

Analytical Use Case Framework for Assessing Data Quality

Anjené Addington,¹ Cora MacPherson,² Brendan Murphy,¹ Naji Younes,² Jason Hunter,¹ David Fetterer,² Ernest Sohn,¹ Dania Shor,¹ Juergen Klenk¹

¹Booz Allen Hamilton, Rockville, MD

²Social & Scientific Systems, Inc., Rockville, MD

Introduction: Data collected as part of a NCS Vanguard Study are being used for data-driven, evidence-based planning of recruitment and retention, study operations and logistics, and study visit measures for the Main Study. To ensure unbiased and reliable analyses it is imperative that data quality issues be identified and corrected as early as possible. However, it can be difficult to define “good” data in a dataset of this complexity and size, especially with a diverse set of end users.

Methods: We have used the concept of Analytical Use Cases to provide a framework in which to systematically define the data elements and linkages that are required to support current and planned uses of the data.

Results: Using these definitions it is possible to rigorously check for data quality issues and report them back to study centers in a timely manner with the goal of obtaining high quality data to produce meaningful analyses.

Figure 1: Analytic use cases allow for proactive identification and resolution of data quality issues

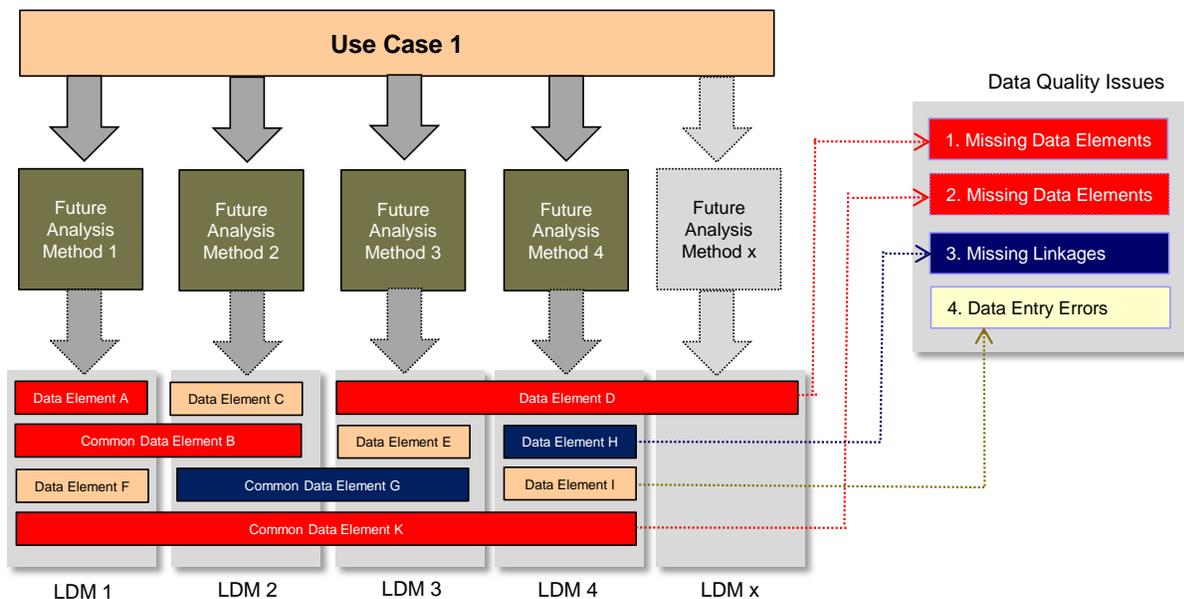
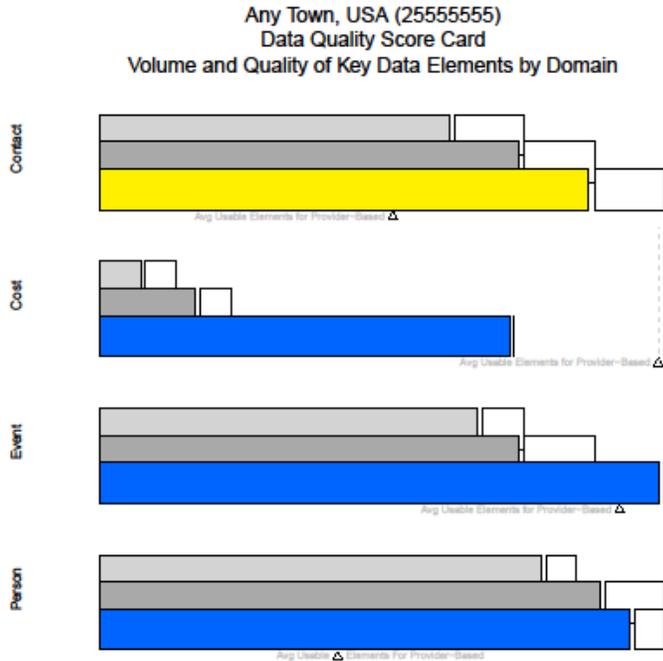


Figure 2: Sample score card report sent to study centers



Filled bars represent usable elements and hollow bars represent unusable elements. Colored bars represent the state of the VDR as of 6/30/2011 : ■ >90%, ■ <50% and ■ 50% to 90% usable data. and bars indicate the state of the VDR as of 6/2/2011 and 6/16/2011, respectively. Your most recent submission occurred on 06/30/2011.