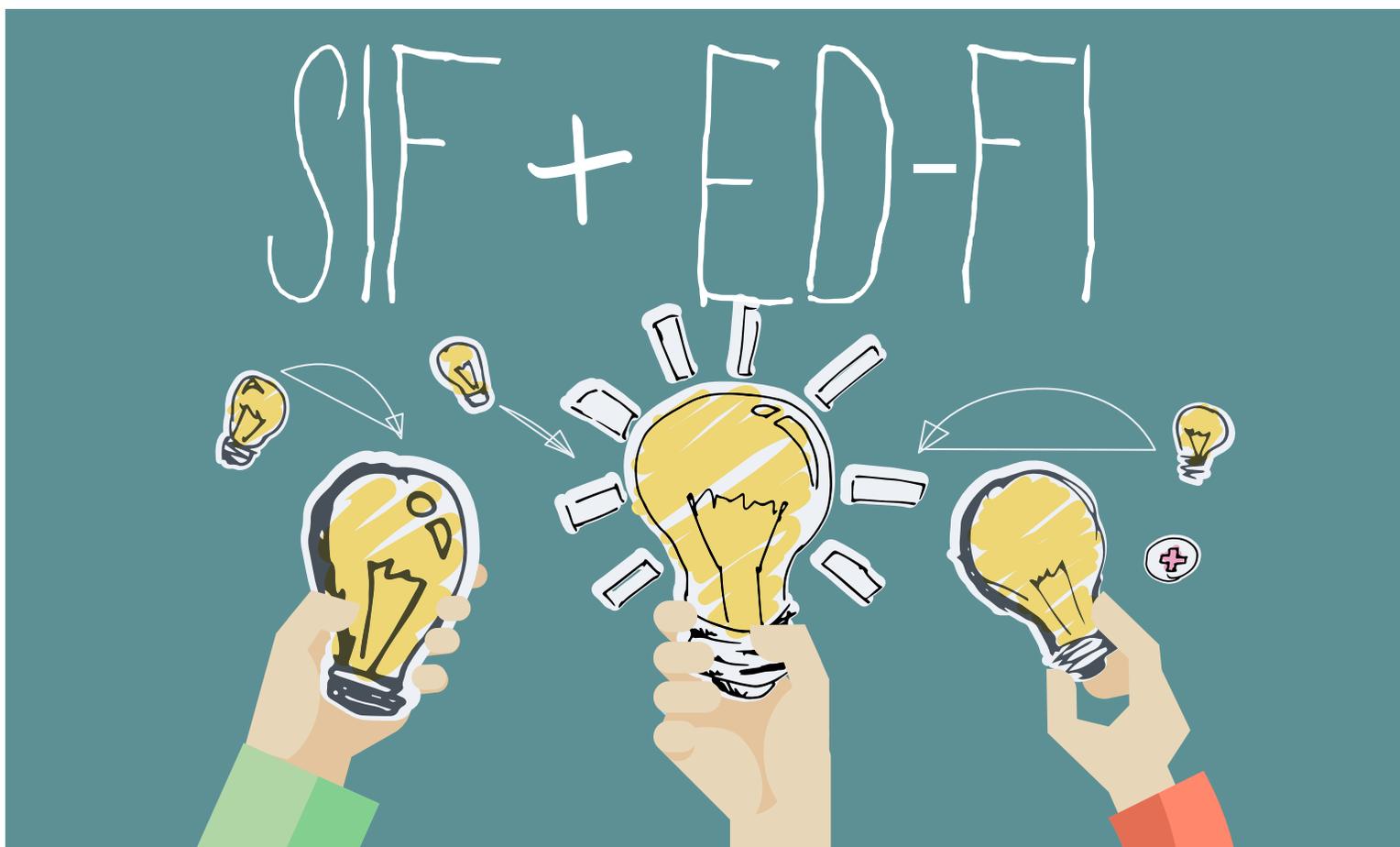# SIF and Ed-FI
Efficiently collect data with SIF while using Ed-Fi dashboards.

While implementing Ed-Fi in a state environment is very popular right now, it is not always the easiest activity to accomplish. Even with their new API, there are issues that will be encountered on a daily basis without the use of a standardized data collection process from the LEAs to the SEA. In fact, while SIF-based implementations can take as little as a year to accomplish, it is safe to say that an Ed-Fi implementation will take no less than two to three years to complete. That is assuming that the districts and their SIS vendors can effectively accomplish the data load process.

Some common issues encountered when implementing Ed-Fi:

1

The Ed-Fi ODS physical data model is highly normalized - 3rd normal form to be exact. This means that the data must be full, accurate, and meet all of the Ed-Fi code sets upon entry into the ODS. The level of data accuracy is extremely difficult to achieve in an educational environment due to the business practices in use at LEA's.

A data collection ODS that can capture data in real-time or near-real-time needs to be designed for data that is in flux, so it cannot be highly normalized. The ODS should also be designed to aid in data integration from multiple sources so that data validation can occur. SIF data collections can easily use an ODS to collect data. From here the data can be standardized, validated and moved into highly normalized data stores.

2

Ed-Fi uses natural keys as the primary keys. This causes a problem with educational data, as the natural keys cannot be considered permanent - for example, correcting state assigned IDs or birth dates. The natural way that educational data is entered does not lend itself easily to a system using natural keys as the primary keys.

SIF uses RefID's to track data. This allows for any data in the student's record to be changed without issues in the record. With CEDS and SIF, tables are linked by UIDs for statewide reference of students. With Ed-Fi, data is easily orphaned when critical fields in records change. With SIF, there is a RefID tracked internally and provided by the data's start of authority.

3

In an Ed-Fi data collection environment, the data update is focused solely on the needs of the Ed-Fi ODS. Nearly every defined element is mandatory, and some values must be provided within the district data for mandatory elements, even if the data is incorrect.

With a SIF data collection, the missing fields can be requested back from the source data and the missing data can be filled in when it becomes available. When using SIF and Ed-Fi, the SIF ODS can be used to collect and maintain complete records that can then be passed on to the Ed-Fi dashboards.

4

Even with the data being correct in the Ed-Fi ODS, the data may not pass the validation rules for the dashboards. Because Ed-Fi uses natural keys, if one key is off the entire record will be inaccurate or not loaded. This means that data that does not pass validation rules will be missing or incorrect in the dashboards.

SIF data collections are required to track students properly across locations and through time. With a SIF ODS, records can be validated and ensured of completeness and accuracy for loading into the Ed-Fi dashboard system.

The Solution

A SIF Based Data Collection System Integrated with the Ed-Fi Dashboards

ONE	TWO	THREE	FOUR
REAL TIME DATA	MULTI-AGENCIES	ONE COLLECTION – MULTIPLE USES	SIF STANDARDS

The SIF-based data collection can be accomplished in real-time making the data available in the Ed-Fi dashboards usable by all of the stakeholders. When data is missing due to incompleteness in an Ed-Fi system, it takes a human to identify the issues and make corrections. Thus, data is rarely in real-time. In fact, many times data is only truly updated in the dashboards every 3 to 4 weeks due to the time it takes to correct the data. With a SIF-based data collection, very little data will be missing and automation can occur with validations and updates.

The SIF-based data collection is a standardized full data set that can be used by systems OTHER than Ed-Fi. In a P-20 system and a system that includes workforce data, the Ed-Fi system cannot really be used due to its limitations in data and timeliness. Using a SIF-based data collection, more data can be collected and used for things like EdFacts reporting, linking with WDQI and higher education, researchers, accountability offices, finance offices, other departments at a state DOE, and other state agencies.

A SIF-based data collection incurs the least work for the school district personnel. With a SIF-based data collection, they only have to send data once for many functions. With Ed-Fi, it is so specialized that a district has to still send state data reports (via uploads or some other way), but now they also have to make sure that their data is correct for the Ed-Fi dashboards. The use of the Ed-Fi ODS more than doubles the work of the district data personnel.

The SIF ODS is populated using SIF agents, as with SIF 2.X, and with REST calls, as with SIF 3.X. Most SIS applications have SIF agents and there are multiple third-party companies that can build SIF agents for vendors that do not have the time or the resources to do it themselves. With all of the states that currently use SIF for data collection, there are very few SIS applications that are not already prepared for SIF vertical reporting.

How does a state implement a SIF-based data collection with Ed-Fi?

CPSI has developed a method for using a SIF-based ODS and data collection and integrating that with the Ed-Fi dashboard system.

First, the SIF-based staging ODS is built using the SIF object based standard. This ODS, using CPSI's xDStore, is capable of collecting real-time data with partial data sets. This database is 2nd normal form with child and parent tables that are not coupled by a foreign key. This allows the ODS to receive data in real-time that is asynchronous and allows for validation feedback on the fly. Data flow that is out of sync is not impacted by the object relationships and the SIF ODS stores partial records. The only mandatory element required by the ODS is the Unique Key for each record. Validation processes, using CPSI's xDValidator, will capture the missing elements and error data and notify the data owners at the LEA that they need to correct data at the source. The SIF ODS is capable of requesting all of the missing data objects without any human intervention so that complete sets of data are made available to the Ed-Fi data system.

Second, the data needs to be validated for business rules and data quality based on the Ed-Fi requirements. Using the xDValidator, all Ed-Fi data validations for mandatory elements and code sets are enforced on the SIF ODS and presented to the data owners at the district for correction at the source of the data. If data is required by Ed-Fi and not collected by the SIF data collection process, the data will need to be added as part of the SIF agents at the districts as mandatory for the state. The data will be validated when entering the SIF ODS in real time. Districts receive reports of data errors that need to be fixed so that data can flow forward into the Ed-Fi ODS for processing into the dashboards.

Third, some data will not be collected via the SIF data collection process, such as with assessments and other agency data. This data will still be loaded into the SIF ODS via web services, a web page that allows for file uploads, an automated SFTP process,

About CPSI?

Schools and districts use data that comes from many different applications and sources. The problem is that the data must be re-entered many times into each application via manual data entry. Without some way to integrate the data, keep it up to date, and check it for accuracy, the data the teachers and staff uses every day becomes stale and inaccurate. This wastes time in the classroom for teachers, extra tech support tickets for IT, and frustrated administrators who are trying to analyze and report on the data.

CPSI helps districts overcome these difficult challenges. Our solution collects data from the many different applications in real time, checks the data for accuracy, and standardizes the data for reporting, BI and analysis. We help schools and districts finally leverage all kinds of data – from rosters to assessments to attendance and more—across the entire district.

With CPSI, all of the school and district staff can be confident that the information on their computer or device always reflects the true "picture" of the data in the district. Our solution provides everyone in the district – from students to teachers to superintendents – with the underlying foundation to provide up-to-date and trustworthy information in all of the applications they use.

Implementing SIF as the data collection layer allows states to create a solid foundation for populating the Ed-Fi dashboards successfully.

or a manual SFTP process for occasional data. The file uploads can be completed via CSV files, JSON, XML, Ed-Fi Entities, or CEDS entities. Non-SIF data is associated to the SIF data based on natural keys or surrogate keys, if they exist. SIF RefIDs are cross-referenced to the non-SIF data. The data is associated to the non-SIF data key to the RefID key that exists in the SIF data objects. All data will be tied back to the proper students. The data loaded into the SIF ODS outside of the SIF data collection process follows the same validation rules required by Ed-Fi.

Fourth, the Ed-Fi ODS and data system is modified to include a RefID (GUID) in the tables. Validated and complete data sets are then moved into the Ed-Fi ODS and made available to the dashboards using the standard Ed-Fi processes via SSIS packages. When data is updated at the district and re-validated, the data is automatically updated in the Ed-Fi ODS.

Some final words:

Many people will say that a REST API is sufficient to use for Ed-Fi data collection and that SIF is not necessary. But there are several reasons why SIF should be used for the data collection process from the districts in order to populate Ed-Fi data tables and the dashboards.

- The SIF staging ODS is designed to be a true staging ODS that can receive real-time data from both SIF agents and REST services. The Ed-Fi ODS is a 3rd normal form ODS that is very rigid in design as an end storage ODS.

- The SIF staging ODS is designed to receive data out of sequence and is able to store the data. The Ed-Fi system does not provide this functionality

- The SIF staging ODS is capable of identifying missing objects and automatically requesting the missing data from the source system with no additional human intervention. The Ed-Fi system does not provide this functionality.

- The SIF staging ODS is designed to implement data quality and business rule validations, log all errors, and report the errors back to the data owners at the districts. The Ed-Fi system does not provide this functionality.

- The SIF staging ODS is designed with a single primary key per object which is immutable and cannot be changed by a human. In contrast, the Ed-Fi primary key is a combination of natural fields that are mutable and can be changed by a human.

- A record can never be orphaned in the SIF staging ODS since the primary keys are immutable and can never change. On the other hand, records are always orphaned in the Ed-Fi ODS if an update is done to any field that is part of the primary key. Unfortunately, this happens quite often. A full deletion of all of the data must occur to keep the Ed-Fi ODS from having orphaned records.

CPSI has the experience you need to implement SIF and Ed-Fi at the state level.

If you would like to learn more about CPSI or would like to schedule a demo of any of our products, please contact us.
http://www.cpsiltd.com
(800) 659-8240
sales@cpsiltd.com.