MELISSA DATA

Melissa Data Quality Integration Toolkit for Microsoft SQL Server SSIS Key to Success of Medical Education Application

The Northern Ontario School of Medicine (NOSM) in the Canadian Province of Ontario is mandated to educate doctors, and offers an undergraduate MD program, various residency programs, and a Continuing Health Professional Education (CHPE) program for practicing physicians.

In order to manage the CHPE program, NOSM had developed a business application to administer information and activities such as course descriptions, participant registrations, and transcripts management. However, the system was manually intensive and time-consuming, and actually consisted of a combination of disparate Microsoft SQL Server applications, Microsoft Access data sources, and paper processes. NOSM wanted to improve its program by creating a single application, and completely automate the process of scheduling and managing its CPHE events.

CRITICAL NEED FOR DATA QUALITY PROCESSES

To accomplish this application development project, NOSM invited various systems integrators to complete requests for proposals (RFPs) to obtain bids for the project. A critical capability that the development partner chosen needed to provide was expertise in the area of data quality because the initial data to be consolidated was not standardized and well-monitored.

Ultimately, two systems integrators that work with Melissa Data were awarded the project in large part because of their expertise in data quality and the strength of the Melissa Data Total Data Quality Integration Toolkit (TDQ-IT) for Microsoft SQL Server Integration Services (SSIS). The TDQ-IT components provide the full spectrum of data quality needs within the SSIS data flow, and can simply be added to SSIS as drag-and-drop icons within Visual Studio.

NOSM was able to integrate data from about 30 distinct source systems.

The two integrators, Fox River Software and Actuality Business Intelligence, won the bid based on a proposal to build a custom application using Ironspeed software as the website application development framework, Microsoft SQL Server as the centralized database, SSIS as the data integration and workflow tool, and Melissa Data TDQ-IT as the data quality solution.

SPECIFIC DATA QUALITY PROCESSES IN THE NOSM APPLICATION

The Melissa Data TDQ-IT provides all of the data quality requirements that have been identified by leading industry analysts. For the NOSM application, the vast majority of these capabilities are being extensively utilized. These include profiling (identifying data quality issues), generalized cleansing (tests to meet business rules), parsing and standardization (e.g., restructuring data into a common format), matching (e.g., finding unique identifiers and performing de-duplication), enrichment (e.g., phone and email validation), and monitoring (checking conformance to data quality requirements).

As one example, Carlier Lauer of Fox River Software indicates that "we utilized the Melissa Data Email and Fuzzy Matching components to take data inputs from six completely different sources (e.g., Microsoft Access, Excel, Flat Files, Novell Groupwise), merge them together, validate the email addresses and create a resulting list of de-duplicated email addresses and names. The total records on the initial load were 50,000 and we came back with 6,000 distinct records that were correct and another 3,500 with invalid email addresses that needed to be manually corrected before they were imported into the centralized system. Since the data sources were from all different departments within NOSM there was quite a bit of duplication which the Melissa Data TDQ-IT detected and eliminated."

SUMMARY

Ultimately, NOSM was able to cleanse and integrate data from about 30 distinct source systems, which is now centralized in a new, automated CHPE system, saving large amounts of time and money. This new capability could not have been accomplished without the TDQ-IT from Melissa Data, and its comprehensive data profiling and data quality capabilities. Because of TDQ-IT's drag-and-drop design, this same result can be accomplished by many other SSIS developers who can easily embed these data quality processes directly into their SSIS data flow transformations. ■

Melissa Data Corp. ■ 22382 Avenida Empresa ■ Rancho Santa Margarita, CA 92688-2112
Phone: 949-858-3000 ■ Fax: 949-589-5211 ■ Toll Free: 1-800-635-4772 ■ www.MelissaData.com