Abstract
22nd International Blaise Users Conference
Trondheim, Norway
September 2026

Automating Missing-Data Classification by Applying FPS IsVisited Metadata Upstream in the Blaise Extraction Pipeline

By
April Beswick
Youhong Liu
Mohammad Mushtaq
The University of Michigan

As the Panel Study of Income Dynamics (PSID) transitioned from a CATI-only design to a mixed-mode CATI–Web environment, a key challenge emerged in the post-collection data pipeline: distinguishing fields that were *off-route* due to instrument skips from fields that were *on-route but unanswered* by respondents. In the relational data output, both conditions initially appear as identical generic missing values. Only the FPS file—specifically the *IsVisited* metadata—provides the information needed to classify these outcomes accurately.

Our earlier workflow applied *IsVisited* information manually after relational extraction, merging FPS information at the individual cell level across all data-out tables. This process was labor-intensive and produced inconsistencies in missing-data coding across modes. To address these limitations, we redesigned the workflow to shift FPS-based logic upstream. Using the FPS file, we now programmatically assign the *Refused* special answer to entire Web cases—or to specific Web sections—prior to extraction. This harmonizes CATI and Web cases at the source, even in Web sections where respondents were never presented with an explicit "Refused" option.

The resulting process is fully automated, eliminates manual intervention, and produces more reliable, mode-agnostic missing-data classifications. This paper describes the revised approach, its implementation within the Blaise environment, and its implications for mixed-mode instrument design and downstream data quality.