Postsecondary Data Infrastructure: What is Possible Today

AUTHOR: AMY O'HARA, GEORGETOWN UNIVERSITY

JUNE 2019

EXECUTIVE SUMMARY

Data sharing across government agencies allows consumers, policymakers, practitioners, and researchers to answer pressing questions. Creating a data infrastructure to enable this data sharing for higher education data is challenging, however, due to legal, privacy, technical, and perception issues. To overcome these challenges, postsecondary education can learn from other domains to permit secure, responsible data access and use. Working models from both the public sector and academia show how sensitive data from multiple sources can be linked and accessed for authorized uses.

This brief describes best practices in use today and the emerging technology that could further protect future data systems. To support decisions facing students, administrators, evaluators, and policymakers, a postsecondary infrastructure must support cycles of data discovery, request, access, analysis, review, and release. It must be cost-effective, secure, and efficient and, ideally, it will be highly automated, transparent, and adaptable. Other industries have successfully developed such infrastructures, and postsecondary education can learn from their experiences.

One such practice, the "Five Safes" framework, is an approach for controlling data access and use. The five safes are:

Safe projects: Building safe projects requires governance protocols to control project requests, review, and approval processes, and may require institutional board or ethics board review and approval.

Safe people: Data users should be screened and trained to become "safe people."

Safe settings: The most important control factors involve the data user's interface and environment. Many current practices regulate data inputs, computation, and outputs, creating safe settings and safe data.

Safe data: Aligned with "safe settings," data users should create "safe data". The practices for both impose restrictions on what an analyst can use, what an analyst can do, the analyst's computing environment, and the analyst's physical location.

Safe outputs: Protect the privacy of data subjects by reducing the risk of individuals being re-identified using the results in the data outputs.