

## Improving Data Quality Using *Estudias*

ZogoTech recommends a process for identifying data inconsistencies in student information systems (Datatel Colleague, SCT Banner, PeopleSoft, etc.), cleaning the data and keeping it clean.

Data quality in many student information systems is notoriously bad. For example, many systems store the same information in multiple locations with different values, values are not checked on data entry and rules describing the connections between data elements are not enforced. *Estudias* discrepancy checks and notifications can provide an important piece in improving data quality.

### Types of Data Inconsistencies

There are several types of data inconsistencies:

1. Information that is definitely incorrect based on a set of rules/constraints (i.e. students with GPA > 4.0)
2. Information that is possibly incorrect, but needing follow-up to confirm (i.e. students who are less than 16 years old may indicate a data entry error for the student's date of birth or there may indeed be students who are taking classes at 16 years old).
3. Information that is incorrect, but can not be automatically flagged (i.e. incorrect phone numbers).

*Estudias* allows schools to create queries to identify the first two types of data errors.

### Data Cleansing

There are a couple approaches to cleaning the data. First, *Estudias* could clean the data after downloading it from the source system (Datatel Colleague, SCT Banner, PeopleSoft, etc.). Second, the data could be cleaned on the source system. ZogoTech favors the second approach. Cleaning the data only on *Estudias* while users continue to use the source system (which is still dirty) creates inconsistencies: users of the source system may come to one conclusion while *Estudias* users come to another.

## ZogoTech's Approach

This is an outline of how ZogoTech recommends institutions improve data quality. It combines processes, tools (*Estudias*) and documentation.

1. Data Analysis
  - a. Top Fifty: Institution and ZogoTech work together to identify the fifty most important data elements (GPA, headcount, etc.)
  - b. Institution and ZogoTech work together to identify rules that describe the constraints of the data elements (i.e. GPA must be  $\leq 4.0$ ) and the rules that connect them (i.e. if a class is "Dropped" or "Cancelled" the GPA credits must equal 0).
  - c. *Estudias* Discrepancy Checks: ZogoTech creates "discrepancy checks" to allow users to identify data elements that do not conform to the rules in (b).
2. Data Cleansing – Institution either updates the source system in batch or by manually changing incorrect values.
3. Continuous Data Integrity
  - a. Processes – users are instructed in the correct values to enter into the source system.
  - b. Notifications – Rules are entered into *Estudias* to send emails to specified users when a user has entered values into the source system that do not conform with rules identified in (1b).

With this approach, data can be cleaned and stay clean. In fact, the source system actually gets cleaner over time. Structured rules describing relationships between data elements serve as a form of documentation for anyone wanting to do reports. Combined with verification against reports like IPEDS, *Estudias* can provide the critical third point in a data integrity triangulation effort.