Variance Estimation for Product Sales in the 2017 Economic Census: Utilizing Multiple Imputation to Account for Sampling and Imputation Variance

Matthew Thompson
Katherine J. Thompson

U.S. Census Bureau*

GASP! Workshop
August 24-25, 2018

**The views expressed in this presentation are those of the authors and not necessarily those of the U.S. Census Bureau*

Economic Census Background

- Not strictly a census
  - Multi-units and large single-units selected with certainty
  - Small single-units sampled

Economic Census Background Data Items Collected

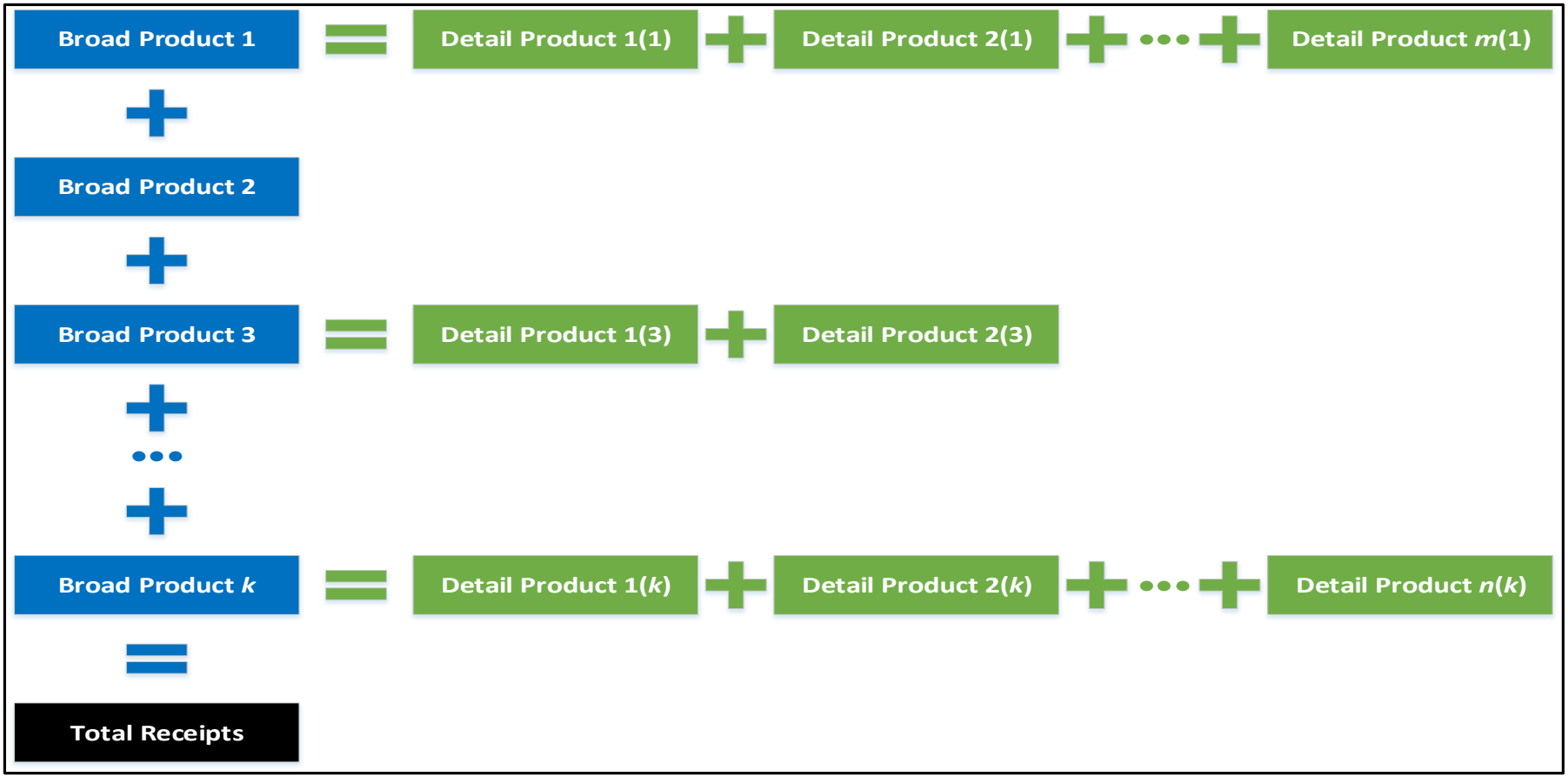
"General Statistics"

- Examples: Total receipts, Annual payroll, and 1st Quarter employment
- Complete universe created using administrative records and imputation

Product Sales

- Only asked of sampled establishments
- Sample weights used to account for non-sampled establishments
- Two types: broad and detail
- Final product sales estimates are produced by calibration to stratum-level receipt totals

Economic Census Background
Product Sales Data



Research Challenges

- Dedicated Team
  - Short time frame (≈ 12-15 months)
  - Relative inexperience of team members with variance estimation
- Magnitude of the problem
  - ≈ 1,000 industries and ≈ 8,000 products
- Historical data limitations
  - Classification differences (to NAPCS)
  - Collection differences (to electronic)
  - Unit collection differences (from varied to $1,000)

Research Team

Research Team

- ≈ 1,000 industries
- ≈ 8,000 products
  - Broad products
  - Detail products
- Calibration Weighting (i.e., post-stratification)

Research Team

Research Team

- ≈ 1,000 industries
- ≈ 8,000 products
  - Broad products
  - Detail products
- Calibration Weighting (i.e., post-stratification)

Research Team

- ≈ ~~1,000~~ 21 industries
- ≈ ~~8,000~~ Top 4 products
  - Broad products
  - ~~Detail products~~
- Calibration Weighting (i.e., post-stratification)

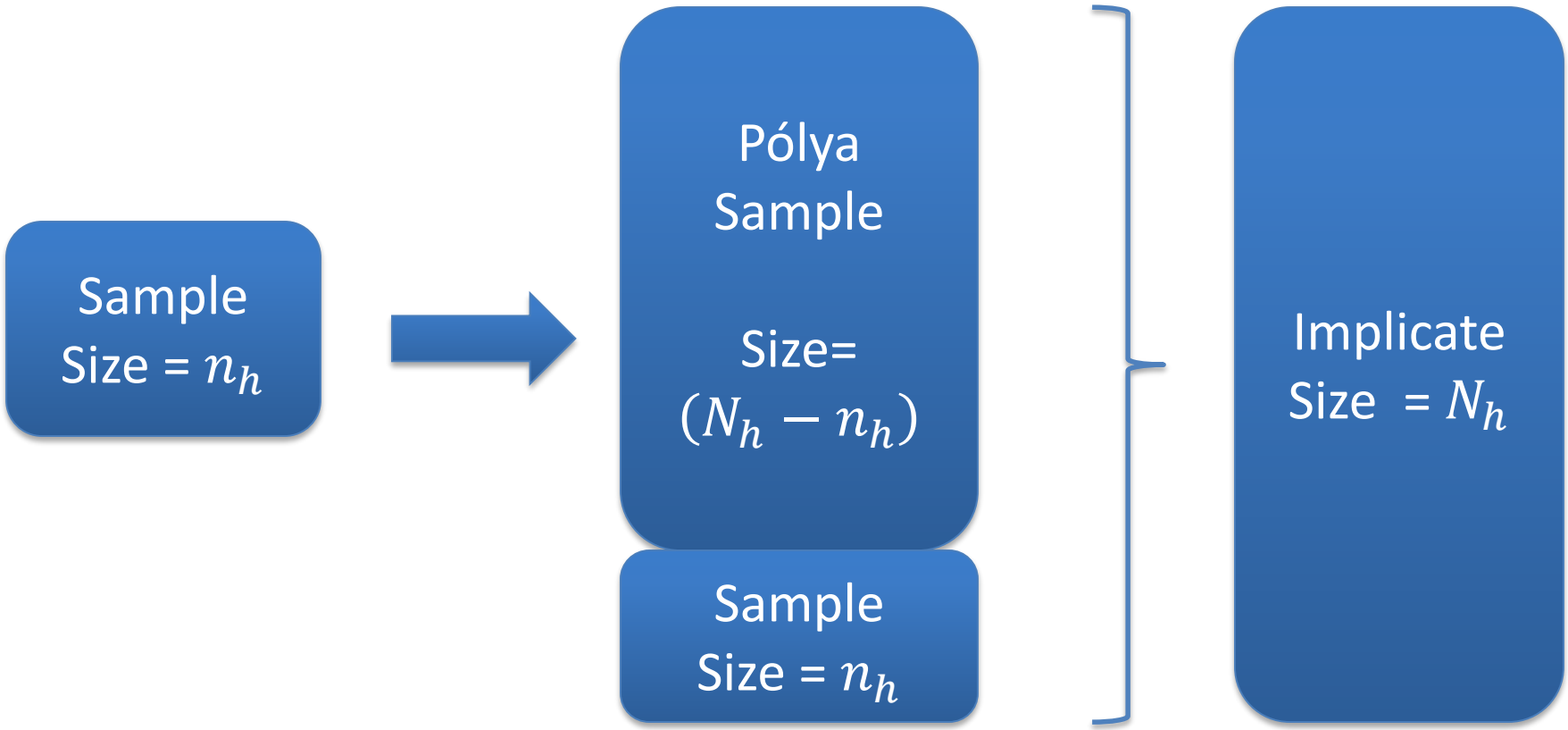
Research Evaluation

Perform simulation studies

- Two initial studies
  - Sampling Variance (Recommend: FPBB)
  - Variance Due to Imputation (Recommend: ABB)
- Final simulation of recommended method

- Recommendation: **FPBB-ABB**

Finite Population Bayesian Bootstrap (FPBB)



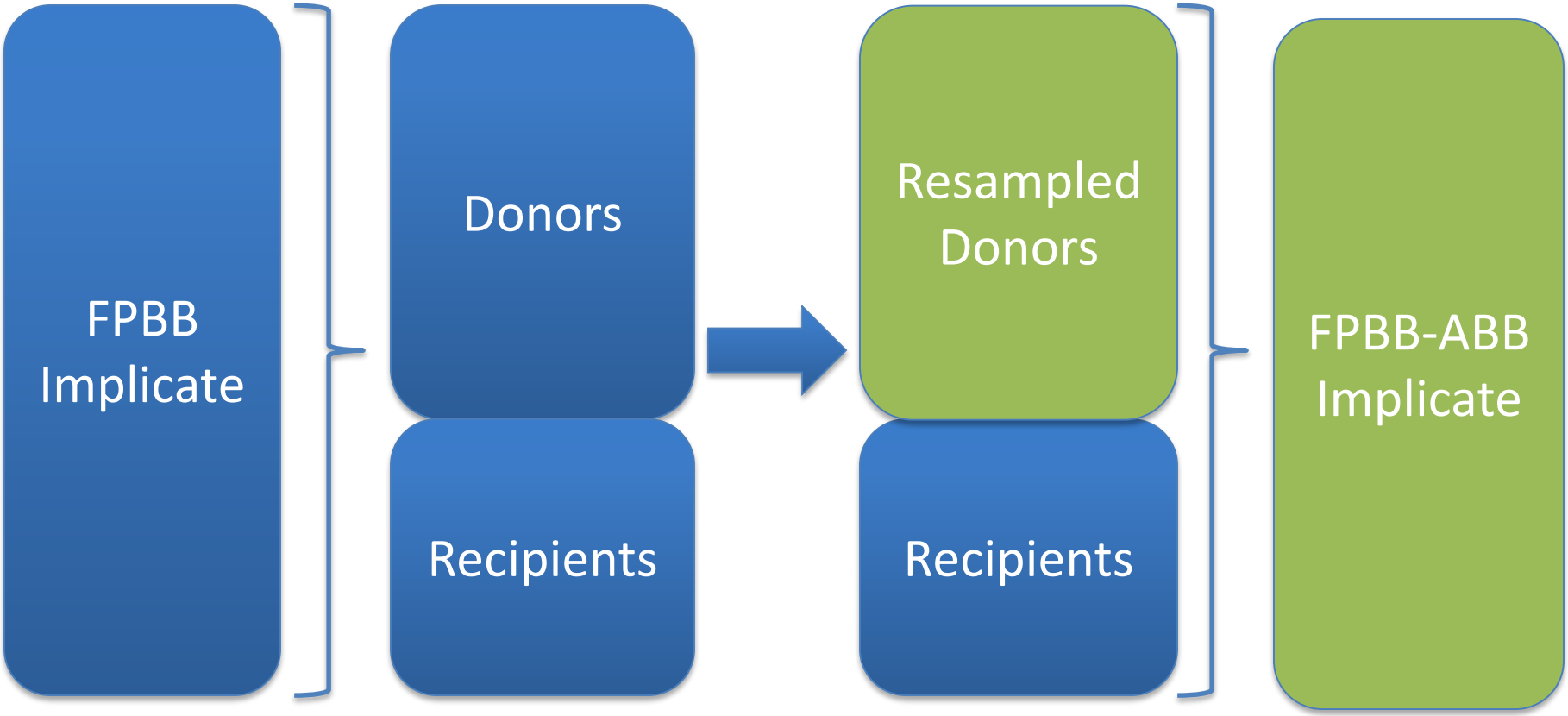
Finite Population Bayesian Bootstrap (FPBB)

- Create an implicate by drawing $N_h - n_h$ establishments from the sample with probability for the kth selection

$$p_{h,k} = \frac{\left(w_i - 1 + \frac{l_{i,k-1}(N_h - n_h)}{n_h}\right)}{N_h - n_h + \frac{(k_h - 1)(N_h - n_h)}{n_h}}$$

- Add the $N_h - n_h$ selected establishments to the original sample to complete the implicate

Approximate Bayesian Bootstrap (ABB)



FPBB-ABB

The FPBB-ABB estimate of variance is

$$\hat{V}_{final} = \hat{V}_{samp} + \frac{1}{B}\hat{V}_{imp}$$

- $\hat{V}_{samp} = \left(1 + \frac{1}{B}\right)\left(\frac{1}{B-1}\right)\sum_{b=1}^{B}[FPBBAVG_b - AVG]^2$
- $\hat{V}_{imp} = \left(1 + \frac{1}{C}\right)\left(\frac{1}{C-1}\right)\sum_{b=1}^{B}\sum_{c=1}^{C}\left[TOT_{b,c} - FPBBAVG_b\right]^2$

- B is the number of FPBB implicates
- C is the number of ABB implicates

Implementation Team

Research Team

- ~~1,000~~ 21 industries
- ~~8,000~~ Top 4 products
  - Broad products
  - ~~Detail products~~
- Calibration Weighting (i.e., post-stratification)

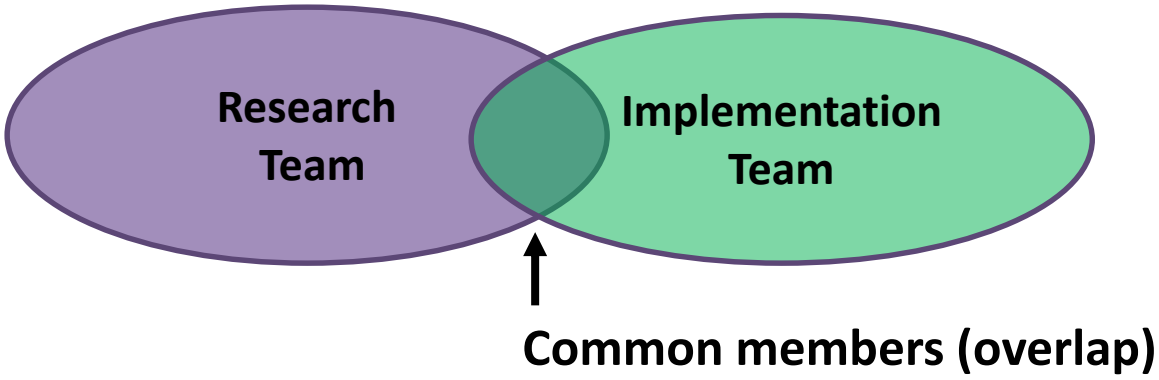
Implementation Team

- ≈ 1,000 industries
- ≈ 8,000 products
  - Broad products
  - Detail products
- Calibration Weighting (i.e., post-stratification)
- "Non-donors"
- Zero Receipts cases
- Processing time …

Implementation Concerns

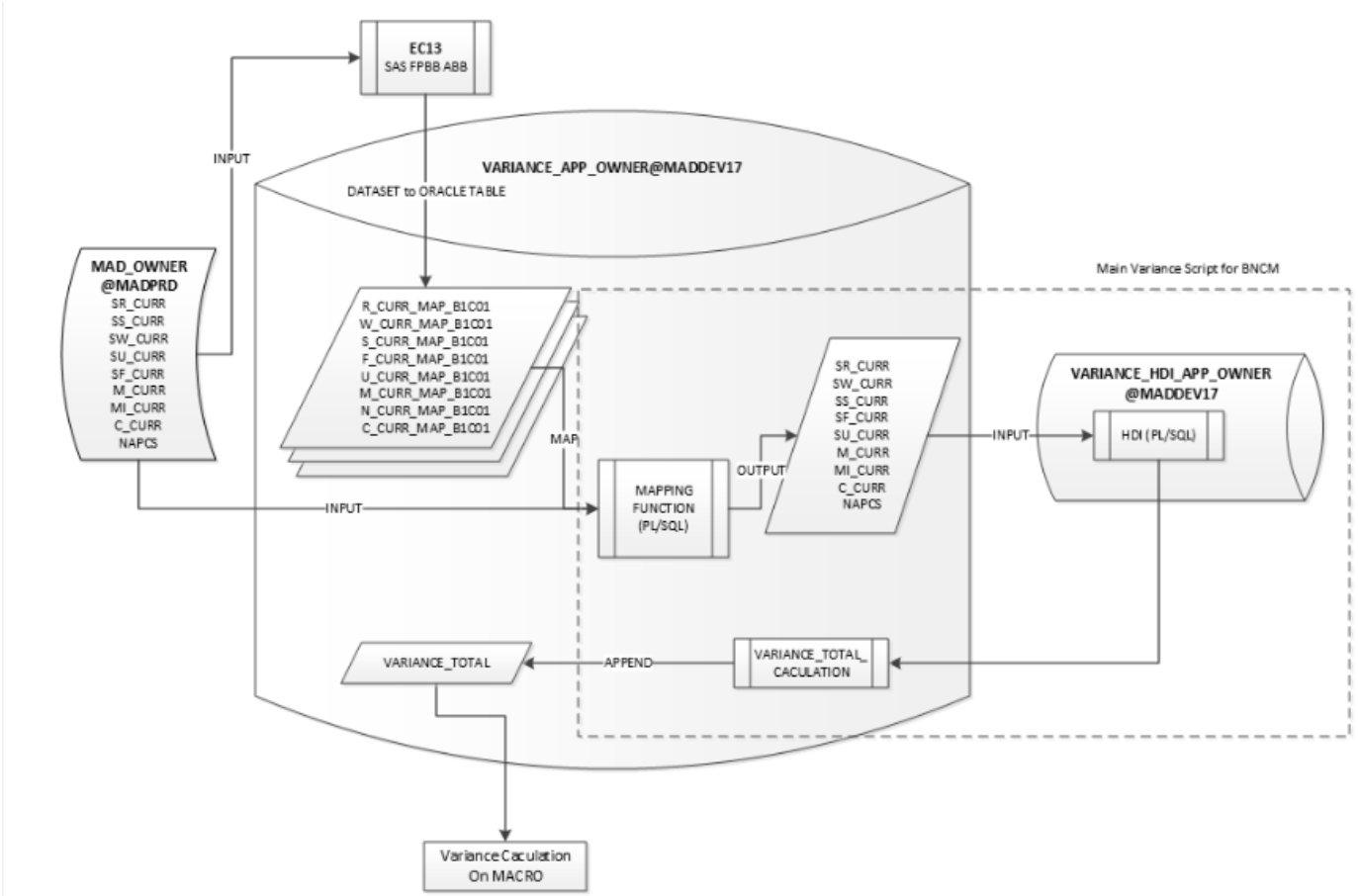
- Time
  - To prepare the system
  - Variance Estimation run time
- Knowledge transfer from research team to production programmers and methodologists
- Inflexibility of existing systems

Implementation Team

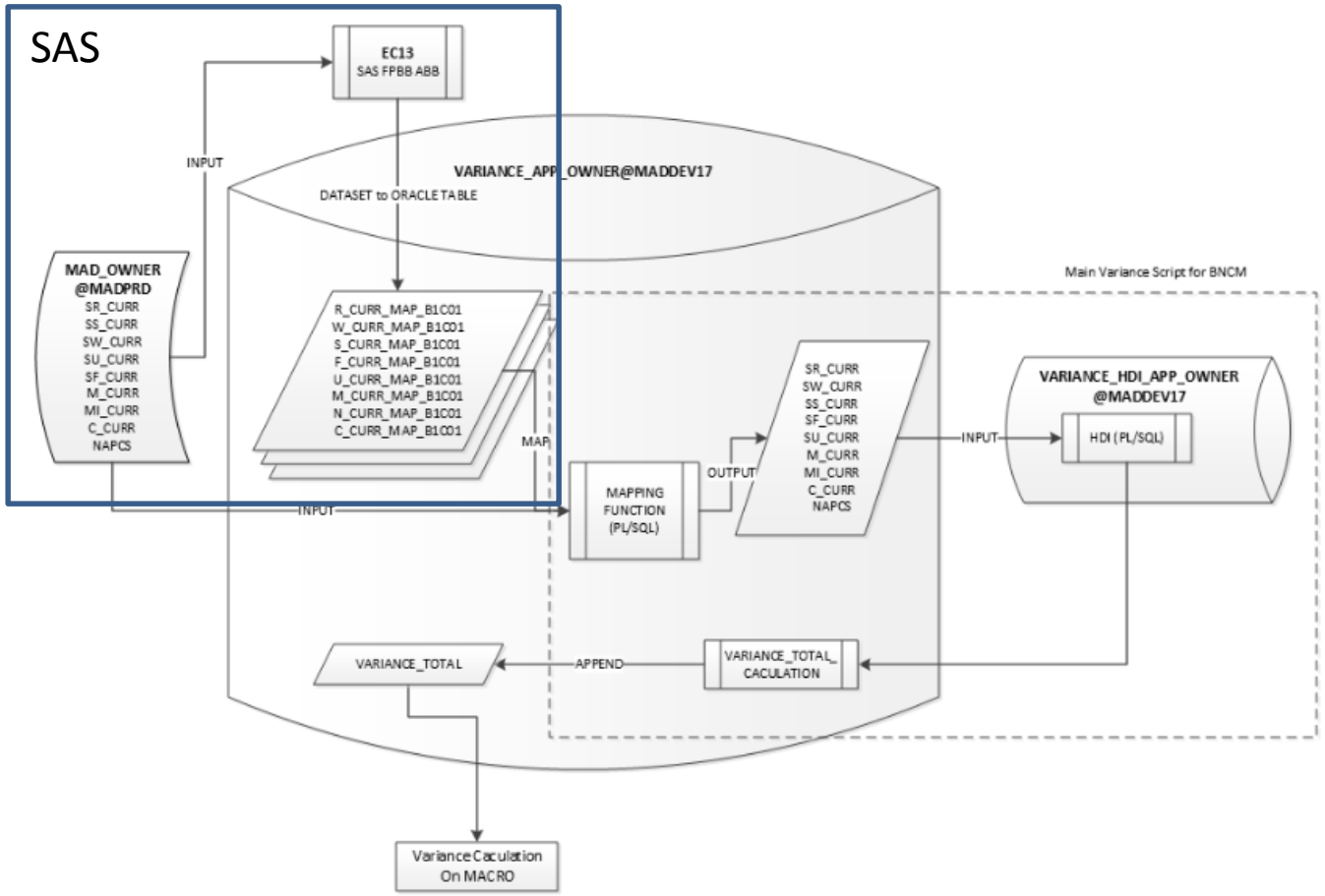


- Overlap
  - Research team leads
  - SAS programmer
  - Project Managers w/ functional requirements
- New members
  - Subject Matter Experts
  - Programmers
  - Methodologists

Time & Knowledge Transfer

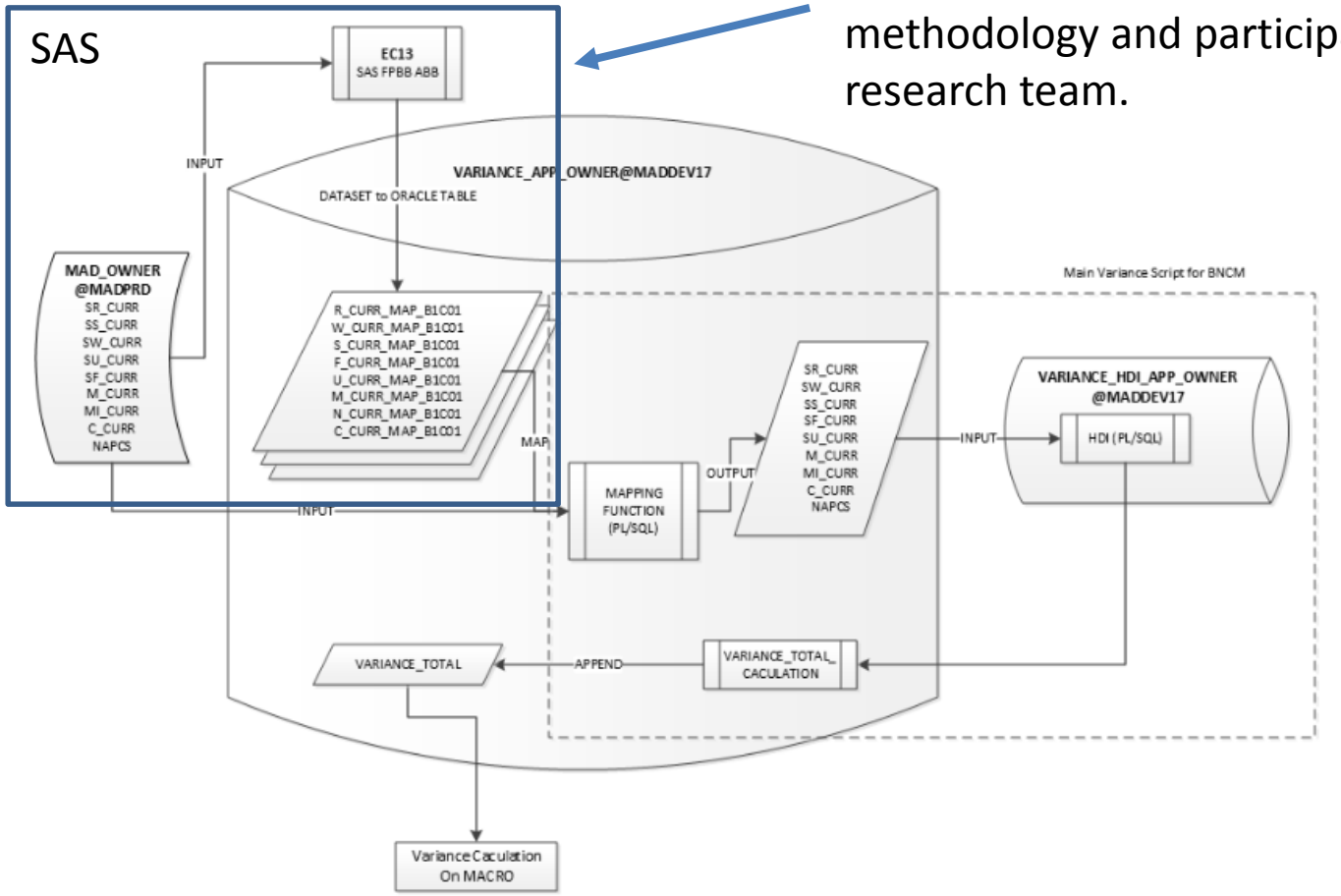


Time & Knowledge Transfer

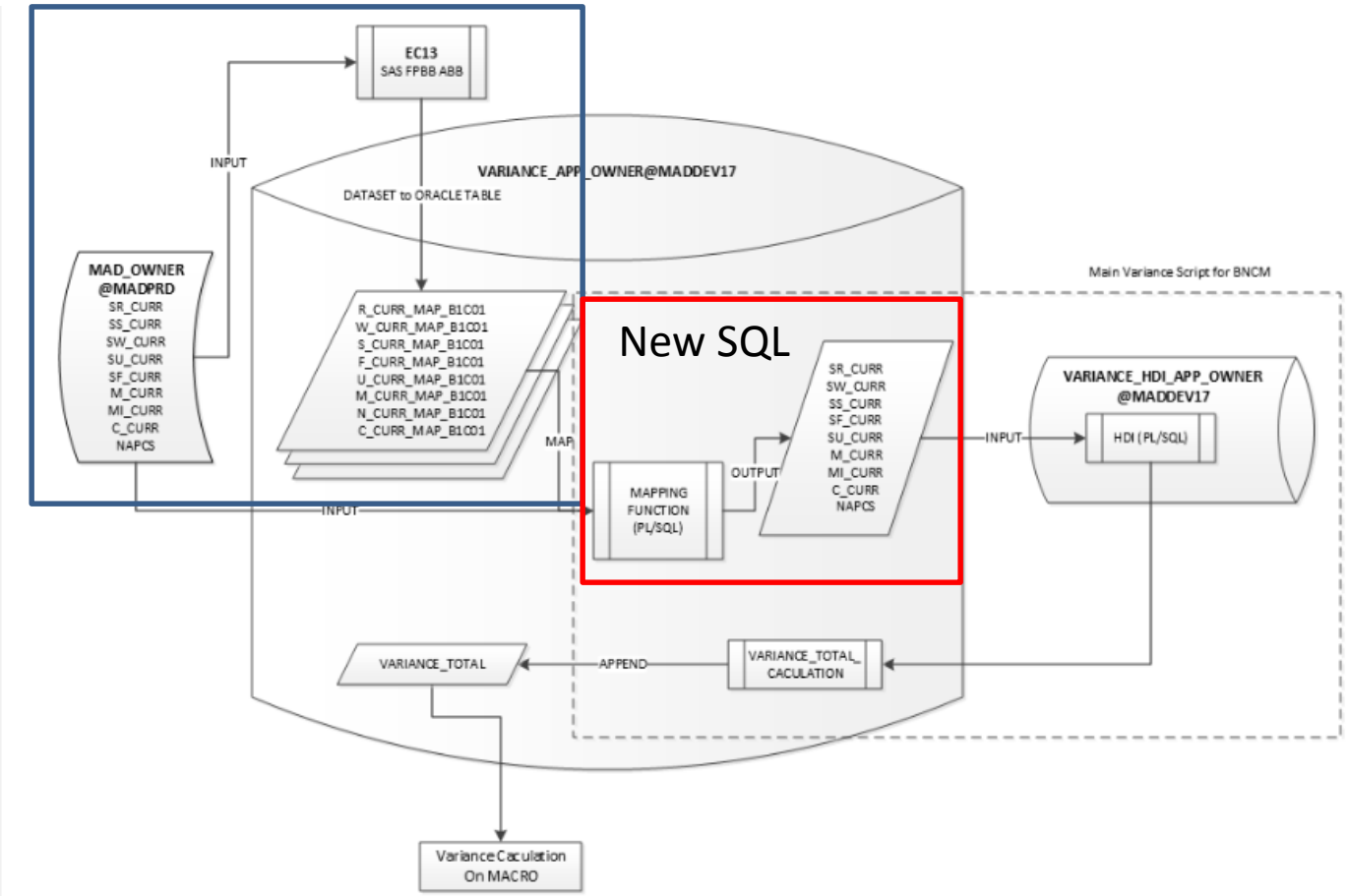


Time & Knowledge Transfer

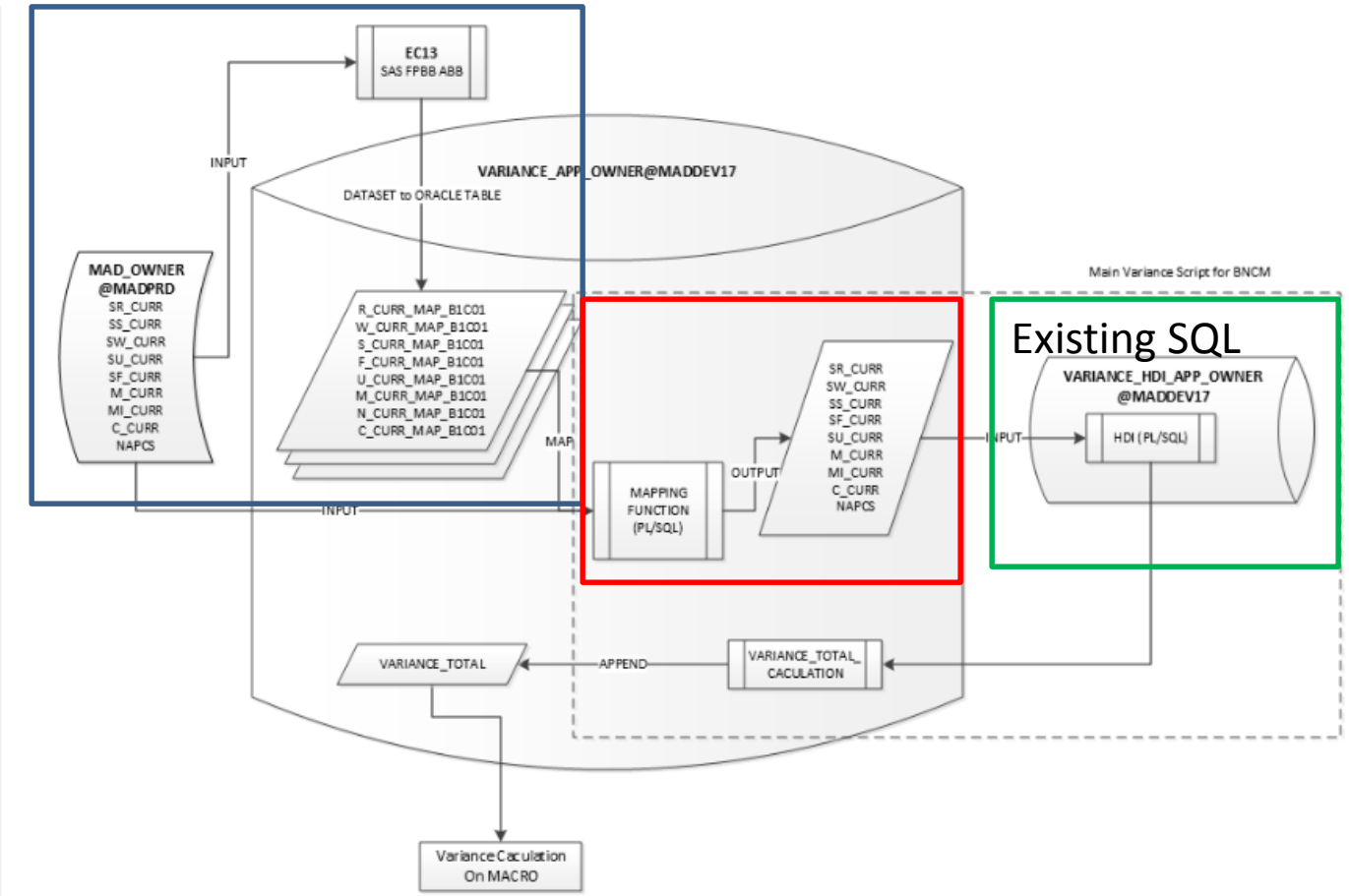
Programmer familiar with FPBB-ABB methodology and participant in research team.



Time & Knowledge Transfer

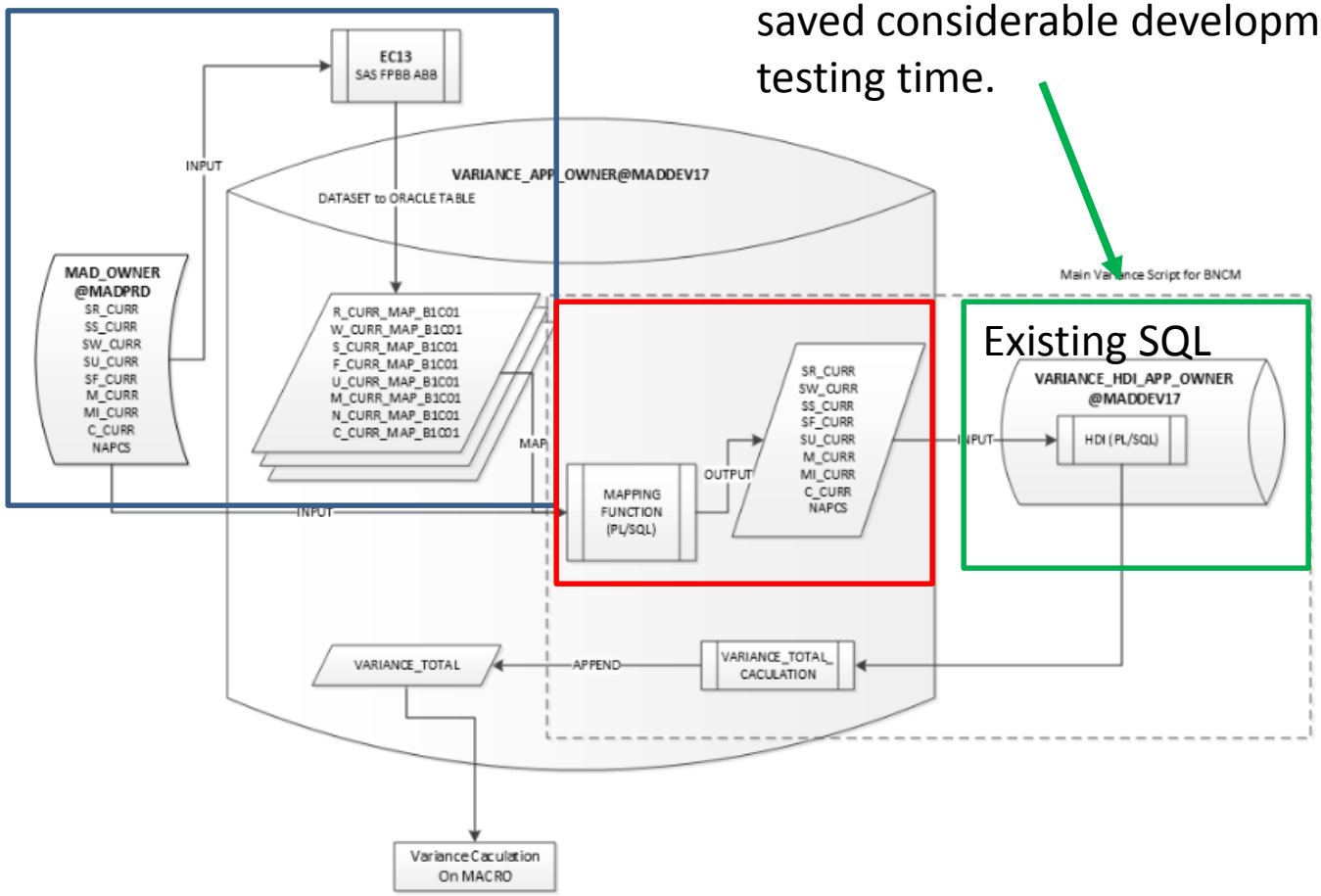


Time & Knowledge Transfer

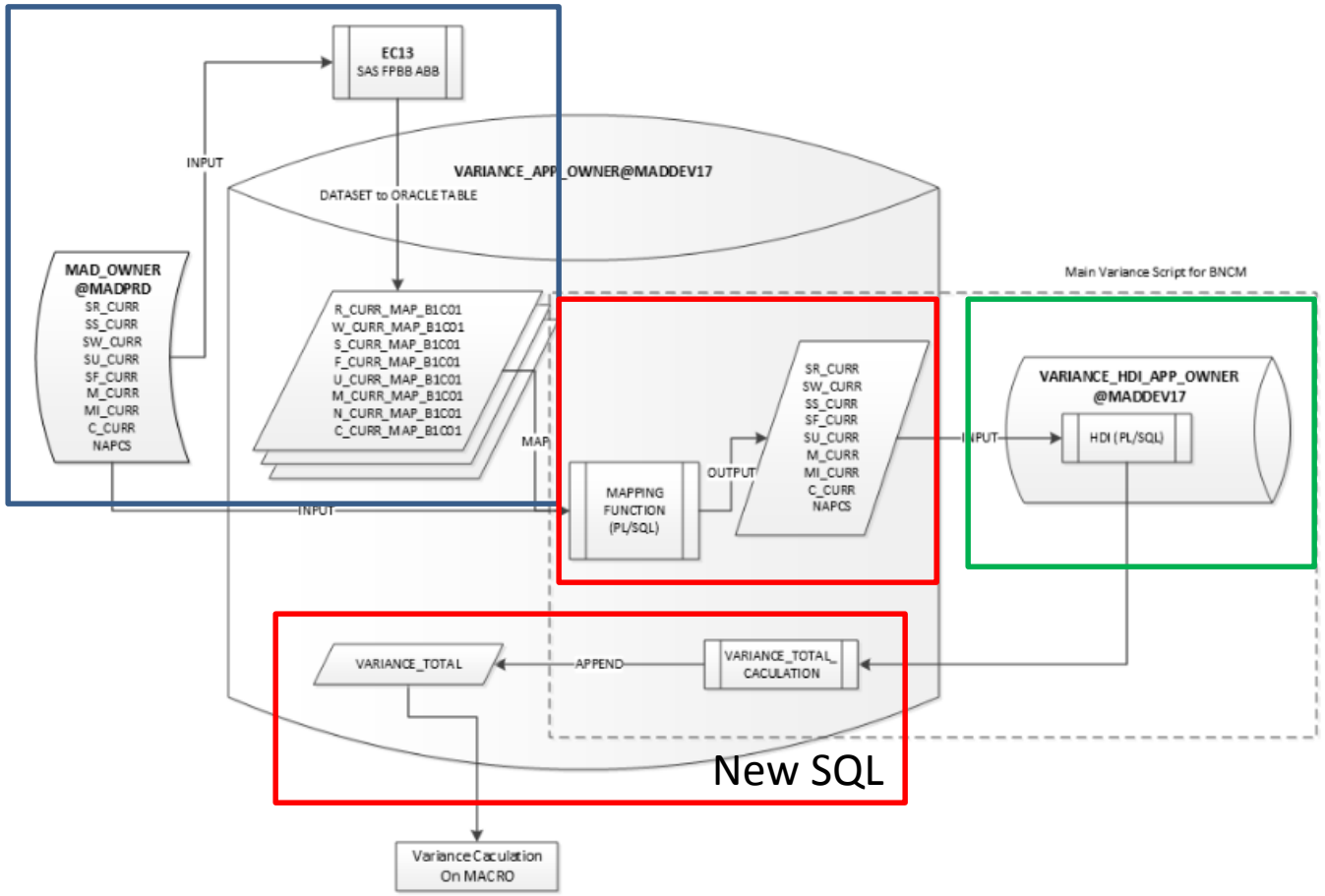


Time & Knowledge Transfer

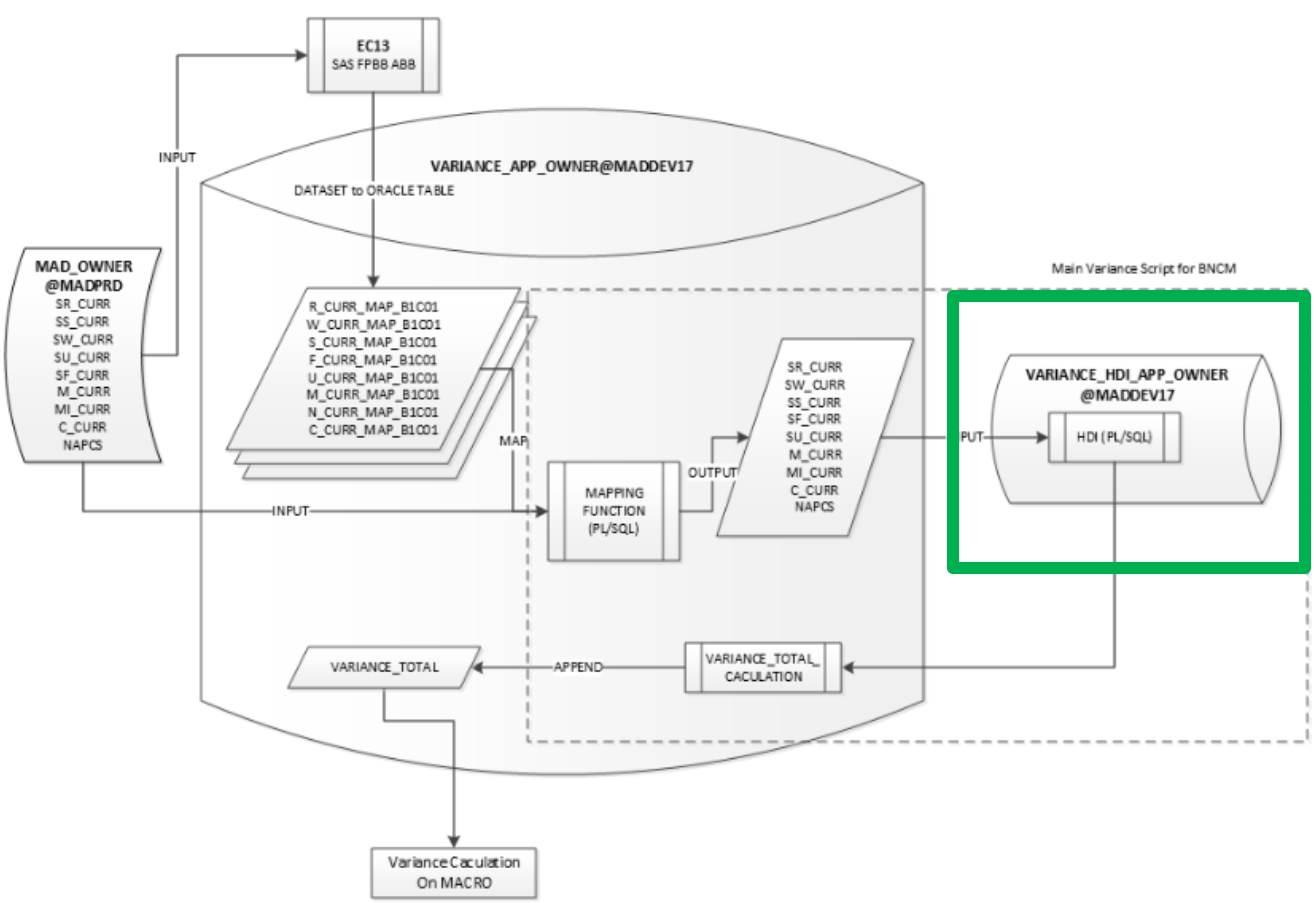
Utilizing existing program and test deck saved considerable development and testing time.



Time & Knowledge Transfer



Inflexibility of Existing Systems



Inflexibility of Existing Systems

- Existing HDI process fixed:
  - Imputation method by imputation cell – Random or NN
  - Cell collapsing methodology
  - Handling of detailed products

Conclusions

- Plan ahead – Involving potential implementation team members in the later phases of research can be a huge benefit
- Leverage existing resources – Give careful thought to areas where existing programs and data can be utilized
- Make sure existing systems are well documented and well understood.

Acknowledgments

Thanks to all of the members of the research and implementation teams.

Contact: matthew.thompson@census.gov