

Data Validation Update

Scope of Validation

- (1) Primary data set validation – statistical analysis of content included in the database, supplemented by meetings and physical verification to ensure data accuracy;
- (2) Supply and Demand validation – confirmation that data integrity is maintained when content is organized into specific supply and demand databases; and
- (3) Integration validation – confirmation that queries to the supply and demand databases return data that are comparable (e.g. if an individual is seeking supply and demand information at a particular geo-code or neighborhood, the data architecture would return relevant, comparable supply and demand data at either level).

Lead Validator

Ganesan Shankaranarayanan, Associate Professor, Babson College

G. Shankaranarayanan (Shankar) obtained his Ph.D. in Management Information Systems from The University of Arizona, Eller School of Management. His research interests cover three primary areas: (1) Modeling and managing data and metadata in information systems, (2) Managing data quality for decision support, and (3) economic perspectives in data management. His research has appeared in journals including Journal of Data Management, Decision Support Systems, Communications of the ACM, Communications of the AIS, Journal of Information Technology Management, Journal of Computer Information Systems, International Journal of Information Quality, and IEEE Transactions on Knowledge and Data Engineering. His research has won the best paper awards at the International Conference on Information Quality (ICIQ) and at the Workshop of Information Technology and Systems (WITS). He serves as the Area Editor of the International Journal of Information Quality and as an Associate Editor of the ACM Journal for Data and Information Quality.

Shankar has taught several courses ranging from introductory courses in information systems to advanced courses in data management and systems design. He received the Computer World Laureate honor from the Computer World Honors Program in 2008.

Executive Summary from November 2012 Interim Data Validation Report

“Boston Public Schools (BPS) provides a set of data to the public via their website (www.bostonschoolchoice.org). This data includes data on choices that parents make to request placement of their children in transitional grades within BPS, data on schools that are part of BPS, and data on students enrolled in the BPS system. We were tasked with validating the accuracy or correctness of these three datasets. Further, we were also tasked with assessing the integrity of the combined dataset that includes school data and student data.

With the help of Ms. Rebecca Frisch from the Mayor's office and with support from Mr. Peter Sloan and several others from BPS, we were able to complete a large part of this validation. We validated the choice data by comparing it with the raw data on choices collected from student applications.

Our conclusion is that there are no discrepancies between the choice data provided to the public by BPS and the choice data collected from student applications. We validated the student data by comparing the student data provided to the public with the raw student data collected and managed by BPS. Our conclusion based on our validation processes is that there are no discrepancies between the student data provided to the public by BPS and the raw student data captured and managed by BPS.

We further analyzed the design of database tables in a database that integrates student and school data. Our conclusion based on our analyses is that the current design that integrates student and school data does preserve the integrity of the two datasets. Further, queries against this database design are guaranteed to provide relevant and valid data.

We have not completed the validation of the School data at this time. We will provide an updated report when this task is completed.”

Next Steps

Validation of school data is ongoing.