost question the organization faces is how to compensate school districts for staff efforts to perform data cleaning and minimization outside of their normal work. Otherwise, data cleaning and minimization can be done by a member of the research team or an external contractor, both of whom usually have to travel to the schools and do the work on-site under supervision. Grants might build in a stipend to compensate districts for data cleaning and sharing. Ongoing costs include data storage, transfer, security, and maintenance.

Risks and Benefits

AIMS identified several risks associated with data sharing partnerships like theirs, with the most significant being unintentionally exposing student data. There are secondary risks that follow something like a data leak or breach, such as losing the trust of students, parents, and community partners, or in extreme cases, the loss of professional reputation, grant funding, or employment.[4] AIMS focused on two tenets identified as important for successful data sharing: first, thoughtful data privacy and security practices, and second, that there must be trust among all stakeholders. The parties recognize that all stakeholders have something to lose if sensitive data is handled improperly, especially students. Additionally, a single leak or breach of student data may incur a significant loss of trust and make future partnerships less likely, even if data privacy and security practices are improved after the fact.

The AIMS Collaboratory is built on the premise that data sharing can help to increase equity in math education. AIMS states that complex problems require a diversity of people and approaches to solve them, and their collaborations rely on ethical and privacy-oriented data sharing. The goal is to benefit students, the community and improve K-12 math education. Additionally, AIMS hopes to support rigorous research methods and improve secondary data analysis practices, both of which will help close critical gaps to make education serve all students.

PARTNERSHIP INFORMATION

AIMS Collaboratory: www.aimscollaboratory.org
J-PAL: www.povertyactionlab.org
University of Toronto: www.utoronto.ca
Khan Academy: www.khanacademy.org
Puerto Rico Department of Education: de.pr.gov

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.

1 www.aimscollaboratory.org
2 The Future of Privacy Forum is supported in part by the Bill and Melinda Gates Foundation.
3 For data disclosed by the school or disclosed by a 3rd party acting as a school official at the direction of a school, FERPA is clear: the school is the data controller.
4 A potential risk to participating schools is sharing data in a way that would subject the school to violating FERPA, which could prevent data from being shared by that district with anyone in the future for five years.