BACK ENDS AND FRONT ENDS

An Overview of Options for Data Load and Login in 3rd Party K-12 Applications

TWO SIMPLE PROBLEMS

1) How do we get educational data out of one system in order to populate another system?
(Back End)

2) How do we use a single identity for log in, authorization, and user information for these systems?
(Front End)

SESSION GROUND RULES

- This session is about best and worst practices.
    - We recognize that there is no perfect solution.
    - The best solution will vary depending on context.
    - But there are some solutions that are simply terrible.
- Feedback from sibling Districts is appreciated.
    - Your mileage may vary.
    - You might have more sophisticated solutions of which we are unaware.
    - Chime in at any time.
- Let us know if we need to make it more interesting.
    - No promises, but Johnnie can sing a little.

PROBLEM CHANGES DEPENDING UPON SCALE

- Is the unit of measurement a single school, a single district, or an entire state?

- How complex an organization? Multiple buildings, multiple departments, multiple schools within a larger school, team-teaching … ?

- How many students? 1,000; 5,000; 20,000; 40,000; 60,000; 100,000?

- How dynamic? Same students all year, different students every six weeks, different students every day?

- How granular is your data? In Florida it is very granular …

- Minimum Situation: Import teachers and students once.

- Maximum situation: Export all data on teachers, students, staff, courses, schedules, institutions at least once a day.

WHERE IS THE CANONICAL DATA?

* Local sites generally have monolithic Student Records, Human Resources, and Finance systems that are the entry point for data and must be correct.
    * Data often copied into these systems from specialized systems (testing, employment screening, etc.).
        * These specialized systems are canonical because they generate the data they hold.
        * But often it is copied into another canonical system that holds a superset of the data, or generates more fundamental data. The copy therefore becomes functionally canonical.
* State data requirements are canonical for ultimate format and data type.
* State systems in the U.S. usually mean State systems, but worldwide the ultimate educational authority is different in every sovereign government.
* Best approach is to project rather than copy canonical data. But no system trusts another system to hold good data.
* Second-best approach is to copy as little as possible: Only two systems should be regarded as canonical for any data set (origin and central) and copy data as little as possible.

DATA STORES

- SQL-Oriented Relational Databases have taken over the world.

    - Even databases that are not natively SQL tend to have a SQL translation layer.

- VSAM and other proprietary formats are on the decline.

- But every now and then something like the 4GL Progress Database used by Skyward turns up.

    - Let's not even consider DataTel and PeopleSoft.

- XML data stores can be thought of as a successor to textual formats.

DATA REPRESENTATIONS

- The protocol/language used to access the data
  - is different from the data representation as structure,
    - which in turn is different from how the data is represented on disk and in memory.
      - All of these things matter in regards to data storage and retrieval — what is easy and what is possible.
- How direct is your access to the data?
- Reads are easy, writes are hard.

ODOM'S COROLLARY
TO THE TURING MACHINE

- Turing said that any programming language meeting his requirement for a Turing machine ("Turing-complete") was functionally equivalent to any other Turing-complete language.

- The Odom Corollary is that any means of storing data so that it can be retrieved programmatically is functionally equivalent to any other means of storage.

- There are "tar pits" in both cases.

- And to use Zawinsky's Second Law:
  "To the man with only a hammer, a screw looks like a malformed nail."

- Are we using SQL in places where it is not efficient just because N-Tier applications are easy to create?

FORMATS FOR DATA TRANSIT
I: TEXT

- Good ol' CSV (or other delimited)
    - Pray that the delimited does not show up in input.
- Good Even Ol'er Fixed Width
- XML
    - Don't forget your DTD
    - And validation.
    - Data types matter!
- What character set are we using?
    - UTF-8 probably best choice, but MS likes UTF-16 in programming. http://www.joelonsoftware.com/articles/Unicode.html
- Line Endings: Windows Or Unix
- Don't forget to escape.
- Get a Good Text Editor!
    - (Notepad++ or TextPad on Windows, BBEdit / TextWrangler or TextMate on OS X. On Linux Gedit, Nano/Pico, Vi EMacs …)

FORMATS FOR DATA TRANSIT II: SQL-TO-SQL

- Direct Query (select ... then update)

- ODBC And JDBC

- And complex data-to-data systems sold by Enterprise vendors.

- Data types will be problematic going between platforms.

- Always test for data existence and, hopefully, sanity before doing updates between databases.

FORMATS FOR DATA TRANSIT
III: PROPRIETARY VENDOR SOLUTIONS

- LINQ is recommended when doing .NET programming.

- SSIS packages on SQL Server for a variety of different data transfers.

- Other database vendors have other options.

- Lots of companies have special data loader programs:

    - Google Apps Directory Sync (GADS) and MS-Office 365 DirSync

    - School Messenger

    - Successmaker/Succesnet

    - Eduphoria

    - Blackboard ConnectEd

FORMATS FOR DATA TRANSIT
IV: BIG PICTURE SOLUTIONS

- Scribe - Sophisticated system for extracting data without writing SQL.

- Clever - Business based on taking the pain out of loads and logins for academia.

- IMS Global - A worldwide consortium of educational institutions and vendors that offer a variety of standards specifically formulated for data interchange between educational systems.

- SIF - The folks at IMS Global intend to learn the lesson of this standard …

- More on Clever and IMS Global in a few slides.

DATA TYPES MATTER

- Databases are strongly typed.

- Programmers tend to choose common data types and switch when their data violates a constraint of those common types.

- But spreadsheets have different common data types.

- And languages vary: PHP data types are generally C-strings. .NET types are quite varied. Javascript data types are a story in themselves.

- Always substitute empty data for null values.

- Proper escaping.

PROTOCOLS FOR TRANSFER

- FTP - Still around because it is the least common denominator.

- FTPS - Simply standard FTP with SSL/TLS bolted onto the front.

- SFTP - SSH as a transfer protocol. Great for Linux systems because it is encrypted, uses existing server security, and grants easy access to full file system paths.

- HTTP/HTTPS - Usually requires another layer of abstraction on top such as SOAP/WSDL or REST. Very popular because of the accommodations made for it due to other services, and because cloud service vendors have developed their services almost entirely in terms of web technologies.

Vendor Black Boxes

- How often can we load data, and how does this affect system uptime?

- Is data altered or replaced when we load changes?

- How easy is it to erase or reload an entire data set?

- What is the architecture of the vendor solution (platform, data store, programming, abstractions, server configuration)?

  - Based upon this we can make assumptions about loads, reliability, capability, etc.

  - Is it entirely cloud, entirely local, or a combination thereof?

- To what degree is collaboration with the vendor's engineering staff possible?

  - The first year we pay for the product. Every year thereafter we are amortizing their development costs, so it is important to get our money's worth in new development that directly meets our specific needs.

  - Don't be afraid to be aggressive with contracts, especially early.

TO WHAT EXTENT CAN YOU ALTER VENDOR DATA STORES?

- For databases: Are new views and indexes allowed?

  - What about tables and fields?

  - It is likely that CTEs, temporary tables, and other trickery will need to be deployed for queries if the vendor is suspicious of alterations.

  - Can always replicate or create a second database loaded from the first.

- For LDAP: are schema extensions supported / encouraged?

- For Servers: Can your code run along side theirs?

  - Are data exports possible using arbitrary tools or only a blessed export system?

MANATEE INTEGRATION LIST

- Collections

- Go Math

- FootSteps2Brilliance

- iStation

- National Geographic

- Pearson SuccessNet

- SuccessMaker

- Plato

- School City

- Study Island

- SunBay

- Taleo

ESCAMBIA INTEGRATION LIST

- OneSource
- Write Score
- Transfinder
- Skyward
- School Messenger
- RTIB
- Renaissance Place
- Read 180
- PEER
- Medicaid
- ThinkCentral
- HRW (Harcourt etc. Textbooks)
- Health Department
- FCATstar
- FastMath
- ESE
- Discovery Education
- Destiny

A FEW EXAMPLES

- Going off-slide …

CLEVER

- The promise of Clever is that they will handle most of the front and back end stuff.

- Send them one master file for data loads.

- You have one login to Clever, which can be a variety of methods.
    - Escambia uses SAML.

- They take care of piping data and logins to disparate vendor products.

- Thereafter access to products (free and licensed) can be done via course, class, school, District, etc.

- Vendors, not educators, pay for Clever's costs.

- Interface design is very important to Clever's business model.
    - Very cloudy.

CLEVER USAGE

- Manatee:
    - Achieve3000
    - Algebra Nation
    - Accelerated Reader
    - Discovery Education
    - iReady
- Escambia:
    - Algebra Nation
    - Khan Academy

IMS GLOBAL

- "The world's leading educational technology advancement and technical standards setting collaborative."

- Consists of institutions, vendors, and governmental authorities working on common standards for specific educational cross-realm challenges.

  - K-12 and Higher Ed

- Standards began appearing around 2005.

LEARNING INFORMATION SERVICES (LIS) AND LEARNING OBJECTS

- Learning Object: A collection of content, practice, and assessment content designed to to address a single learning objective.

- Learning Information Service (LIS): "The definition of how systems manage the exchange of information that describes people, groups, memberships, courses and outcomes within the context of learning."

- LIS has 6 core services:

  - Bulk Data Exchange Management

  - Course Management Service

  - Group Management

  - Membership Management

  - Outcomes Management Service

  - Person Management

- Binds via WSDL/XSD or LDAP.

- There is also a Core Profile that must be implemented, all other functionality is optional.

SCHEMAS AND METADATA

* Dublin Core: 1995 from Dublin Ohio workshop. Fifteen properties for resource description.

* Learning Object Management (LOM): IEEE standard for metadata to facilitate search, evaluation, acquisition, and use of learning objects.

* schema.org: Machine-readable schemas for searching supported by the major search engine providers and given free usage for educational vendors and institutions.

LEARNING REGISTRY

- Joint project of U.S. Dept of Education and Dept. of Defense.

- Means of providing indexes of learning objects and other curricular items to systems that make use of them.

- Chaotic at present because implementation is being hashed out.

  - Standard search-engine model?

  - RDFa, Microformats, or Microdata?

  - Push or Pull?

LEARNING TOOLS INTEROPERABILITY (LTI)

- Container standard for sharing a very abstracted set of services across educational platforms.

    - Bi-directional with the ability to launch and share content from multiple platforms into a single interface.

- Version 2.0 released January 26, 2014.

- Version 1.0 actually subset called Basic LTI (bLTI) that exists to launch tools and content from within an LMS.

    - bLTI can also be used as ad-hoc SSO.

- RESTFUL API using XML as data format.

- Consumer (LMS) and Provider (Curriculum Content or Testing) Model

- Parties inform each other via product profiles.

- Consumer->Provider messaging via browser POST.

- Provider->Consumer messaging via REST directly between parties using OAuth.

- Authentication and Authorization passed via standard, but underlying mechanisms are pre-existing and specific to each party.

QUESTION TEST INTEROPERABILITY (QTI)

- ❖ Format to provide banks of assessment items, full tests, and result reporting (bi-directional).

- ❖ Provides modeling to accommodate most types of assessment (simple questions, essays, multiple choice, etc.) and contain them via XML.

- ❖ Works in conjunction with APIP: Accessible Portable Item Protocol

COMMON CARTRIDGE

- Packing format for importing and exporting content to an LMS.

  - Largely uses HTML for glue.

- Designed to replace SCORM.

- Supports LTI, QTI, and Metadata (LOM, Dublin, etc.)

- Provides the following pieces:

  - Rich content

  - Integrated assessments

  - Discussion forums

  - Metadata – both descriptive metadata for cartridges and roles-based for resources

  - Authorization for protected content

- Future Plans:

  - Web/enterprise integration with third-party tools and services

COURSE PLANNING AND SCHEDULING (CPS)

- Application profile for LIS.

  - Bulk Data Exchange Management Service

  - Membership Management Service

  - Course Section Service

  - Section Association Service

- Defined via UML to support WSDL and REST.

- Uses a central Planning And Scheduling (PAS) system that might be moderated by a Student Information System (SIS)

INTERACTIVE WHITEBOARD/COMMON FILE FORMAT (IWB/CFF)

* File format for shared whiteboard with ability to move objects across a whiteboard space.

* .iwb zip file containing contents.xml and other data.

* Makes heavy use of SVG.

* Supports many other media types.

ESCAMBIA'S HIERARCHY OF LOGINS

1) Full LTI from LMS per IMS Global

2) SAML

3) LDAP

4) District Username and Other Password from Data Load

5) Vendor Username and Other Password from Data Load

MANATEE'S HIERARCHY OF LOGINS

1) SAML

2) LDAP/ADFS

3) ADFS

4) District Username and other password

5) Vendor Username and other password

6) District and Vendor username and password are mitigated by the use of Stoneware's WebNet product.

LOGIN ISSUES

- Most applications keep their own user stores.

- Only trust District to provide authentication based upon username / password.

- User data is copied rather than projected.

- More sophisticated forms of authentication, such as SAML, keep most user information (including passwords) away from vendors.

- Federation with third parties such as Google and FaceBook is an ongoing issue, but attractive mainly due to convenience.

  - These still rely upon the third party only for authentication and passing of an identifier.

  - Cross-organization initiatives are underway throughout the world to create federated groups that are non-commercial.

CONCLUSION

- This is a space where higher vendor expectations and a willingness to follow through on those expectations can be highly beneficial to educational institutions.

- Loads and logins both moving targets.

- Questions?