Centralized vs. Distributed Educational Solution Architectures -
Data Confederacies compared to Data Unions

The Issue

There are real and growing privacy concerns about Local and Government Education Authorities ceding, or being forced to cede, control over exactly who can access their student's sensitive data and for what purpose. Educational solutions can be constructed, which address such data privacy concerns, without impacting their value to teachers, parents, administrators and others who use the saved data to more effectively meet the needs of individual students.

Many educational institutions (local, state/territory and federal agencies or even foundations) see the value of a single “grand Data Store” containing all the information they must gather and maintain. The reality is that it is often far more effective to architect the Data Store as just one “citizen” in a community of interoperating applications because, for security reasons, it may not contain all the data it needs to answer every question that it might possibly be asked! (Example: Give me the identities of all special education students in District X with more than two discipline incidents).

Ideally a Data Store is a Consumer (gathers data) and a Provider (gives it to analytics applications for analysis and reporting), but its access to data should be controlled the same way as any other application. In a “Data Union”, the Data Store gets unrestricted access to all the data and enforces who can access and change it. In a “Data Confederacy” the Data Store only gets access to the data that the local data administrators have determined it can acquire.

Compare and Contrast: Multi-tier Educational Data Sharing Architectures

<table>
<thead>
<tr>
<th>Architecture</th>
<th>“Data Union”</th>
<th>“Data Confederacy”</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>• A single central Data Store is given access to all (demographic, health, discipline, etc.) student data in a State or large District and identification data that ties that information to a specific student.</td>
<td>• Every data point has an “owner of record” (ex: an SIS/LMS data provider) responsible for its accuracy and relevance. “Master Data Management” (MDM) allows for distributed control.</td>
</tr>
<tr>
<td></td>
<td>• All requests to access Student-related data go to this one central source.</td>
<td>• Consumers wishing to access student-related data may be transparently redirected to the actual owner of that data.</td>
</tr>
<tr>
<td><strong>Data Security</strong></td>
<td>• The Providers of the data are not involved in restricting data access or enforcing data security.</td>
<td>• Data security policies are locally defined and enforced by local administrators before the data is send to Consumers.</td>
</tr>
<tr>
<td></td>
<td>• The “cloud” may contain student data from multiple states, each of which “shares” the same security policies (ex: a single representation of a student identifier).</td>
<td>• Consumers have access to the ID element it was given. Updates return to Consumers with associated analytics, resources, etc. may not be linked to the actual student record.</td>
</tr>
<tr>
<td><strong>Accountability</strong></td>
<td>• A centralized Data Store seldom guarantees more than “best effort” responsibility for ensuring the privacy of student data it has been entrusted with.</td>
<td>• Local Administrators traditionally have been held accountable for any security breaches of the data they control.</td>
</tr>
<tr>
<td><strong>Flexibility</strong></td>
<td>• Since the data store is the owner, it must contain identity information for all students.</td>
<td>• Data security policies may be imposed and enforced by the Data Store, the Data Owner (ex: SIS) or the responsibility for ensuring data privacy may be ceded to a Data Consumer such as a trusted Portal.</td>
</tr>
<tr>
<td><strong>Market Penetration</strong></td>
<td>• Successfully deployed wherever a single organization was in control of data consumers and providers, and a single security policy could be mandated (Banking, Higher Ed, etc.)</td>
<td>• Successfully deployed wherever data was being provided and utilized by multiple organizations and/or by multiple levels within the same organization (Healthcare, K-12, etc.).</td>
</tr>
</tbody>
</table>
Why Consider a Data Confederacy?

- Increased Security through local control of Data Sharing: Data owners maintain control of their data and must “opt in” to share it with other entities (application or organization). Note the marked contrast with a “Data Union” based around a single central Data Store that every application must share its data with.

  Local administrators of a Data Confederacy determine which applications (local and remote) can access or update sensitive data such as student discipline or health information, and exactly which parts of that data will be made externally visible. This level of local control over all data that is to be shared, directly addresses the increased public concerns surrounding student data privacy.

- Master Data Management and local accountability for Data Quality: Master Data Management (MDM) is about ensuring accountability. An MDM solution has an assigned application responsible for ensuring the accuracy and correctness of the value of each data element.

  In a Data Confederacy, that application is the “owner” of the element, and the one to which other applications go to access it or update its value in accordance with local business rules (ex: Student data elements are “owned” by the local Student/Management Information System (SIS/MIS)).

  In a Data Union, all applications go to a central Data Store to access or update data. This effectively makes the Data Store responsible for the correctness of all elements, even where they were supplied by another application, possibly one not even in the same level of the organization (ex: local student data elements contained in a regional/federal Longitudinal Data System (LDS)). While data element values may be transferred across organizational boundaries, the business rules on sharing and updating those elements never are, and so must be recreated and re-imposed. This duplication is a significant factor in introducing data inconsistencies and data security violations into a purely Data Union operational solution. For example, the list of applications that a school grants access to its identifiable sensitive student data might differ from the list of applications granted access to the exact same data when it is contained in an LDS.

- Trade-offs between Educational Web Services and an Educational Enterprise IT Backbone: Data Confederacies are generally implemented as an enterprise IT backbone behind a firewall, leveraging middleware such as an Enterprise Service Bus or Message Broker. Such an enterprise backbone can have many “owners” of various types of data, including SIS, Learning Management Systems, Library Systems and local Data Stores, each of which can freely access, update and monitor the data maintained by other owners. A set of Services provides a centralized administrative control over application authentication and authorization rights, Identity Management groups, error logs, diagnostics and service governance and maintenance. The fact that all components utilize common middleware allows huge scalability advances (ex: 1000 daily student attendance records can be packed into a single “create” request).

  Data Unions may be part of such an IT backbone, but they are most often deployed “in the cloud” as an isolated web service directly accessible over the internet, often without the benefits of a firewall or any “denial of service” protection mechanisms. They generally follow the RESTful Web Service design paradigm which is a lowest common denominator method of data exchange well supported by standard developer toolsets. Each such “Data Union” Web Service has an individually defined set of security policies and permissions, and is totally independent of any other data sources, other than those it accesses as a web service client.
Request-Response vs. Real-Time Events

The most common Message Exchange Pattern (MEP) is “Request-Response”. A client makes a request to access or change data held by the service and the results are immediately returned in the response. In general this is the only design paradigm supported by a Data Union solution featuring a Data Warehouse supported in the cloud.

An additional MEP is Publish-Subscribe which is used by client applications to monitor data and react to server-side data changes in real time. In Data Confederacies, multiple clients can “subscribe” to data changes from a variety of Data Owners, and be instantly informed whenever such changes occur. This allows several important functionality advances, including the ability for the Service to limit subscribers to a pre-approved list (critical for security) and to eliminate message latency delays by “notifying” the client when new changes are made, rather than waiting for the client to periodically poll for such changes.

Two Solution Paths

When constructed effectively, Confederacies are very flexible, and can add real value in addressing a variety of different use cases. There are two main topologies incorporating Data Confederacies, which are successfully deployed at many educational sites today:

- A Data Collector inserts itself into each Local Data Confederacy and requests and subscribes to change event notifications from the Student Data owner – and receives them in real time! The Local solution administrators still have the option to remove student identifying information before being passed to the Collector level subscriber, or to restrict collection of irrelevant data. The data being returned is filtered “raw” Student Data.

- A Data Collector reporting application publishes a “report manifest” which reaches subscribers representing each Local Data Confederacy, indicating the desired format of the report. A Local report collector gathers the data, calculates necessary aggregates, and sends back the report. The report data being returned in this configuration is usually aggregate rather than individual.

In many cases, a combination of a Data Confederacy and a Data Union (mixed mode) offers the most promise to solving the needs of a particular Educational Community. An example of such a Solution Profile might be:

- Each school operates as a Data Confederacy and supplies data to the LDS in real time with those elements it is “willing” to share. In this case, that might be all Student Data minus those elements that would allow specific identification of a student.

- At the Collector Level, the LDS functions as a Data Union, making its data available to data mining applications and supplying the results of analytics back to the schools in accordance with established security policies.

The Enabling of “Confederacies and Unions”

Effective Master Data Management is critical for the success for any data system on the ground or in development today. It must consist of business processes, policies and the technical tools/standards to successfully manage the data not only needed for day to day activities (ideally supporting student learning) but also the essential reporting required by another governing authority.

The SIF Association, and its unique global community of 3,200 schools, governments and marketplace providers, have been developing the technical standards to allow tools to be created that allow for both “Data Unions” and “Data Confederacies”. For over 15 years, this open community has been sharing the effective practices in the identification, management, movement
and usage of learning information and has established tools for developers to build, and end user to validate their current and future IT investments. It is THE only technical standard to enable true master data management at all levels of the educational ecosystem.

The recent release of the community’s Implementation Specification 3.0 is now taking that enabling to the next level. This new technical blueprint allows locally developed data models to be moved over various reference infrastructures for not only the “right data” but also in “the right and secure way”. For the global community this means that the government led projects in the United States (Common Education Data Standards), United Kingdom (Information Standards Board), and Australia (Student Baseline Profile) can all be used over the same “wire” that has been implemented and tested globally.

Summary

We all know that rapid technology change impacts all of our work. Layered on top of that is the expectation from our stakeholders on “why can’t I get that data now”, which is rapidly changing our current roles and work focus. The technology you need now may not be the same technology you need in the coming months and years. In our collective change management, it is critical that we think to the future and how we can optimize our data ecosystems to work together in a safe and secure manner using widely adopted effective practices.

There are viable alternatives to the pure “Data Union” architectural approach to collecting and providing access to sensitive student data. A mixture of “Data Union” and “Data Confederacy” architectures are often the optimal solution, allowing for broad data access, while still conforming to more locally defined data security policies. The key to either strategy is the utilization of community developed and open technical standards which have been in use for an extended period of time – like the SIF Association Implementation Specifications!