Section 1: Week 2: Organizational Data Management Problems

Nate Bachmeier

TIM-7020: Database and Business Intelligence

December 14th, 2019

North Central University

EdTech Data Management

Educational institutions need to maintain student rostering information, such as the class enrollments and the instructional hierarchy of their courses. In some instances, these institutions must associate their rosters with parent organizations, such as county and state districts. Specific aspects of these hierarchical structures are made available to various third-party providers that use these feeds to expose an ecosystem of personalized student learning, teacher dashboarding and classroom analytical tooling.

# Challenges of Data Sharing

Innovating within the Education Technology (EdTech) ecosystem is challenging because of non-uniform exchange protocols, insufficient information governance, and inefficient synchronization strategies. These issues complicate data management solutions and force businesses to design solutions that operate in a heterogeneous environment that cannot adequately address the needs of its audience.

## Use of Proprietary Protocols

Student enrollments are not static, and mechanisms need to exist for notifying partners of these changes. These notifications started life as proprietary unstructured messages that lacked a focus on interoperability. The Instructional Management Solutions (IMS), an international standards organization, has proposed OneRoster to standardize the communication protocol (IMS, 2019). However, these efforts need broader adoption by the tool ecosystem and more strict enforcement of data values. Many schools degrade the protocol to semi-structured messages because they lack sophisticated data quality processes (Herald, 2016). Herald provides an example that different systems might know the same student as Jim, Jimmy Smith, and James. Another challenge arises from the reuse of student numbers (primary key) that might be unique within a given school (local context) but overlap at the district level (global context).

## Insufficient Privacy and Security Controls

Parents, administrators, and legislatures have concerns around their privacy of student’s information (Regan & Jesse, 2019). Regan and Jesse state that federal regulators oversee sharing between organizations using powers granted under the Family Educational Rights and Privacy Act of 1974 (FERPA), Children’s Online Privacy Protection Act of 1998 (COPPA), and related laws. “The central tension in Edtech is between the need to protect student data privacy on the one hand, and Edtech companies’ ability to innovate on the other (Peterson, 2016, p. 962).” Peterson’s statement is missing the critical point that students have entrusted their data to their schools explicitly, and the EdTech implicitly. These additional protections necessitate an information governance policy that is natively part of the roster sync protocols, at both a coarse (e.g., entity filters) and fine-grained (e.g., attribute filters) level of control.

## Inefficient Synchronization Protocols

A third aspect of the problem is an inability to perform data integrity checks and efficiently remediate in real-time. Many third-party consumers handle this drift through a full periodic synchronization with the source of truth. However, this can be challenging to scale across large customer bases with millions of students. The delays also reduce the customer experience, as students cannot immediately enjoy the third-party products (ClassLink, 2018). Specific sources of truth providers expose proprietary protocols for streaming update deltas (Clever, 2019). These solutions introduce external system dependencies, such as Lightweight Directory Access Protocol (LDAP), that can be clunky for specific workloads.