

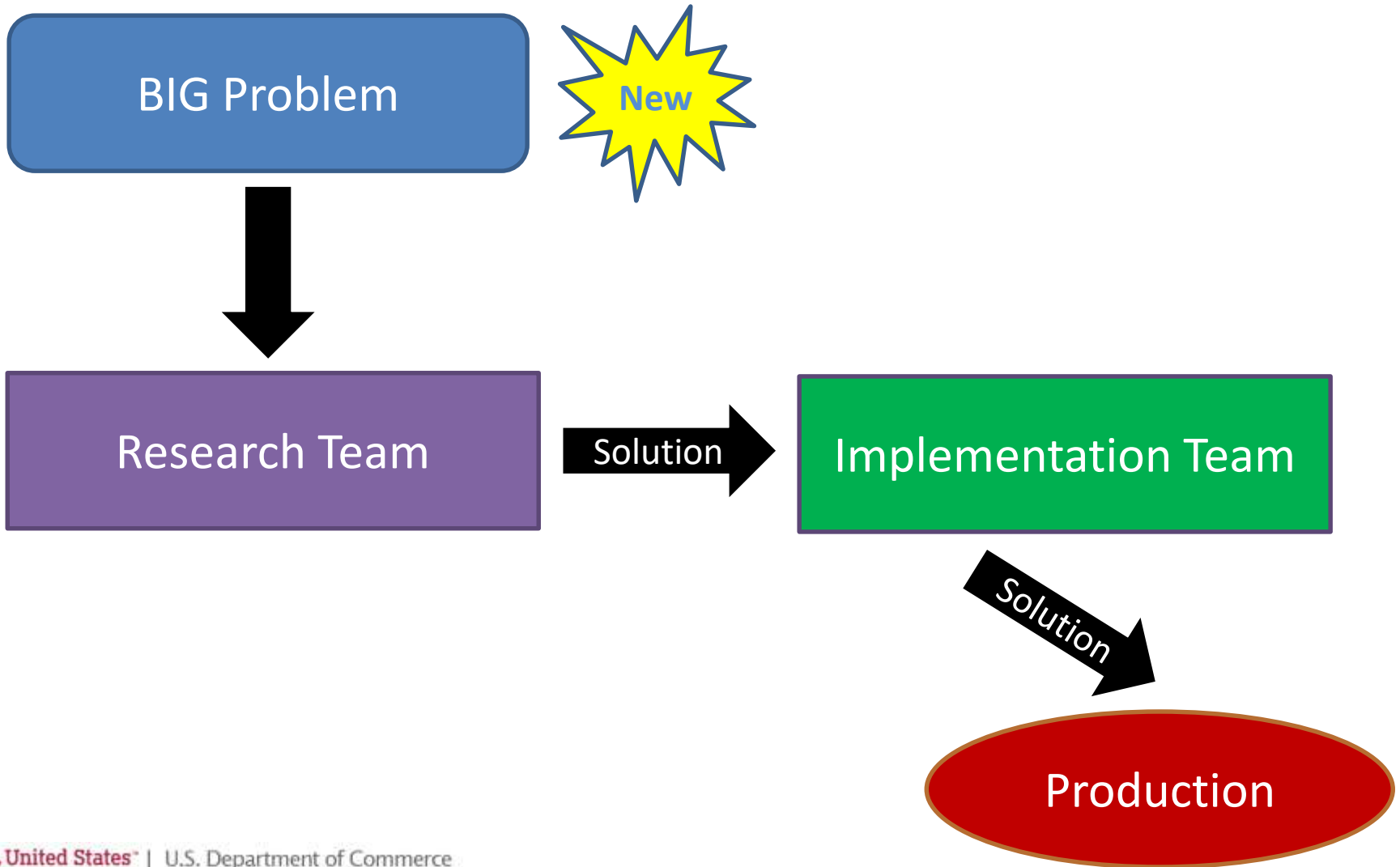
From Research to Implementation of Product Estimation in the 2017 Economic Census: Hard, Harder, and Hardest

**Federal Committee on Statistical Methodology
2018 Research and Policy Conference
March 9, 2018**

**William C. Davie Jr.
Economic Statistical Methods Division**

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

General Set Up



General Set Up

BIG Problem



New

Examples:

- Survey expansion
- Preliminary release
- Missing data treatment
- New survey questions

General Set Up

BIG Problem

New

HARD

General Set Up

BIG Problem



Research Team

- **“Dedicated” staff**
 - **Relevant job series**
 - **Differing experience levels**
- **Fixed duration**
- **Team responsibilities**
 - **Scope out problem**
 - **Obtain data**
 - **Conduct research**
 - **Provide “data-driven” recommendation**

General Set Up

BIG Problem

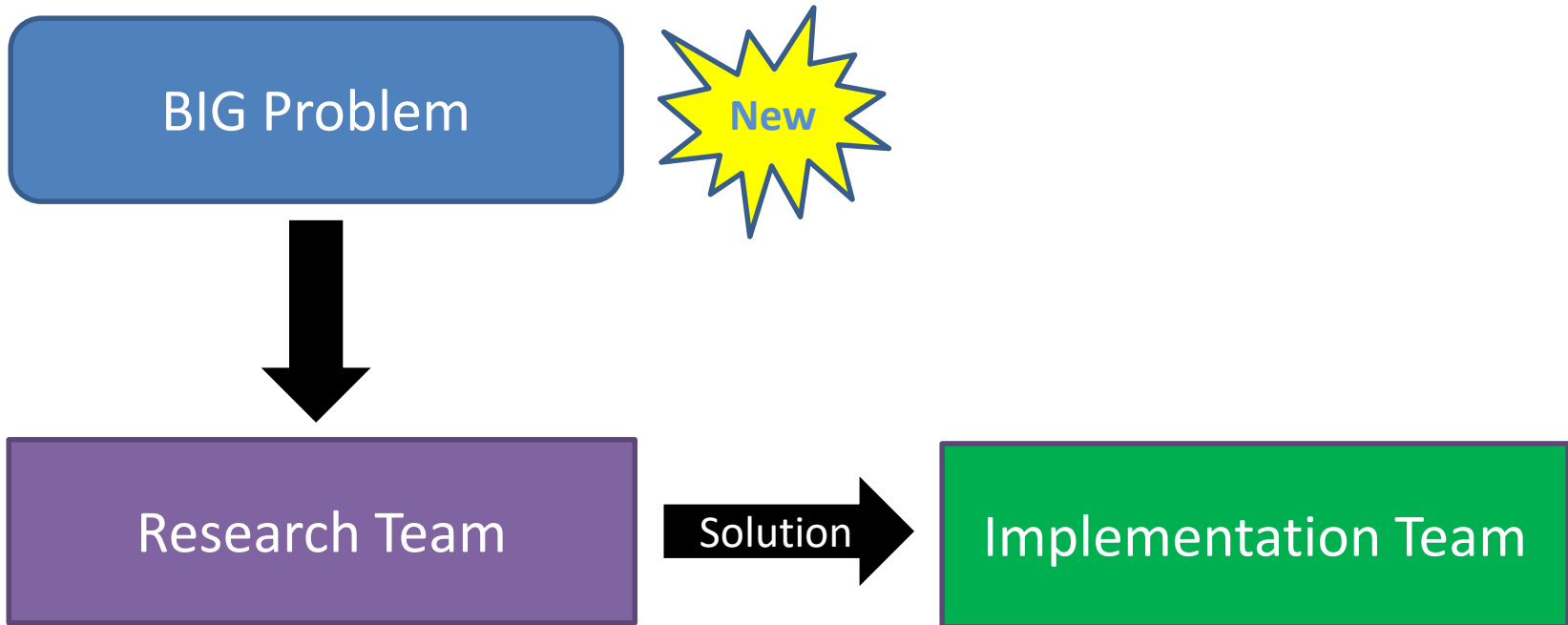
New



Research Team

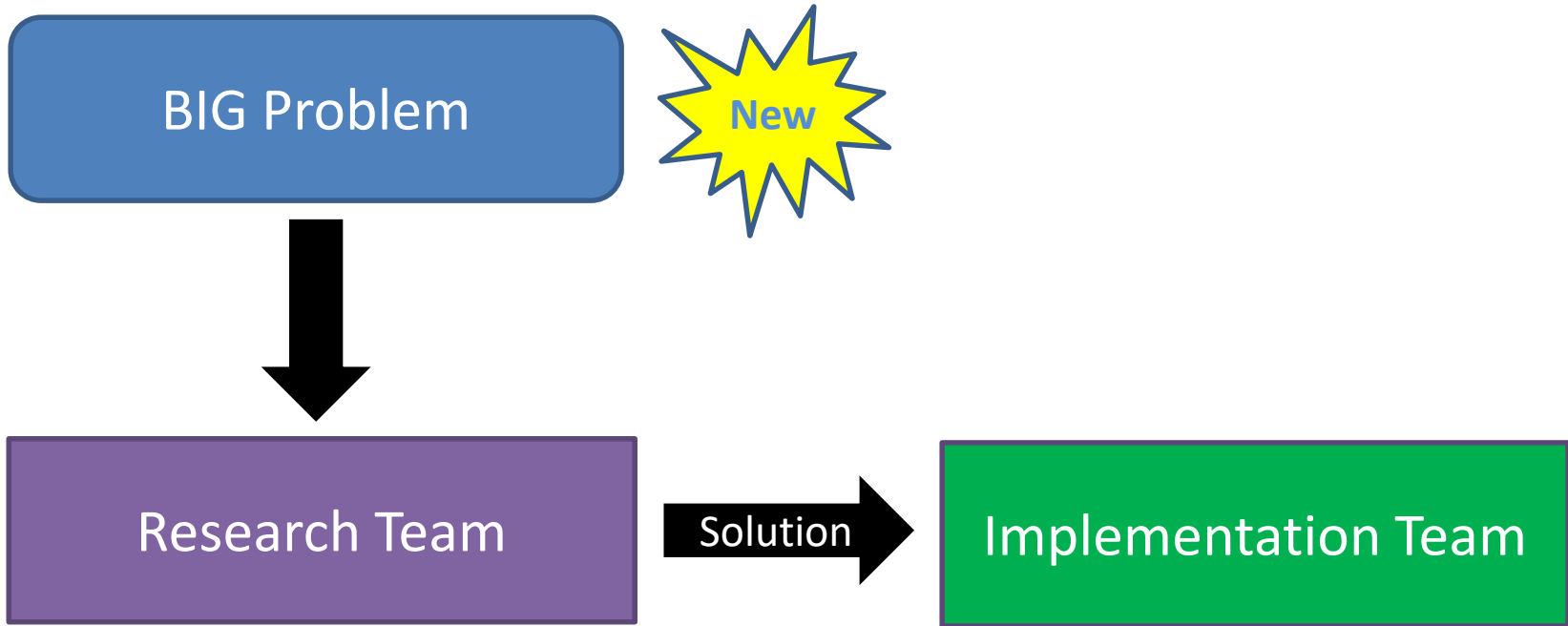
HARDER

General Set Up



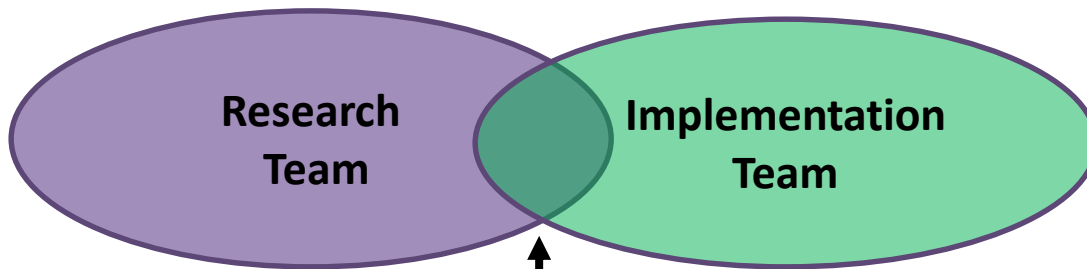
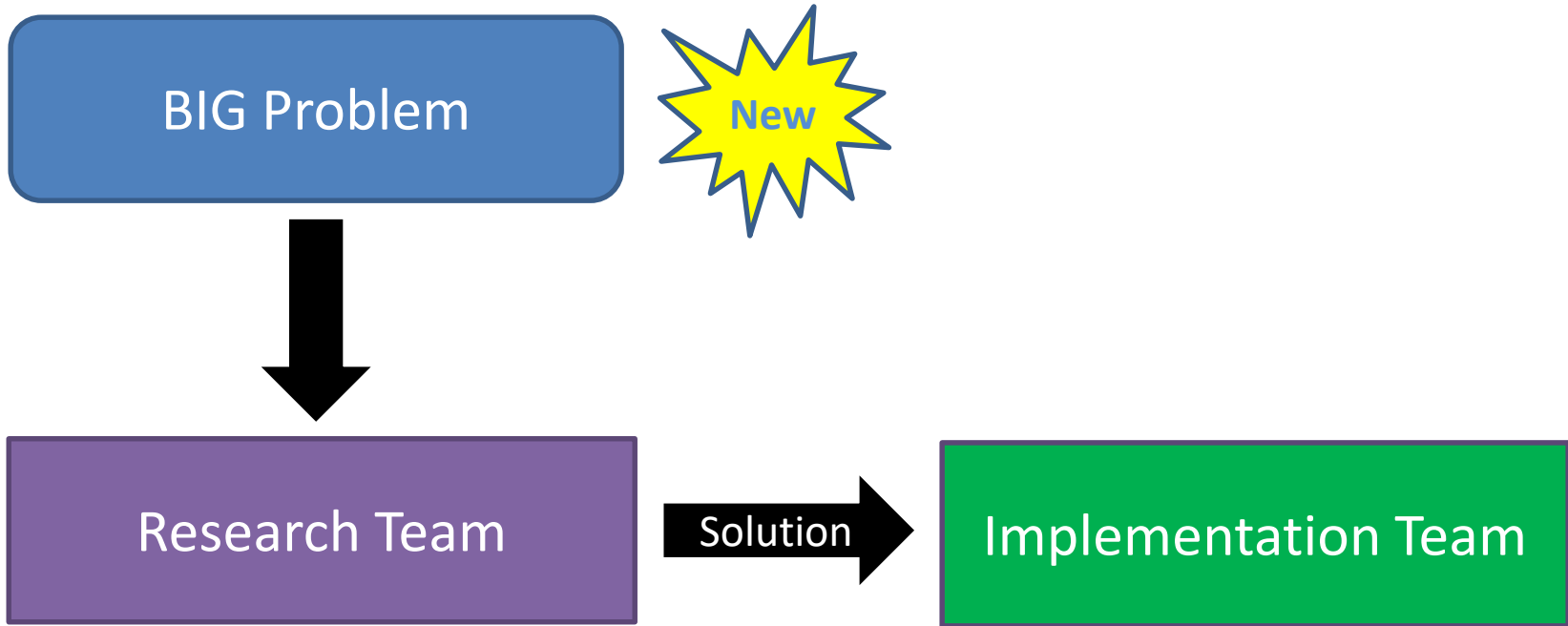
- Expert staff
- Fixed duration
- Team responsibilities
 - Address “ignored” issues from research
 - Write specifications
 - Logistics (coding, testing, validation)

General Set Up



HARDEST

General Set Up



↑
Common members (overlap)

Today's Discussion

Methods for treating missing product data



Research Team



Implementation Team



Production

Major Changes for the 2017 Economic Census

- North American Product Classification (NAPCS) - nearly 3,000 broad products and 5,000 detailed products
 - Missing data treatment problem
 - New data item(s) problem
 - New economy-wide product-by-industry table problem
- All electronic data collection
- Only one option for reporting product sales data – rounded to \$1,000

Example of Product Questions

- Retail Trade Industry – New Car Dealers

ITEM 22: DETAIL OF SALES, SHIPMENTS, RECEIPTS, OR REVENUE

Of the \$,000.00 of Sales, Shipments, Receipts, or Revenue reported in **Item 5**, what was the value for each product or service?

Description	Value
<p>1. Retail sales of automobiles and light-duty trucks (Include all outright sales plus transportation charges, dealer preparation charges, and dealer-installed options. Deduct discounts granted to the purchaser as an increase in trade-in allowance over fair market value and manufacturers' rebates and incentives.) (Report rental of automobiles and light-duty trucks on line 9, leasing of automobiles and light-duty trucks on line 10, and maintenance and repair services and the value of service contracts for automobiles and light-duty trucks on line 11. Report wholesale sales of new and used automobiles on line 3 and wholesale sales of new and used light-duty trucks on line 4.)</p>	
<p>a. Retail sales of new passenger cars (except fleet)</p>	\$ <input type="text"/> ,000.00
<p>b. Retail sales of new passenger cars -- fleet</p>	\$ <input type="text"/> ,000.00
<p>c. Retail sales of new vans and light-duty trucks, including minivans, cargo vans, sport utility vehicles (SUVs), and light passenger trucks (except fleet)</p>	\$ <input type="text"/> ,000.00
<p>d. Retail sales of new vans and light-duty trucks, including minivans, cargo vans, sport utility vehicles (SUVs), and light passenger trucks -- fleet</p>	\$ <input type="text"/> ,000.00
<p>e. Retail sales of used passenger cars (Include sales of passenger cars previously rented or leased.)</p>	\$ <input type="text"/> ,000.00
<p>f. Retail sales of used vans and light-duty trucks, including minivans, cargo vans, sport utility vehicles (SUVs), and light passenger trucks (Include sales of vans and light-duty trucks previously rented or leased.)</p>	\$ <input type="text"/> ,000.00
<p>g. Retail sales of all other powered transportation vehicles (Report new motorcycles, motor scooters, and motor bikes on line 2a; used motorcycles, motor scooters, and motor bikes on line 2b; utility trailers, boats, and other sports vehicles on line 7; and motor homes, travel trailers, and campers on line 8.)</p>	\$ <input type="text"/> ,000.00
<p>Subtotal</p>	\$ <input type="text"/> ,000.00

Detailed Products

Broad product subtotal

5. Retail sales of automotive parts, supplies, and accessories
 (Report parts installed in repair for automobiles and light-duty trucks on line 11.)

a. Retail sales of automotive lubricants, including oils, greases, etc.	\$	<input type="text"/>	,000.00
b. Retail sales of new automobile and light-duty truck tires and tubes	\$	<input type="text"/>	,000.00
c. Retail sales of new medium- and heavy-duty truck tires, including industrial, off-the-road, and farm tractor tires	\$	<input type="text"/>	,000.00
d. Retail sales of retreaded or used automobile and light-duty truck tires	\$	<input type="text"/>	,000.00
e. Retail sales of retreaded or used medium- and heavy-duty truck tires, including industrial, off-the-road, and farm tractor tires	\$	<input type="text"/>	,000.00
f. Retail sales of automotive parts, new and rebuilt, including wheels (except batteries)	\$	<input type="text"/>	,000.00
g. Retail sales of automotive parts, used, including wheels (except batteries)	\$	<input type="text"/>	,000.00
h. Retail sales of automotive batteries	\$	<input type="text"/>	,000.00
i. Retail sales of automotive accessories, including safety- and comfort-related items	\$	<input type="text"/>	,000.00
j. Retail sales of automotive supplies, including appearance and maintenance chemicals, automotive paint, antifreeze, functional fluids, etc. (Report automotive lubricants, including oils and greases, on line 5a.)	\$	<input type="text"/>	,000.00
k. Retail sales of automotive audio equipment, components, parts, and accessories (Include radios, stereos, compact disc players, mp3 players [audio only], and other sound reinforcement and recording equipment.)	\$	<input type="text"/>	,000.00
Subtotal		<input type="text"/>	,000.00





 22 / 30
 



14. All other products and services, not elsewhere classified

a. All other products and services, not elsewhere classified - write-in #1		
Pick one	Describe	\$,000.00
b. All other products and services, not elsewhere classified - write-in #2		
Pick one	Describe	\$,000.00
c. All other products and services, not elsewhere classified - write-in #3		
Pick one	Describe	\$,000.00
<input type="button" value="Add Additional Products"/>		
TOTAL - Sum of lines should equal total Sales, Shipments, Receipts, or Revenue reported in Item 5		\$,000.00

Multiple blank spaces to write-in products not listed

TOTAL

Notes:

- Respondents can “write-in” products that are not pre-listed
- All **detail** product values should sum to their associated broad product value
- All **broad** product values should sum to the reported total value of sales (within a tolerance) for the establishment

Product Data

Collection Rules

- Different products collected by industry
 - Same (1-5) products frequently reported within a given industry
 - Reported in \$1000 (in 2017)
- Reported products must add to total sales for establishment after editing/imputation

Characteristics

- Low response rates
- Limited availability of predictors
- No auxiliary or administrative data readily available for imputation

Missing Data Treatment

Research Team

- \approx 1,000 industries
- \approx 8,000 products
 - Broad products
 - Detail products
- Sample data (i.e., sampling weights)
- Imputation cells
 - Cell collapsing rules
 - Minimum number of donors
- Restrictions on value ($>$ \$1,000)

Research Challenges

- Dedicated Team
 - Short time frame (\approx 6-9 months)
- Magnitude of the problem
 - \approx 1,000 industries and \approx 8,000 products
- Historical data limitations
 - Classification differences (to NAPCS)
 - Collection differences (to electronic)
 - Unit collection differences (from varied to \$1,000)

Research Evaluation

- Perform a simulation study
- Two evaluation statistics
 - Imputation error
 - Fraction of Missing Information (FMI)
- Recommendation: **Hot Deck**
 - Nearest neighbor or random

Research Decisions

- Restrict research to about 25 industries with common products under NAPCS
- Consider only **broad** products
- Develop imputation cells and collapsing rules that did not adequately consider industry- or geography-specific differences
- “Top ten” products per industry

Missing Data Treatment

Research Team

- \approx 1,000 industries
- \approx 8,000 products
 - Broad products
 - Detail products
- Sample data (i.e., sampling weights)
- Imputation cells
 - Cell collapsing rules
 - Minimum number of donors
- Restrictions on value ($>$ \$1,000)

Missing Data Treatment

Research Team

- ~~1,000~~ 25 industries
- ~~8,000~~ Top 10 products
 - Broad products
 - ~~Detail products~~
- ~~Sample data (i.e., sampling weights)~~
- LOCAL Imputation cells
 - Cell collapsing rules
 - Minimum number of donors
- ~~Restrictions on value (\rightarrow \$1,000)~~

Missing Data Treatment

Research Team

- ~~1,000~~ 25 industries
- ~~8,000~~ Top 10 products
 - Broad products
 - ~~Detail products~~
- ~~Sample data (i.e., sampling weights)~~
- LOCAL Imputation cells
 - Cell collapsing rules
 - Minimum number of donors
- ~~Restrictions on value (\rightarrow \$1,000)~~

Implementation Team

- \approx 1,000 industries
- \approx 8,000 products
 - Broad products
 - Detail products
- Sample data (i.e., sampling weights)
- Imputation cells
 - Cell collapsing rules
 - Minimum number of donors
- Restrictions on value ($>$ \$1,000)
- “Must” products for industries
- Choice of hot deck method by industry
- Maximizing use of reported data
- Backup methods
- Processing time ...

Missing Data Treatment

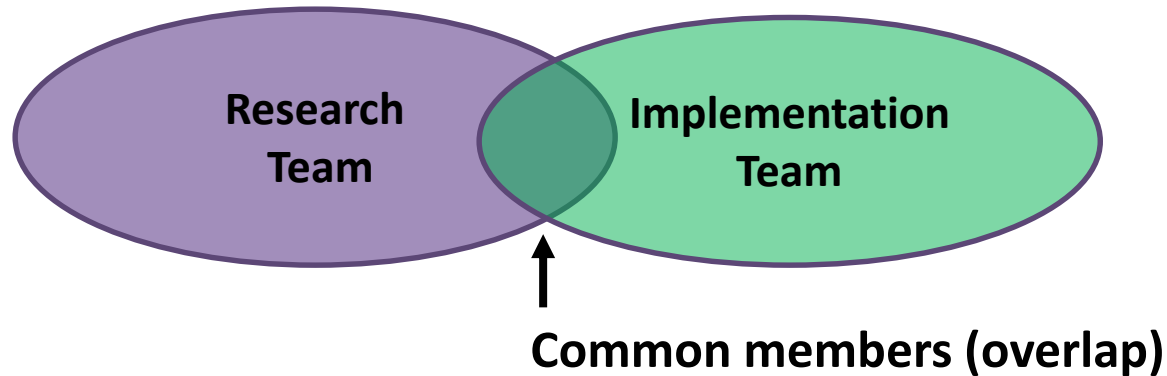
Research Team

- ~~1,000~~ 25 industries
- ~~8,000~~ Top 10 products
 - Broad products
 - ~~Detail products~~
- ~~Sample data (i.e., sampling weights)~~
- LOCAL Imputation cells
 - Cell collapsing rules
 - Minimum number of donors
- ~~Restrictions on value (\rightarrow \$1,000)~~

Implementation Team

- \approx 1,000 industries
- \approx 8,000 products
 - **Broad products**
 - **Detail products**
- Sample data (i.e., sampling weights)
- Imputation cells
 - Cell collapsing rules
 - Minimum number of donors
- Restrictions on value ($>$ \$1,000)
- “Must” products for industries
- Choice of hot deck method by industry
- **Maximizing use of reported data**
- Backup methods
- **Processing time ...**

Implementation Team



- Overlap: Team leader and consultants
- New members
 - Subject Matter Experts
 - Programmers
 - Methodologists

Implementation Decisions

- Group decision (“consensus”)
 - Support by previous census counts when possible
 - Take subject-matter experience/expertise into account
- Testing methods
 - Test deck of 2.4 million donor records (with over 20 million products) and 1.1 million full recipients
 - Independent validation of production code

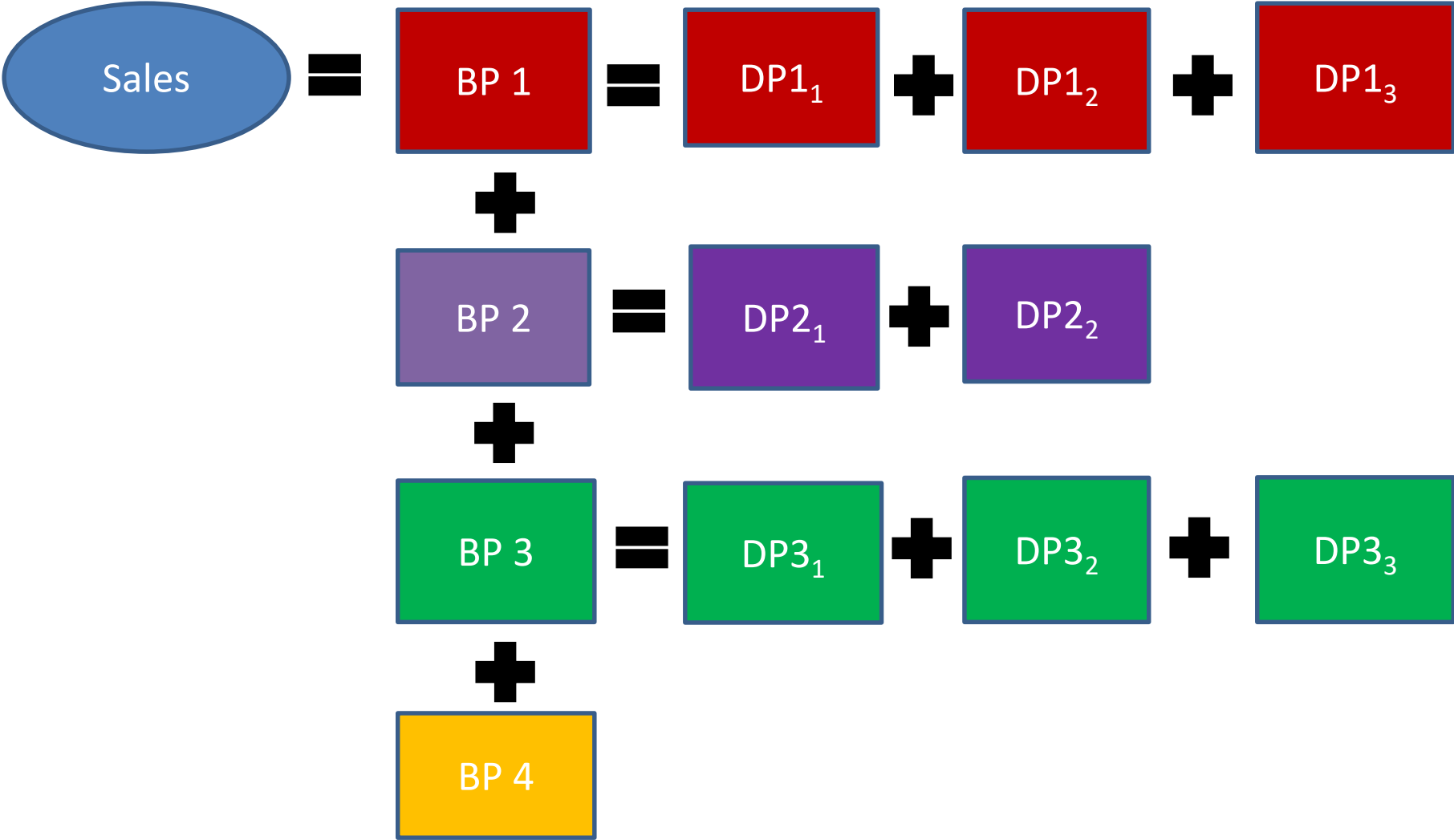
Imputation of Detailed Products

- Businesses more likely to report broad products than detailed products
- Subject matter experts wanted to maximize use of valid reported data in imputation
- Implementation team decision
 - Group establishments into types
 - Use category average imputation
 - Statistical model frequently used for business data
 - NOT part of the research for product data
 - “Hot deck” imputation – all products (broad & detail)

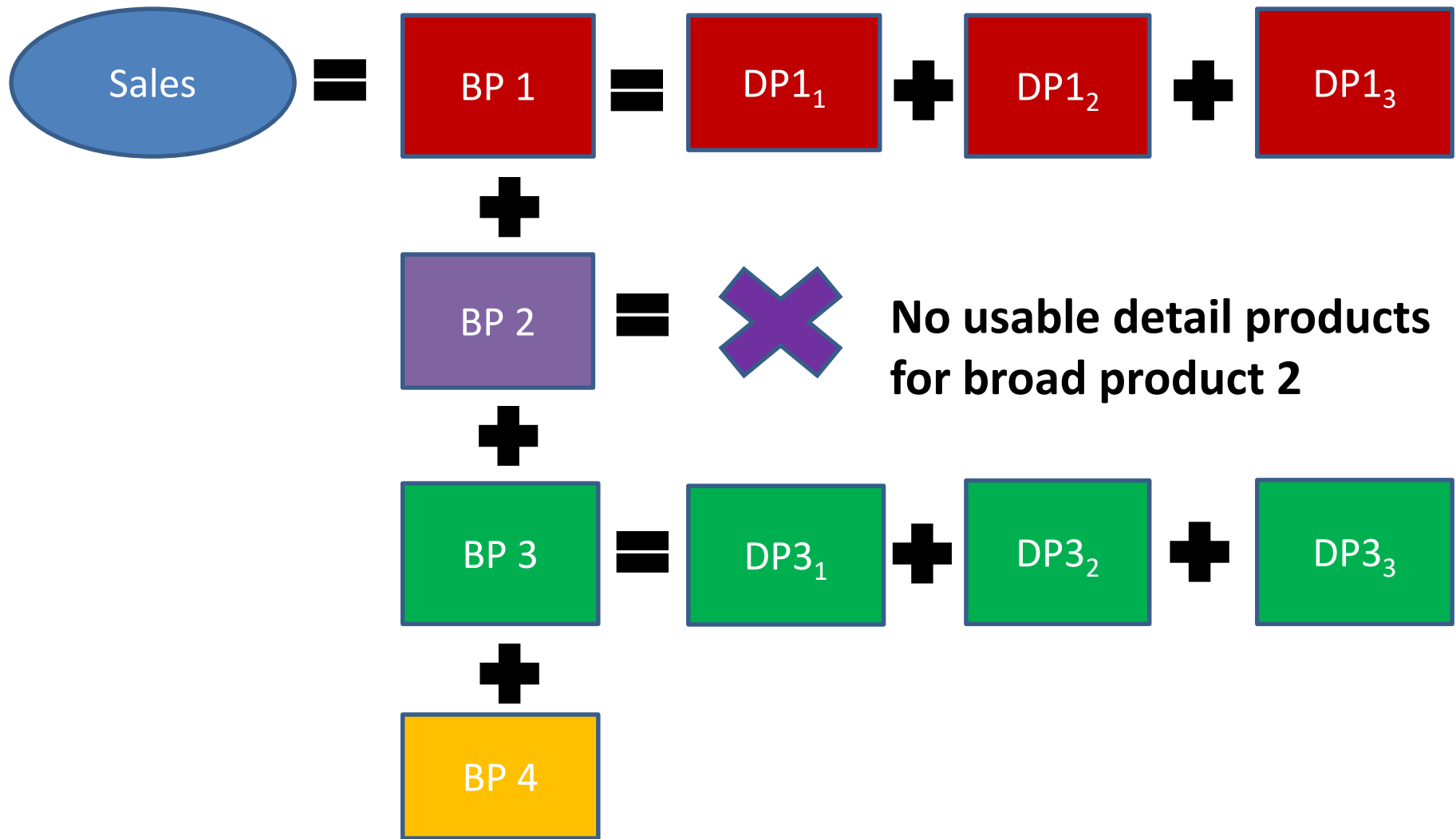
Establishment Groups

Donors	Broad products usable
Complete	All required Detail products usable and in balance to Broad products
Partial	Some usable Detail products
Minimal	No usable Detail products
Recipients	Missing products
Full	Need Broad products and Detail products
Partial	Need some (designated) Detail products
Minimal	Need all designated Detail products
Ineligible	All products usable, but not “typical”; excluded from donor pool

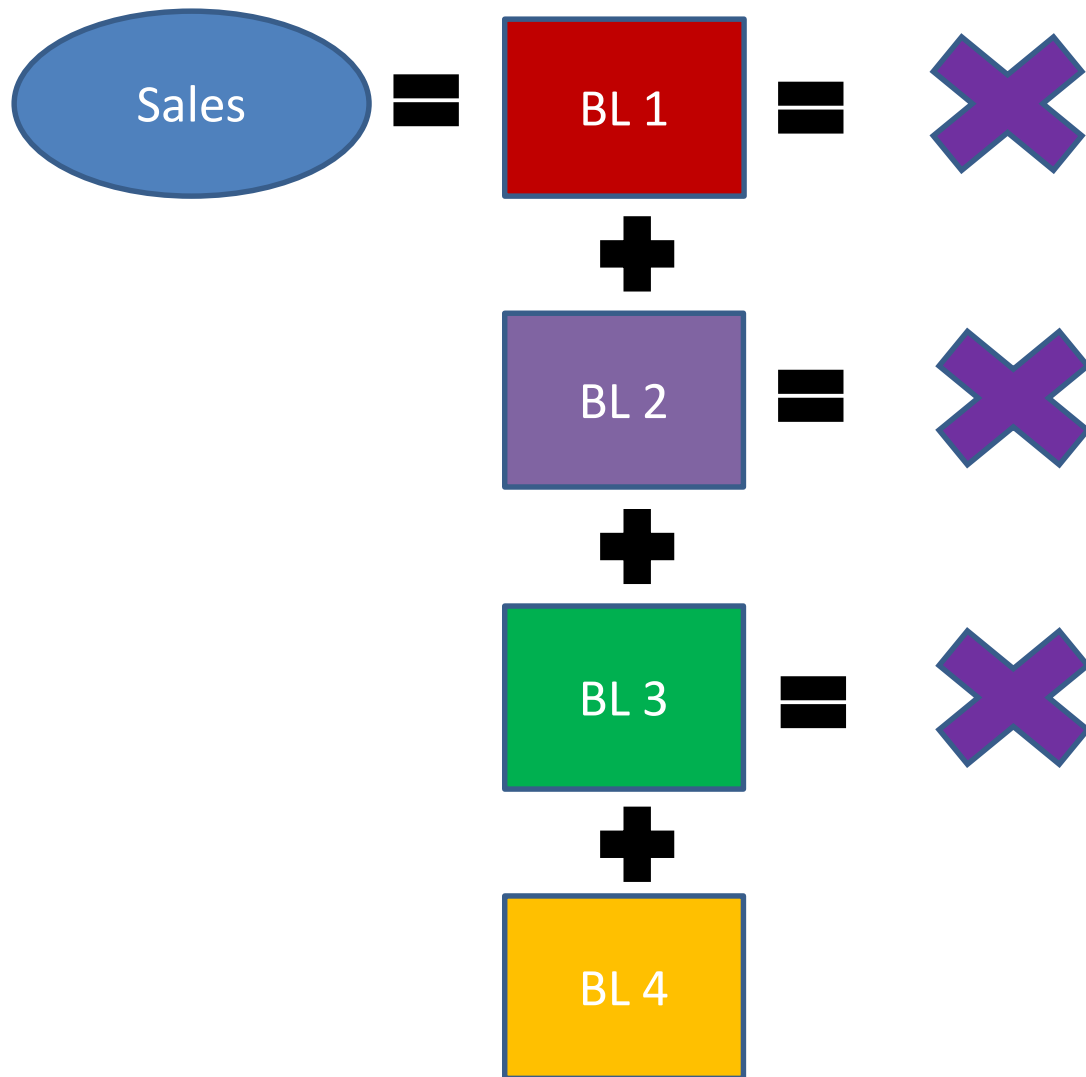
Complete Donor Example



Partial Donor Example



Miminal Donor Example



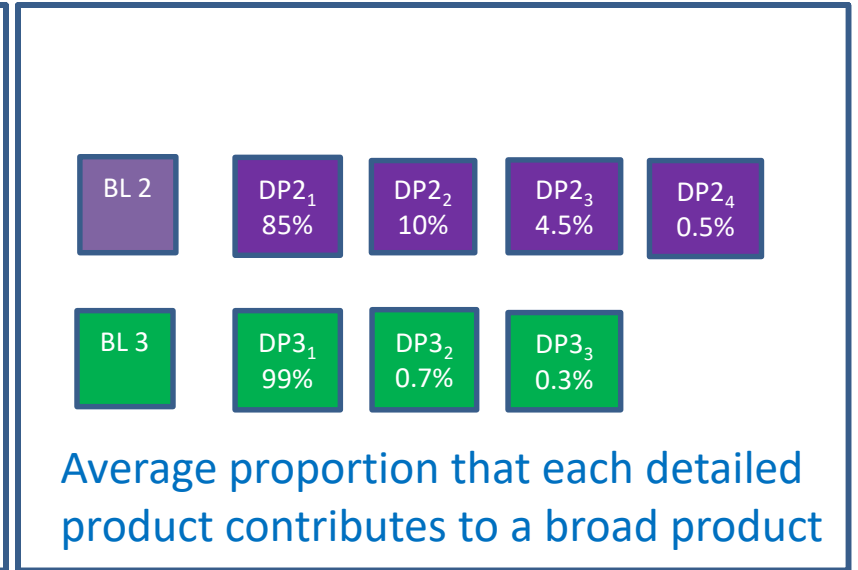
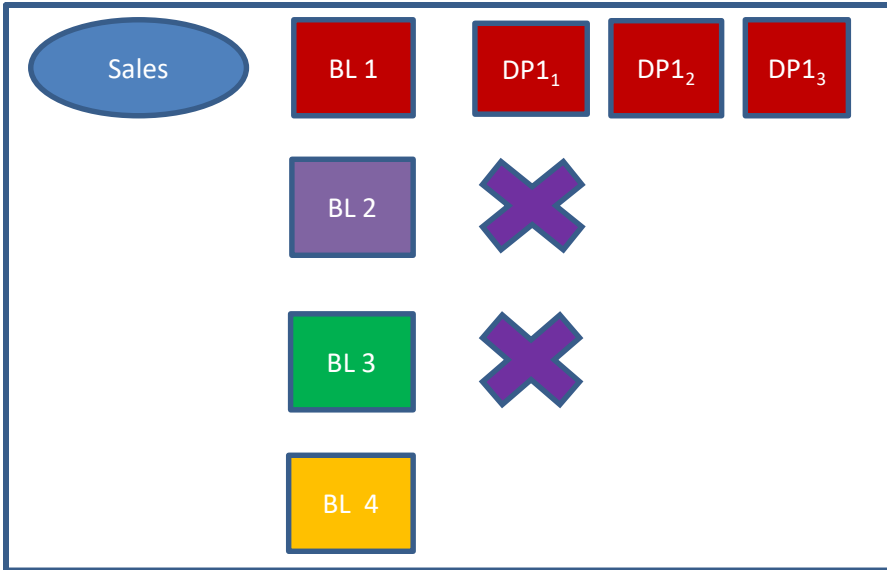
**No usable detail products
for any broad product**

“Completing” Partial Donors

Missing detailed products for BL 2 and 3

Category Averages for BL 2 and 3

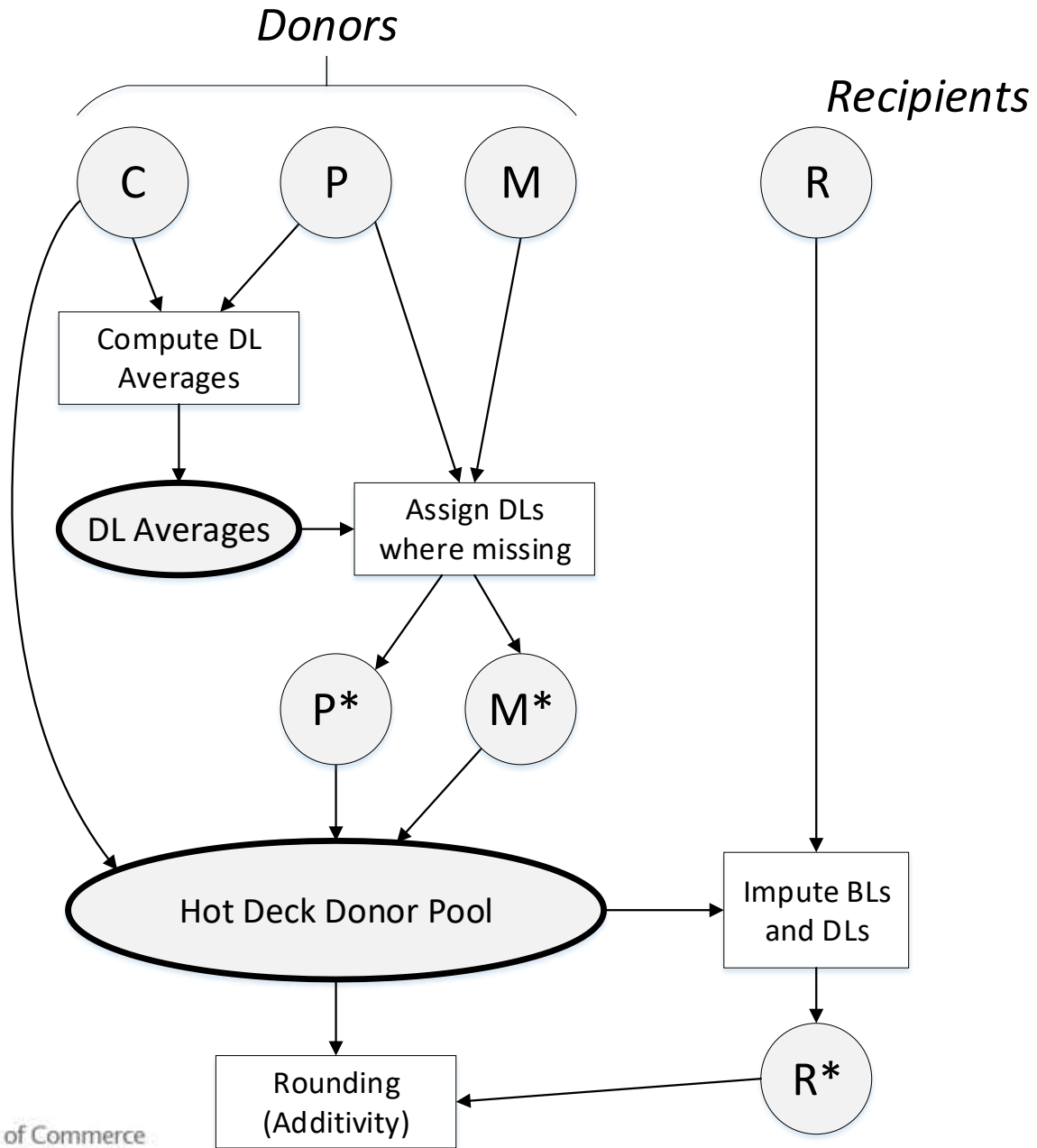
I
n
p
u
t
s



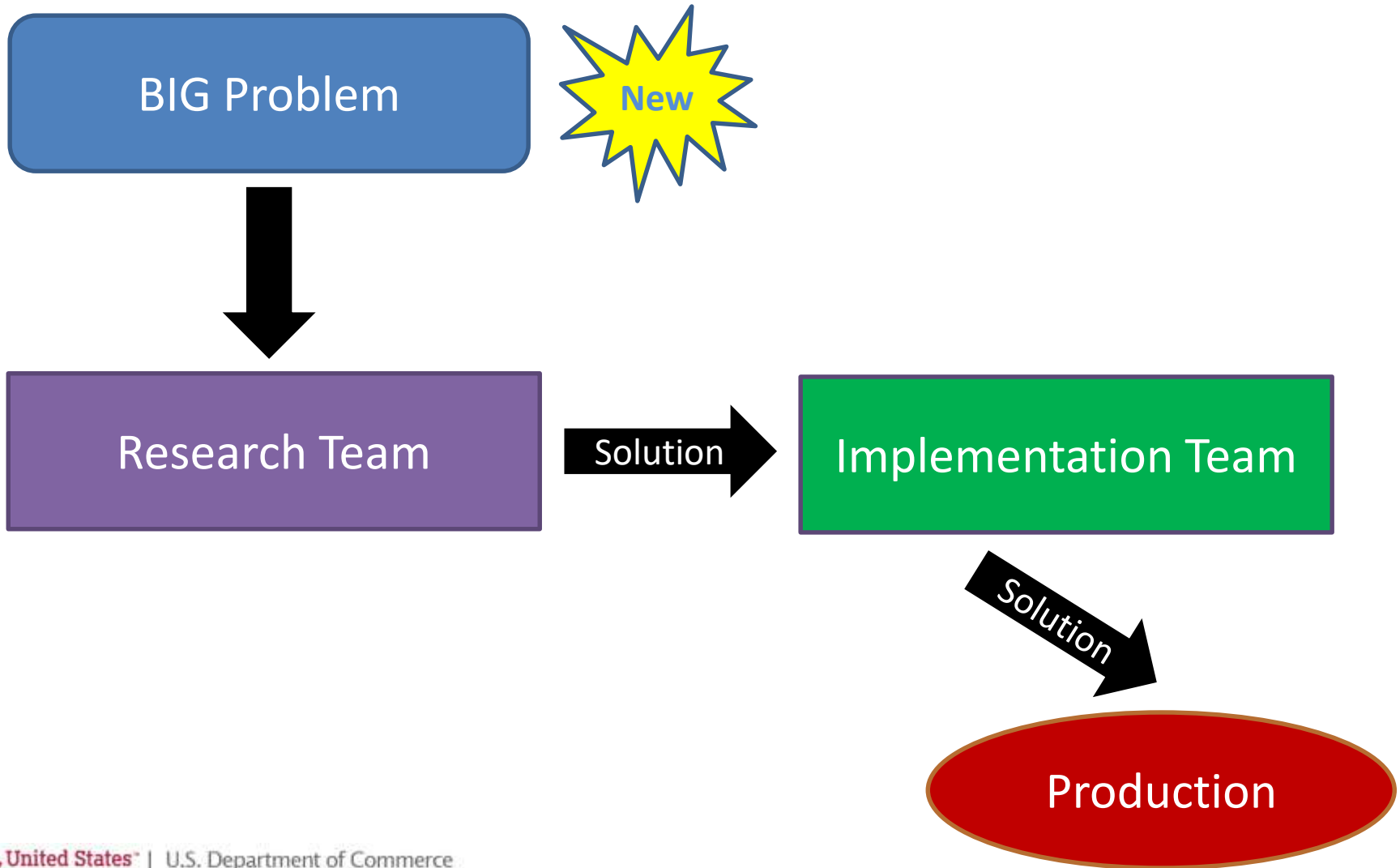
O
u
t
p
u
t



Completed record is now a donor



Going Back to Original Picture



Drawbacks

- Limited scope in research
 - Missed requirements
 - Unexpected results in implementation testing
- Delayed decisions
 - No time for detailed investigation
 - Quick decisions made for convenience
 - Anecdotal reasons
 - No alternatives tested
- Implications on other statistical processes

Benefits

- Workable solution
- Buy-in on methods
- Broadened knowledge
- Research-based decision
- Ideas for 2022

Thank you

Contacts:

William.C.Davie.Jr@census.gov

Katherine.J.Thompson@census.gov

Scot.A.Dahl@census.gov

References (Previous Research)

- Bechtel, L., Morris, D.S., and Thompson, K.J. (2015). Using Classification Trees to Recommend Hot Deck Imputation Methods: A Case Study. *Proceedings of the FCSM Research Conference*.
- Ellis, Y. and Thompson, K.J. (2015). Exploratory Data Analysis of Economic Census Products: Methods and Results. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Garcia, M., Morris, D.S., and Diamond, L.K. (2015). Implementation of Ratio Imputation and Sequential Regression Multivariate Imputation on Economic Census Products. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Knutson, J. and Martin, J. (2015). Evaluation of Alternative Imputation Methods for Economic Census Products: The Cook-Off. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Thompson, K.J. and Liu, X. (2015). On Recommending a Single Imputation Method for Economic Census Products. *Proceedings of the Government Statistics Section*, American Statistical Association.
- Tolliver, K. and Bechtel, L. (2015). Implementation of Hot Deck Imputation on Economic Census Products. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

References (Previous Research)

- Knutson, J., Thompson, M., and Thompson, K.J. (2017). Developing Variance Estimates for Products in the 2017 Economic Census. *Proceedings of the Government Statistics Section*, American Statistical Association.
- Thompson, K.J. and Thompson, M. (2016). Estimating the Variance Due to Hot Deck Imputation for Product Value Estimates in the 2017 Economic Census. *Proceedings of the Government Statistics Section*, American Statistical Association.
- Thompson, M., Thompson, K.J., and Kurec, R. (2016). Variance Estimation for Product Value Estimates in the 2017 Economic Census Under the Assumption of Complete Response. *Proceedings of the Government Statistics Section*, American Statistical Association.
- Fink, E.B., Beck, J.L. and Willimack, D.K. (2015). Data-Driven Decision Making and the Design of Economic Census Data Collection Instruments. Proceedings of the FCSM Research Conference.