Data Warehousing and Web 2.0 Data Dissemination

Joel M. Lehman

United States Department of Agriculture/National Agricultural Statistics Service

Email:joel_lehman@nass.usda.gov

Address: USDA/NASS, Data Services Branch

1400 Independence Ave

Washington, DC 20250

Abstract N.49

In 2008 the United States Department of Agriculture's (USDA) National Agricultural Statistics Service (NASS) was looking to enhance the electronic data dissemination products for the 2007 U.S. Census of Agriculture. NASS improved upon the legacy web based data dissemination tool and underlying database by creating a generalized data model and a Web 2.0 query application utilizing hierarchical metadata. NASS approached the issues posed by the legacy system by redesigning the data model, web application, and the aggregate metadata at the same time to create a fully integrated data dissemination platform. The redesign included a generalized data warehouse model capable of storing all of NASS' published estimates in a simplified data structure. A data model utilizing data warehousing technology was specifically designed for very fast data retrieval while maintaining ease of browsing. A metadata repository was developed to standardize metadata across NASS' aggregate data processing systems. Standardized metadata, a simplified data model, and a purely data driven application have reduced the time and effort involved in adding new data sources for public dissemination, reduced the potential for errors and provides the ability to compare data at every point in aggregate data processing stream. A data driven, Web 2.0, ad-hoc query application (Quick Stats 2.0) was developed in a rapid development (SCRUM) environment using open source development tools to provide public access to all NASS published statistics. In addition to being a web based ad-hoc query tool, Quick Stats 2.0 was designed to be a data dissemination engine capable of feeding data to external applications. Quick Stats 2.0 provides the ability for public users to view data, create maps, download large datasets, and create and save custom queries. Since the public release in February 2009, Quick Stats 2.0 currently provides access to over 24 million data points from 23 thousand different statistics dating back to 1866. Over 50 thousand users have accessed Quick Stats 2.0 from over 100 countries issuing over 200 thousand queries since February 2009.