

Student Data Privacy and Automatic Textbook Billing

by Tina Weyland
(she/her)
Reference & Instruction Librarian,
Rogue Community College
tweyland@roquecc.edu



Tina Weyland (she/her) is a Reference and Instruction Librarian at Rogue Community College and is a member of the college’s Textbook Affordability Group.

The textbook market in U.S. higher education is changing. In recent years, publishers have developed an automatic billing model, in which colleges and universities negotiate deals with publishers to provide ebooks and courseware to students, folding the cost into student fees. This model is commonly known as “inclusive access.” Because it offers students first-day access to course materials—important to student success—as well as some savings over full-priced standard textbooks, it is becoming popular with faculty and administrators. But textbook publishers are promoting these plans for another reason: The data they can collect with digital materials opens a lucrative new market, allowing them to diversify into analytics services.

Publishers’ textbook revenues have been hurt in recent years by the resale marketplace, Open Educational Resource (OER) adoptions, and lower enrollments. Shifting to automatically collected access code fees allows publishers to recoup some of those earnings, as “inclusive access” contracts provide a higher sell-through rate per course (Aspesi et al, p. 36). Students aren’t able to save money in traditional ways—for example, buying used books or older editions, renting, sharing, using library reserves, or selling books back—and publishers likely gain revenue overall. Some educators are pushing back against automatic billing, and not only for cost reasons. Students usually don’t retain access to materials after a course ends, and if they need to drop and take a course later, they will be charged again. The contracts can include high quotas for student purchasing and uncapped annual price increases (Vitez, 2020, p. 11).

But as important as these concerns are, the considerable student data these plans allow publishers to capture, as well as the lack of any real option for students who would prefer to protect their privacy, is just as troubling.



Students can only opt out of this data collection by opting out of the purchase. They are essentially a captive market.



Once students transition to digital materials it enables both their institutions and the commercial vendors to collect vast amounts of data on them: their physical location when they use them, their study habits, their learning profile, and granular knowledge on their performance. (Aspesi et al, p. 40)

Students can only opt out of this data collection by opting out of the purchase. They are essentially a captive market. While they can sometimes find another way to access the textbook, if they need to submit assignments or take quizzes through bundled courseware, opting out could mean trading a portion of their grade for data privacy. Students do forgo textbooks because of the expense (in a recent study, 63 percent of students had skipped buying for this reason [Nagle & Vitez, 2020]), but with courseware, opting out—for cost or privacy reasons—could mean accepting a lower grade before the course even begins.

Publishing companies are quickly moving toward services that allow them to collect data. Pearson, one of the largest college textbook publishers (Pearson, Cengage, and McGraw Hill together hold 80 percent of the market [Vitez, 2020, p. 1]), has announced it is moving to a “digital first” model in the U.S. (McKenzie, 2019b), and Cengage is aggressively marketing its digital library (Aspesi et al, p. 46). Pearson and Cengage have also developed mobile apps for their content which, while helpful for students without reliable access to a device other than their smartphone, also allow substantial data harvesting.

Institutions should be concerned about what these plans expose their students to—vulnerability to breaches, potential sale of data to third parties, or data being surrendered to governmental authorities, like local police or Immigration and Customs Enforcement (ICE), without judicial process. “The collection of massive amounts of data about faculty and students poses a significant legal and reputational risk for institutions, along with potential privacy and security threats for individuals” (Aspesi et al, p. 8).

This automatic billing model, sometimes presented as an equity solution to the high cost of commercial textbooks, may in fact amplify existing disparities. Publishers tout the convenience of getting materials directly to students; however, this is true only for students with reliable devices and internet connectivity. But, publishers’ data collection is its own equity concern. Learning analytics products promise improved student learning through data collection and proprietary algorithms. But algorithms carry the biases of their designers, and can reinforce existing disparities. In one example, COMPAS, an assessment program used to predict prisoners’ risk of reoffending, predicted that Black defendants would reoffend more often than they did, and that White defendants were less likely to reoffend than actually occurred (Angwin et al, 2016). And, when the Apple Card was launched by Goldman Sachs, it reportedly offered lower credit limits to women (Vigdor, 2019). The company insisted that a person’s gender was not one of its data inputs. But just as neighborhood can be a proxy for race, shopping history might be a proxy for gender. While companies may believe that their (proprietary, secret) algorithms are not considering prohibited characteristics, both the data and the algorithms reflect society.

It is likely that publishers’ products could profile students in similar ways. Could student performance data be sold to potential employers, with both the products and their baked-in algorithmic biases entirely hidden from students? While the data collected by publishers may be de-identified, “it could be matched with other third-party databases, leading some to worry that assigning access codes is tantamount to signing students up for surveillance” (Nagle & Vitez, 2020, p. 9).

Best Practices

Students included in automatic-billing plans should be clearly informed, optimally by their instructor, about the data collection allowed. Terms of use should be viewable by faculty and students before sign-up. Plans should be opt-in, but where contracts are opt-out students should receive repeated reminders (through more than one channel) of opt-out dates. Students should be able to meet all course requirements without opting in. And, institutions must consider the unintended consequences of using publishers' automatic billing plans for course materials. The Scholarly Publishing and Academic Resources Coalition (SPARC) suggests the following risk mitigation measures in negotiating contracts:

- The sale of data to third parties should be prohibited.
- Contracts should prohibit the surrender of students' data to authorities without judicial review.
- Institutions should maintain ownership of collected data.
- The procurement process should be open—no nondisclosure agreements.
- Contracts should require that algorithms using student data be “fully transparent” (Aspesi et al, p. 53).

Students don't know how much data is being collected about them. Surveyed about their knowledge of vendors' data collection, most students rated their understanding at the low end on a scale of 1-10 (“10 being fully aware and able to explain to a peer”), with a median rating of 2 (Nagle & Vitez, 2020, p. 3). Students do need to click through end-user license agreements to access their materials, but the agreements are long and complex, and clicking through is routine for most people. Most of us make decisions about which entities we find trustworthy, but for students who need an assigned textbook, it is not really a choice. Not agreeing to publishers' terms may mean not having what they need to be successful in a course.

Terms of use often include everything and the kitchen sink, as far as what companies are allowed to do.

Generally speaking, it is standard for terms of use for digital products to include a clause allowing the provider to change terms at any time without notice, possibly retroactively. Faced with increasing financial pressures and tempting opportunities to monetize data, could publishers resist? (Aspesi et al, p. 49)

Further, contract language may give publishers “the option to veto language in institutional communications that give students more context and information” (Vitez, 2020, p. 9). A recent study found that 42 percent of the 31 institutions reviewed “had signed at least one contract that appears to give a publisher final say on any public communications about the automatic billing program.”

Resources

Open Oregon’s Course Materials Adoption Best Practices—
checklists for administrators, faculty, and students

<https://tinyurl.com/mrx87nz9>

SPARC information page

www.inclusiveaccess.org

SPARC Automatic Textbook Billing Contract Library

<https://tinyurl.com/yc3pn2by>

Protecting Student Privacy While Using Online Educational Services:
Model Terms of Service, from the U.S. Department of Education’s
Privacy Technical Assistance Center

<https://tinyurl.com/35nh3s6a>

While federal law requires that publishers’ automatic billing plans allow students to opt out, this has its own equity implications for publisher analytics. Opt-in/opt-out frameworks are affected by consent bias, so any products built on the resulting data will be skewed. Those who opt out “may differ systematically, such that the conclusions or actions taken based on the data will unfairly bias one of the groups of students” (Brooks, 2021). A 2019 survey at the University of Michigan, for example, showed that women as a group may be more likely to opt in, with Black students as a group more often opting out (Li et al, 2021).

Most students trust their colleges and assume they have an ethic of care, but this ethic is compromised if decision makers are not considering potential harms. Administrators and instructors seem often to be choosing these plans while unaware about the data collection piece. Many institutions likely need to take a more comprehensive approach to data collection in general, with a wider set of stakeholders (faculty, librarians, staff, and students) included in decision-making. “Policies governing student data collection and use have lagged behind technological and cultural changes in higher education” (Brown & Klein, 2020, p. 4).

The Family Educational Rights and Privacy Act (FERPA) does not prevent data exploitation by publishers; constraints apply only to educational institutions, not vendors or other third parties. Passed in 1974, the act is commonly viewed as preventing institutional disclosure of student data. However, the law was originally motivated not by worries over improper disclosure, but by the impact of the data collection itself on students’ lives—concern about “secret gatekeepers, arbitrary categorizations, and bureaucratic errors that, unchecked, could become permanent liability” (Igo, 2018, p. 250). Lawmakers worried that “inaccurate information or biased judgements about students would linger ... creating a ‘records prison’ that follows students” (Brown & Klein, 2020, p. 5).

At issue was not so much whether a pupil would be documented in a variety of ways ... but whether that student’s record would be documented accurately and fairly, how long it would be maintained, who else would have access to it, and how the subject of that record would go about finding out what it contained (Igo, 2018, p. 250).

Protecting user privacy is one of librarians' core values. Most higher education institutions in Oregon have librarians working on textbook issues, and librarians should be advocating within their institutions for students' data privacy interests. Librarians have an important role to play in helping students, faculty, and administrators understand how this data is being collected and how it might be used. As Nicole Allen of the Scholarly Publishing and Academic Resources Coalition (SPARC) said of automatic billing plans, "Higher education owes it to students to grapple with the ethics of this new course content landscape" (McKenzie, 2019).

References

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Aspesi, C., Allen, N. S., Crow, R., Daugherty, S., Joseph, H., McArthur, J. T., & Shockey, N. (2019, March 29). *SPARC landscape analysis: The changing academic publishing industry—implications for academic institutions*. <https://doi.org/10.31229/osf.io/58yhb>
- Brooks, C. (2021, September 20). Privacy opt-out may lead to inequities. *Inside Higher Ed*. <https://tinyurl.com/yc7rmnpx>
- Brown, M., & Klein, C. (2020). *Whose data? Which rights? Whose power? A policy discourse analysis of student privacy policy documents*. Iowa State University Digital Repository. <https://dr.lib.iastate.edu/handle/20.500.12876/104715>
- Igo, S. E. (2018). *The known citizen*. Harvard University Press.
- Li, W., Sun, K., Schaub, F., & Brooks, C. (2021). Disparities in students' propensity to consent to learning analytics. *International Journal of Artificial Intelligence in Education*, 1–45.
- McKenzie, L. (2019a, April 26). Scrutiny of a financial relationship. *Inside Higher Ed*. <https://tinyurl.com/2p85nwpe>
- McKenzie, L. (2019b, July 16). Pearson's next chapter. *Inside Higher Ed*. <https://tinyurl.com/3hx86nkk>
- Nagle, C., & Vitez, K. (2020). *Fixing the broken textbook market* (2nd ed.). U.S. PIRG Education Fund.
- Vigdor, N. (2019, November 10). Apple Card investigated after gender discrimination complaints. *New York Times*. <https://tinyurl.com/mwcpu8cp>
- Vitez, K. (2020). *Automatic textbooks billing: An offer students can't refuse?* <https://tinyurl.com/2b737yjx>

