

## March 2008    Features

### Mix Master

by John K. Waters

*The Schools Interoperability Framework (SIF) has emerged as a cornerstone of K-12 data warehousing, enabling once-isolated information to be shared among diverse systems.*



The **Champaign Community School District** Unit 4, based in East Central Illinois, is building a new database management system (DBMS). The district plans to first assemble a centralized data warehouse to store student information currently isolated in a range of custom-built software systems scattered around the district's 17 schools—everything from attendance management applications to transportation systems and food-service programs. Once these information "silos" are integrated, the district intends to provide a web-based data access system that offers an especially high level of transparency for administrators,

teachers, parents, and students.

It's a big project, and a tricky one. Data integration is one of the single most challenging tasks any district can face. Fortunately for Champaign's Unit 4, and the many other school districts throughout the country with data scattered in disparate systems, an open specification known as the Schools Interoperability Framework is mitigating that challenge. For roughly 10 years, SIF has been providing an evolving set of rules and definitions for sharing data among applications to a growing community of K-12 software vendors.

"The SIF specification is critical to this project," says Michael Harden, Unit 4's data analysis manager. "It pulls all the different pieces together so that our district's homegrown databases and proprietary technologies can talk to each other."

First proposed in 1997, SIF is a set of open software and data specs that describes how information can be exchanged among applications in a K-12 setting. It's based on the Extensible Markup Language (XML) and service-oriented architecture (SOA). XML is a platform-neutral, generalpurpose markup language; SOA refers to an architectural approach that loosely couples software systems to provide a set of linked, repeatable business tasks, or "services."

The first SIF-based products began appearing in 2000. At that time, SIF was competing for industry mindshare with another standard: the Electronic Data Interchange, which is a set of

information-structuring standards for electronically exchanging data. EDI was used mainly by corporations and government agencies, and it's still widely used for e-commerce. But in K-12 environments, SIF has emerged as the dominant data integration standard.

"Data isolation is a serious problem, wherever you find it," Harden observes. "If the data is difficult to access and share, it fosters problems like data redundancy, problems with data integrity, and inefficient or incomplete reporting. My name might appear as 'Michael Harden' in one DB, and 'M. Harden' or 'Mike Harden' in another. You can't make good decisions with incomplete or inaccurate information."

The world is choking on data, says Gartner analyst Janelle Hill. Her firm reports that a typical enterprise stored 10 times more data in 2007 than it did in 2000. A lot of that data is simply trapped in the systems that created it. Data integration connects these burgeoning islands of information. According to Hill, the process involves extracting data from a source-usually a database, but it could be files, web services, and even emails- transforming it with "joins," "lookups," or calculations, and then loading the transformed data onto target systems.

### **Interoperability:**

The ability of two or more systems to exchange information and to use the information that has been exchanged.

Two main components make up the SIF architecture: software servers called *Zone Integration Servers*, and message gateways called *agents*. The servers provide messaging among the agents within a network known as a zone. A zone could comprise a single school building, a group of schools, an entire district, and even an entire state. Essentially, a Zone Integration Server is the message bus, whose primary job is to guarantee delivery of messages. It also provides a channel for transport and access control to data. Agents are extensions of the individual applications that serve as intermediaries between the applications and the Zone Integration Servers. Agents move messages to and from the message queue, translate between the SIF data model and an application's native data model, and then express that translation in XML as SIF Data Objects. In other words, the Zone Integration Servers act as the data integration brokers among applications that support the SIF spec.

"It's about comparing apples to apples," says Larry Fruth, executive director of the Schools Interoperability Framework Association (SIFA). "It's about getting everyone not so much on the same page as access to it."

Fruth's organization is the nonprofit, independent standards body that advocates for, and maintains, the SIF spec. SIFA emerged from a project launched in 1997 under the auspices of the Washington, DC-based Software & Information Industry Association. The spec's first big backer was Microsoft, which today describes SIF as "an industry-supported technical blueprint for K-12 software." Today, SIFA boasts a membership exceeding 1,100, Fruth says. Though it started as a

vendor-dominated organization, 75 to 80 percent of its members are now end users. The group's membership roster includes schools, states, K-12 software vendors, systems integrators, and others interested in helping advance interoperability in the K-12 space.

"As schools utilize technology more comprehensively for teaching, learning, and administration, the need for this kind of interoperability becomes paramount," Fruth explains. "With the increased accountability and expectations in schools today, linking accurate, interoperable educational data to the right learning resources for individuals is the holy grail for using technology for real and measured improvements in learning."

Fruth says that the push for interoperability doesn't come easily to a school system in which things are done in isolation. "Education tends to be made up of a lot of islands, generally. At the classroom level, a teacher never really has to interact with an adult. They close their doors and they're on an island. Those teachers go on to become administrators, and those administrators move on to the state department, and eventually, they get hired by the Fed. So it's a model that everyone is used to, and it has become endemic across education."

"With the increased accountability and expectations in schools today, linking accurate, interoperable educational data to the right learning resources for individuals is the holy grail for using technology for real and measured improvements in learning."

*-Larry Fruth, Schools Interoperability Framework Association*

Keep in mind that SIFA produces no products itself, but rather coordinates member contributions to an open specification, which software vendors use to guide them in the development of their commercial offerings. SIFA supports and promotes the specification, certifies products, hosts a website, and provides training and tools.

The first official SIF specification, version 1.1, was unveiled in February 2003. Since then, SIFA has approved three upgrades. The latest upgrade, version 2.1, was released last September.

"The two big changes in the 2.1 version are in the student records exchange, and teaching and learning, all under this umbrella of greater out-of-the-box interoperability," Fruth explains. "Out-of-the-box interoperability is our overarching goal with this release. It's one of those goals we're all shooting for in high tech, but one we rarely get to. But some of the features and functionality in this release get us a lot closer."

In February of this year, SIFA unveiled the Zone Integration Server 2.0r1 product standard, which makes it possible—for the first time—for vendors to test the mandatory features of a Zone Integration Server as defined in the SIF Implementation Specification, and to document support for any optional/ conditional features in the associated conformance statement. The new standard will contribute to greater out-of-the-box interoperability in SIF implementations, says Mark Reichert, SIFA's chief technology officer.

One of the reasons SIF has survived and thrived for so long is that it holds the support of an active and involved community, or ecosystem. It's one of the keys to the success of any open specification. The SIF ecosystem consists of individuals and organizations investing time and resources to refine and improve the spec—essentially, the SIFA membership. In its roughly 10-year stewardship of the SIF spec, SIFA has overseen the development of an expanding set of technologies and services, including among others a vendor registry, a group of online tools, a readiness assessment program, and certification services for SIF-based products. There are currently more than 86 SIF-certified products on the market.

Within SIFA, there's also a subcommunity called EdSTART (Educational Software and Technology Application Registry Tool). The EdSTART community focuses on, as the SIFA website puts it, "providing end users and vendors with accurate and up-to-date information on data initiatives, preK-12 software tools, and procurement opportunities for true 'solution sharing.'" The EdSTART community includes SIFA members and non-members. It offers listings of requests for proposals, school software vendors, and vendors' products.

Another sign that SIF has come of age: The association is accelerating its version-release timeline. "We've actually changed our processes because of the end users' demands," Fruth says. "They want a spec every six months, so we're shifting from our old schedule of releasing updates every year and a half to two years. Now we're turning out new versions every six months. That doesn't mean that someone has to build to that new version, but we want to be more receptive to what the end users need."

The SIF data model is large and expanding, Fruth says. The recent addition of organizational profiles to that model allows the set of data elements to be virtually as large as a school needs.

How widely implemented among American K-12 school districts is SIF today? Getting accurate statistics about the proliferation of the spec is almost as challenging as the data integration itself, Fruth quips, because vendors are notoriously closedmouthed when it comes to sharing their own data. "If it gives them a competitive advantage, they don't want to talk about it," he says.

However, according to SIFA's 2006-2007 activities report, district-level implementations of SIF-based solutions now number in the thousands. As of the publication of that report, five states are implementing the spec statewide. "We can say with confidence that we now have more than 9 million students and teachers utilizing some information coming from a SIF-enabled application," Fruth adds.

In fact, notes Dave Moravec, business development manager at Integrity Technology Solutions, SIF-based solutions have become what he calls "me-too" products.

"School districts need a certain comfort level," Moravec says. "They don't want to be a beta deployment of anything. Early adopters of SIF were really on the cutting edge. But now, because the SIF spec has grown into the SIF 2.0 standard, because more school districts are adopting it,

and because it's becoming a standard across states, it has increased awareness. The more school districts that implement it, the more that say, 'If they can do it, then I guess we can too.'"

Bloomington, IL-based Integrity was selected by Champaign's Unit 4 to develop its new database management system. The company is a well-known provider of custom application development and integration solutions for schools. It's also a long-time member of SIFA (Moravec is the chairman of the membership and marketing committee) and offers a range of commercial, SIF-based products and support services. It is, in fact, a nationally recognized developer of SIF agents for K-12 software vendors.

"Because we build the agents that connect other people's software, we're often seen as a third-party solutions provider," Moravec says. "We think of ourselves as Switzerland, because we're decidedly vendor-neutral."

According to one industry analyst, the world is choking on data. A typical enterprise stored 10 times more data in 2007 than in 2000.

Among other reasons, Moravec likes the SIF methodology because it exposes business processes—those procedures or activities that support the objectives of a company or school. That's one of the great things about SIF, he says. "And I think that's going to be more important in the schools in the future. In the private sector, business processes have been well defined over the years. Now we're seeing an increase in the number of administrators entering the school systems from corporations, and those managers are looking at processes."

This is Unit 4's first SIF implementation, so Integrity will have to implement the SIF infrastructure as part of the first phase of the project. That implementation will require the district to define its own business processes, as SIFA puts it, "for and around data." Describing its processes will help Unit 4 to define so-called data-quality best practices for the district, and enforce the implementation of those practices as they are automated by Integrity's solution.

That said, Fruth hastens to add that SIF is not a business specification being used in education, but rather a data integration spec designed specifically to meet the needs of schools, districts, and their governing agencies.

SIF also allows Integrity to bring best-of-breed technologies to this project, Moravec says, including components from Microsoft, Edustructures, and technology integration services provider CPSI. The result will be a complete data warehouse and business intelligence solution designed to help Unit 4 improve its teaching, learning, and administrative processes.

"This project is definitely a collaborative effort between our staff and the staff at Champaign," says Moravec. "They want to be able to own the solution and not be locked in with a vendor. Using the Microsoft SQL platform and the reporting services associated with it, we will be able to maintain that neutrality. And they'll own the tools after we walk away."

One of the reasons Champaign's Unit 4 chose Integrity, Harden says, is the amount of customization the district will require to connect the disparate data in its systems. "We have a lot of homegrown systems," he says, "and Integrity understood that. They offered the best customization package, which, given our particular reporting challenge, made it pretty clear which vendor we should go with."

## links

- **CPSI**
- **Edustructures**
- **Gartner**
- **Integrity Technology Solutions**
- **Microsoft**
- **Schools Interoperability Framework Association**

For the past several years, Champaign Unit 4 schools have been operating under a voluntary consent decree that requires them to eliminate unwarranted disparities between white and minority students in academic achievement, discipline, assignment to special education, and participation in gifted classes, among other things. According to Roger Grinnip, the district's director of technology, the driving force behind Unit 4's data warehouse project is the exacting data reporting standards required by that consent decree.

"The ability for students, parents, and educators to have real-time access to accurate data, in order to make data-driven decisions, is critical to our efforts," Grinnip said when the project was announced. "We are confident that SIF is going to enable Unit 4 to streamline a lot of its processes and greatly reduce any possible errors and redundancy in our data."

That consent decree was good news for Harden, whose position was created because of it. And in the end, he believes, it will also be good for his district.

"I guess that's the bad news and the good news," Harden says, noting that Champaign's Unit 4 expects to have the first phase of its DBMS project—the data warehouse—completed by the beginning of the 2009 school year. The interface implementation will follow later that year.

"Because of the consent decree, we really have to make sure that our data is consistent, that our numbers are correct, and that we provide as much access to real-time information as possible to improve upon any type of disparities that might exist. But in the end, how we got there probably isn't as important as *that* we got there.

"The changes we're making to our systems will allow us to maintain the consistency and integrity of our data, and it will allow parents, students, teachers, and administrators to have the information they need, in real time, to make data-driven decisions. And that's a very good thing."

*John K. Waters is a freelance writer based in Mountain View, CA, and a regular contributor to T.H.E. Journal.*