

Growing a Modern Edit and Imputation System

Megan Lipke

Darcy Miller, Vito Wagner, Karl Brown &
Vikas Agnihotri

National Agricultural Statistics Service
UNECE 2022

Acknowledgements

Edit and Imputation Modernization Team (EIMT)

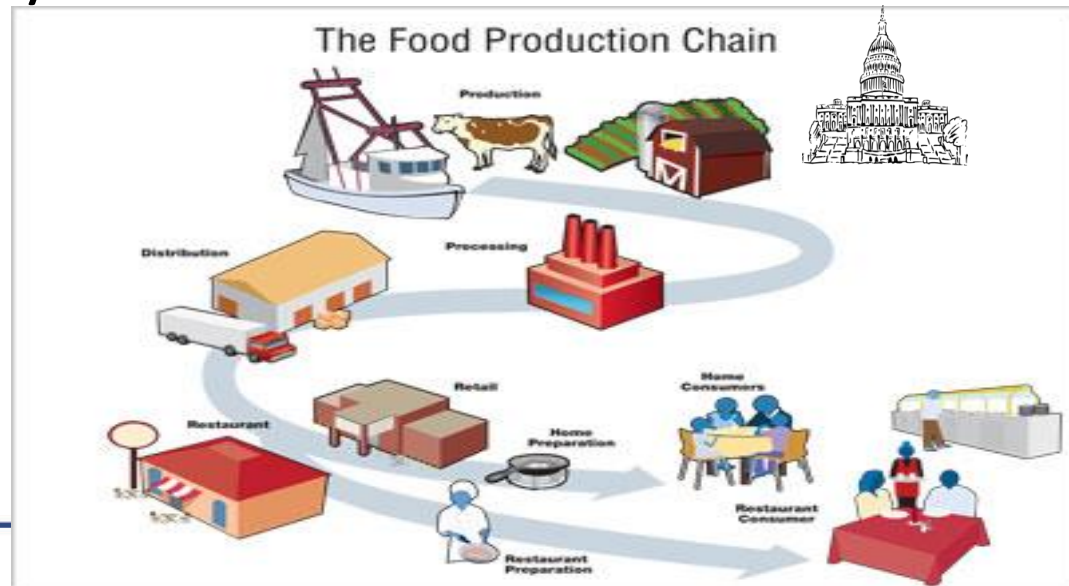
Members from all divisions of NASS

- **Executive Sponsors:** Joe Parsons and Linda Young
- **Business Council Sponsors:** Denise Abreu
- **Strategic Planning Office Representative:** Nick Pallotta
- **NASS Staff**

THE FINDINGS AND CONCLUSIONS IN THIS PRELIMINARY PRESENTATION HAVE NOT BEEN FORMALLY DISSEMINATED BY THE U.S. DEPARTMENT OF AGRICULTURE AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY AGENCY DETERMINATION OR POLICY

National Agricultural Statistics Service (NASS)

- Agency in the United States Department of Agriculture (USDA)
- Mission: “The National Agricultural Statistics Service provides timely, accurate, and useful statistics in service to U.S. Agriculture.”
- Over 400 reports annually
- Census of Agriculture
 - 5 years



Editing and Imputation

- Data collected contain missing or erroneous values
- Primary methods to mitigate bias:
 - Unit Nonresponse – re-weighting
 - Item Nonresponse – imputation
- Often, customized code and/or manual process is used
- Major goal is a ‘clean’ dataset where edit logic is met

Current Editing and Imputation Systems

PRISM

- A few large surveys + Census of Agriculture
- Provides editing rules and error flags with interface for some manual imputation
- Automated imputation is largely handled in a separate step
- Analysis step follows editing and imputation

Blaise

- Remaining small-to-medium surveys (>100)
- Provides an interactive editing interface for error flags and an interface for manual imputation
- Most changes to data occur manually through interactive edit screens
- Analysis step follows editing and imputation

Editing and Imputation Review

- NASS continually seeks to improve its products
- 2017 Report
 - document editing and imputation processes at NASS
 - produce a wholistic vision for NASS editing and imputation

Goal

- *Modularize* and automate editing and imputation process for **Blaise** surveys in a *generalized* system
 - Imputation, Deterministic Edits, Automation and Logic (IDEAL)

IDEAL

Fully Automated Edit + Imputation for Blaise

Warning Pass-
Through

Estimation Tools

Automated
Administrative Code
Editing

Warning Pass-Through:

- Two types of errors in Blaise
 - Critical: review, must change value to make record “clean”
 - Warning: review, must change value or “suppress” to make record “clean”
- No longer require suppression for warning errors
- Increase efficiency and allow focus on records that impact estimates

Editing and Imputation Tools:

- Impute must-strata records (primarily the largest producers) that are unit nonresponses
 - Currently, whole-record imputation outside of interface
 - Automate process developed to:
 - Reduce clerical work/save time, minimize human error, provide tracking and repeatability
 - Implemented for two survey programs

Automated Editing for Administrative Codes:

- Automate edit changes needed for some of the administrative codes
- Reduces need for staff intervention for edits with consistent business rules
- Tested and implemented for all Blaise surveys

Imputation, Deterministic Edits, Automation, and Logic (IDEAL):

- Goal: To modularize and automate editing and imputation process for Blaise surveys in a generalized system
 - Decrease staff time editing
 - Increase data quality

IDEAL: Accomplishments

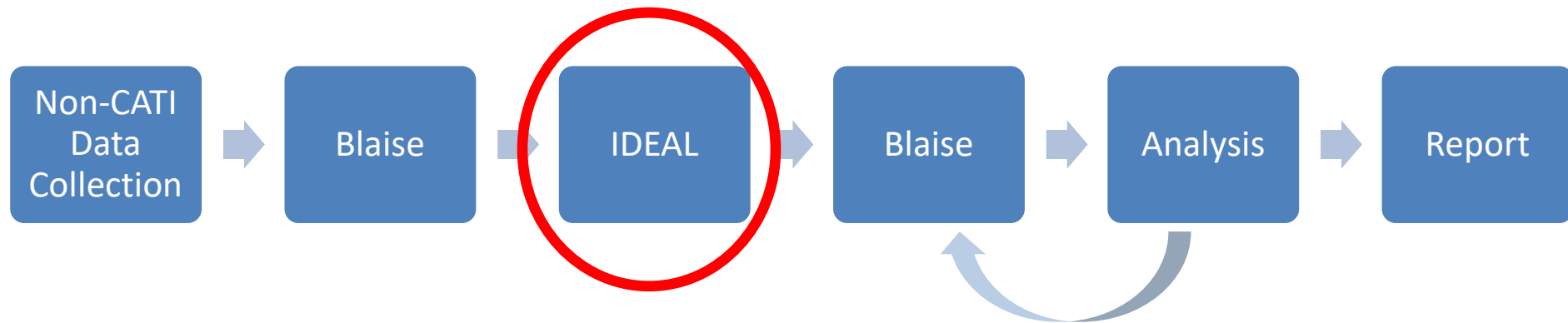
- Making strides!
 - Automated parser that extracts, documents and organizes edit logic from Blaise code for multiple surveys
 - Documented business requirements and conversations with methodological experts
 - Initial application developed for first phase implementation was tested on subsets of previously reported data

IDEAL:

Phase 1 Implementation

- Utilizes Statistic Netherlands R packages to apply rules and make automated changes to data
- User interface (UI) developed to manage/share edit logic and track status of values
 - Ability to change values available but not being implemented in first phase
- Processing occurs in the USDA cloud environment to increase processing speed

IDEAL: Phase 1 Survey Data Flow



IDEAL:

Phase 1 IDEAL Processing



Conclusion

- NASS is implementing a modern editing and imputation process
 - Leveraged internal resources to implement high-impact changes requiring minimal effort to establish
 - IDEAL
 - Start testing full datasets in system for evaluation in September 2022
 - Plan to implement in March 2023

Thank you!

