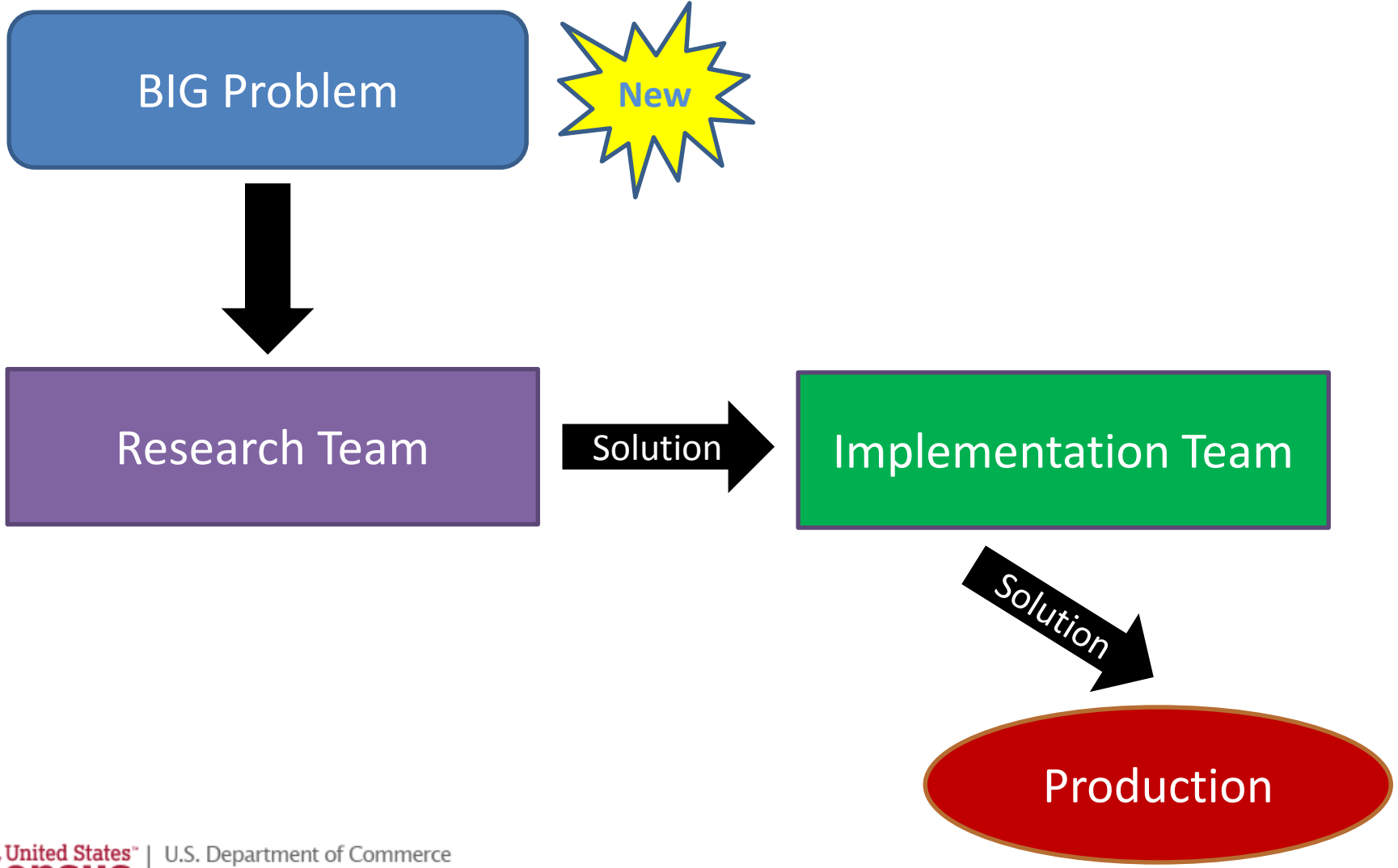# On Implementing a New Imputation Method into Production in the 2017 Economic Census Illustrated through Selected Vignettes
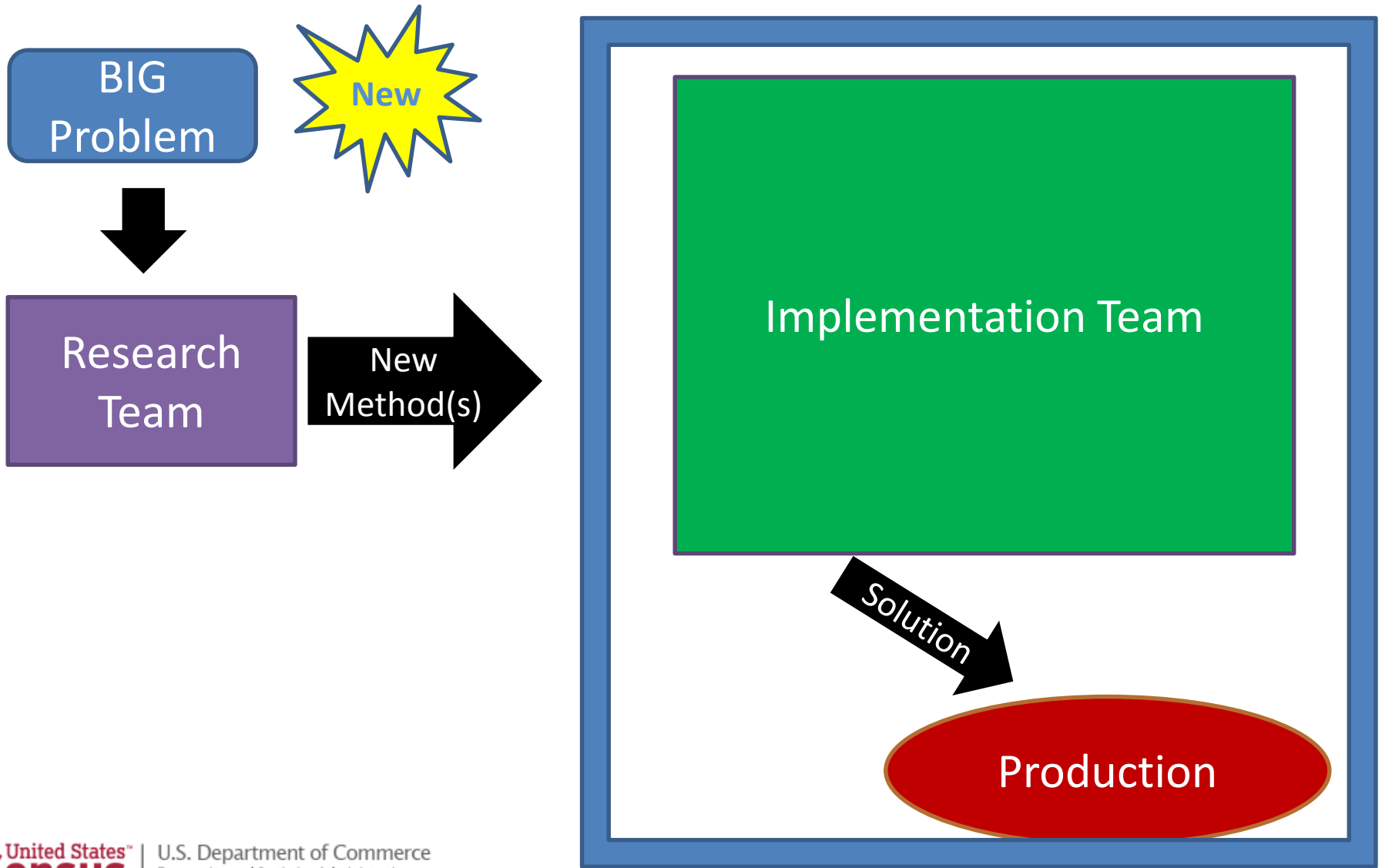
Katherine Jenny Thompson and Laura Bechtel
Economic Statistical Methods Division

General Set Up

The Next Three Presentations

BIG Problem

New

Research Team

New Method(s)

Implementation Team

Solution

Production

The Next Three Presentations

BIG Problem

New

Research Team

New Method(s)

Implementation Team

Additional Requirements

Existing Processes

Integration challenges

Storage
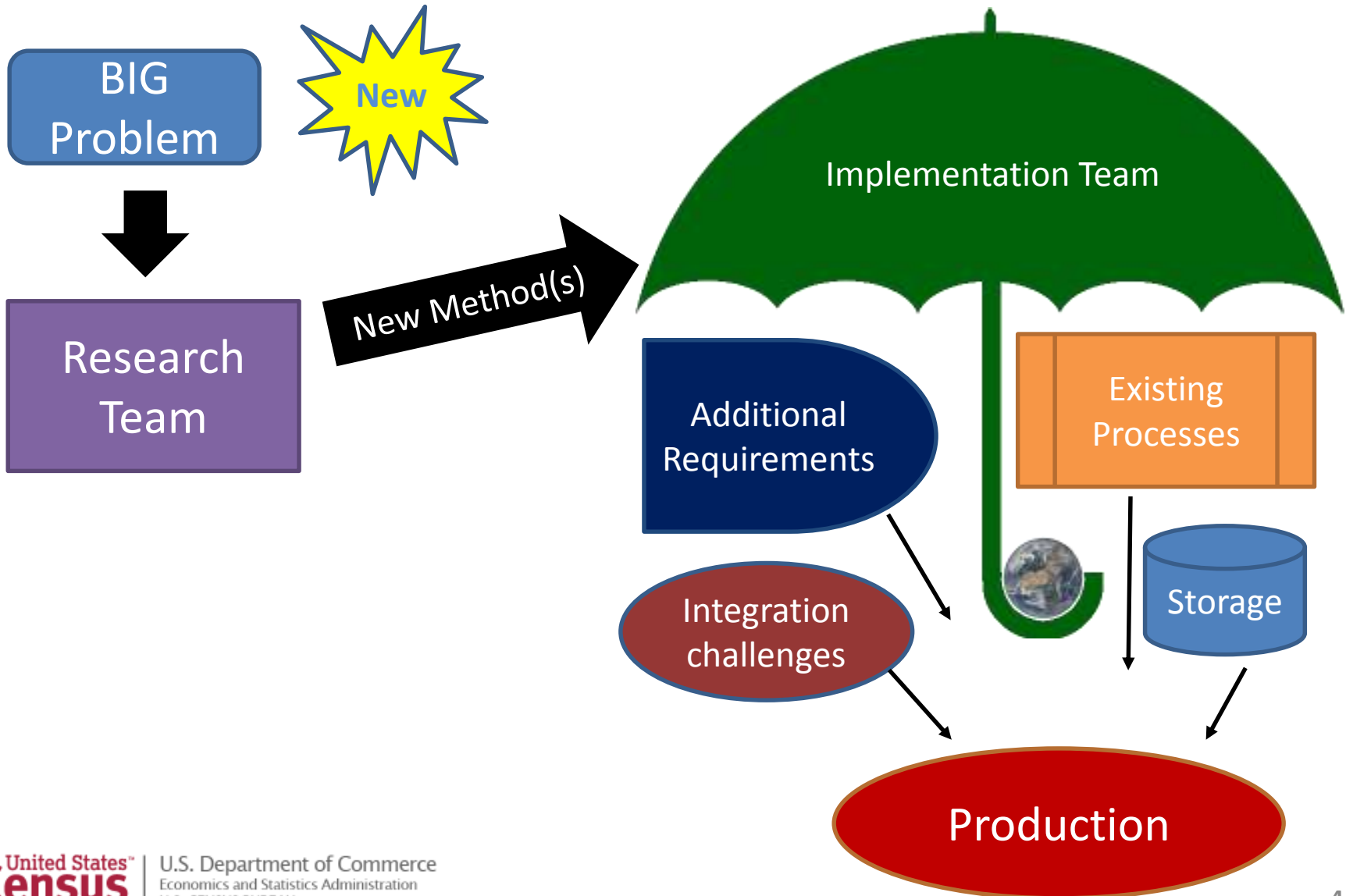
Production

My Discussion

Methods for treating missing product data

New

Research Team

Solution

Implementation Team

Solution

Production

Major Changes for the 2017 Economic Census

- North American Product Classification (NAPCS) - nearly 3,000 broad products and 5,000 detail 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

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

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
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.)	
a. Retail sales of new passenger cars (except fleet)	$,000.00
b. Retail sales of new passenger cars -- fleet	$,000.00
c. Retail sales of new vans and light-duty trucks, including minivans, cargo vans, sport utility vehicles (SUVs), and light passenger trucks (except fleet)	$,000.00
d. Retail sales of new vans and light-duty trucks, including minivans, cargo vans, sport utility vehicles (SUVs), and light passenger trucks -- fleet	$,000.00
e. Retail sales of used passenger cars (Include sales of passenger cars previously rented or leased.)	$,000.00
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.)	$,000.00
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.)	$,000.00
Subtotal	$,000.00

Detail 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.

$ _____ ,000.00

b. Retail sales of new automobile and light-duty truck tires and tubes

$ _____ ,000.00

c. Retail sales of new medium- and heavy-duty truck tires, including industrial, off-the-road, and farm tractor tires

$ _____ ,000.00

d. Retail sales of retreaded or used automobile and light-duty truck tires

$ _____ ,000.00

e. Retail sales of retreaded or used medium- and heavy-duty truck tires, including industrial, off-the-road, and farm tractor tires

$ _____ ,000.00

f. Retail sales of automotive parts, new and rebuilt, including wheels (except batteries)

$ _____ ,000.00

g. Retail sales of automotive parts, used, including wheels (except batteries)

$ _____ ,000.00

h. Retail sales of automotive batteries

$ _____ ,000.00

i. Retail sales of automotive accessories, including safety- and comfort-related items

$ _____ ,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.)

$ _____ ,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.)

$ _____ ,000.00

Subtotal

$ _____ ,000.00

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 |

Add Additional Products

TOTAL - Sum of lines should equal total Sales, Shipments, Receipts, or Revenue reported in Item 5 $,000.00

TOTAL

Multiple blank spaces to write-in products not listed

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

Missing Data Treatment

Research Team

- ≈ 1,000 industries
- ≈ 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

- Team Composition (next slide)

- Short time frame ($\approx$ 6-9 months)

- Magnitude of the problem
  - $\approx$ 1,000 industries and $\approx$ 8,000 products

- Historical data limitations (new collection)

Research Team Composition

- Methodologists
  - Very limited experience with Economic Census
  - Even less experience with "products"

- Subject matter experts
  - Very limited experience with imputation methodology

- NO Programmers/IT Specialists

Missing Data Treatment

Research Team

- ≈ 1,000 industries
- ≈ 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 (> $1,000)~~

Quick Summary

- Research conducted under restricted conditions

- SAS code developed in-house
  - Not ready for prime time!

- Recommendation: **Hot Deck**
  - Nearest neighbor or random
  - Unaddressed production requirements

Implementation Team

Research Team — Implementation Team

↑

Common members (overlap)

- Overlap
  - Team leader: Subject matter expert
  - Consultants: Methodologists (4)
- New members
  - Subject Matter Experts
  - Programmers
  - Methodologists (1 new)

United States™ Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

17

Implementation Team

Research Team

Implementation Team

Common members (overlap)

Implementation Team

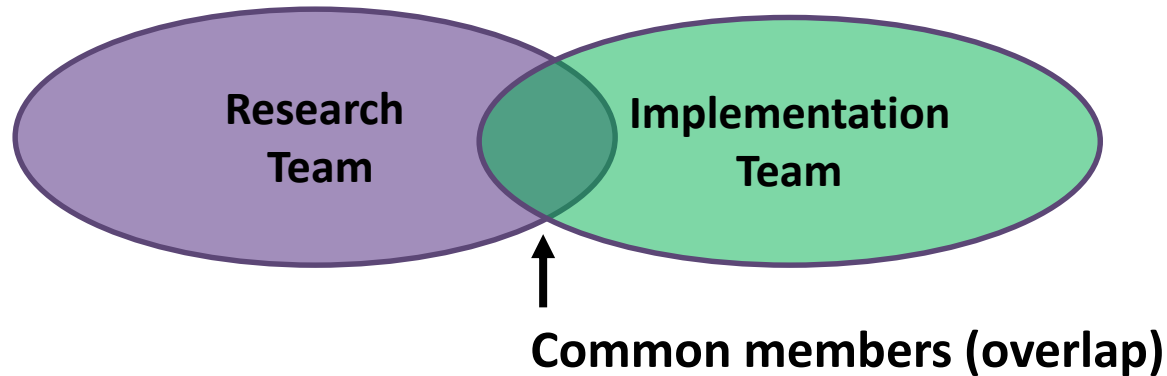| Topic	Team Members	Knowledge Level
Economic Census Processing	Subject Matter	Expert
	Methodologists	Varied
	Programmers/IT	Expert
Hot Deck Imputation	Subject Matter	Low
	Methodologists	Expert
	Programmers/IT	Low

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 (> $1,000)~~

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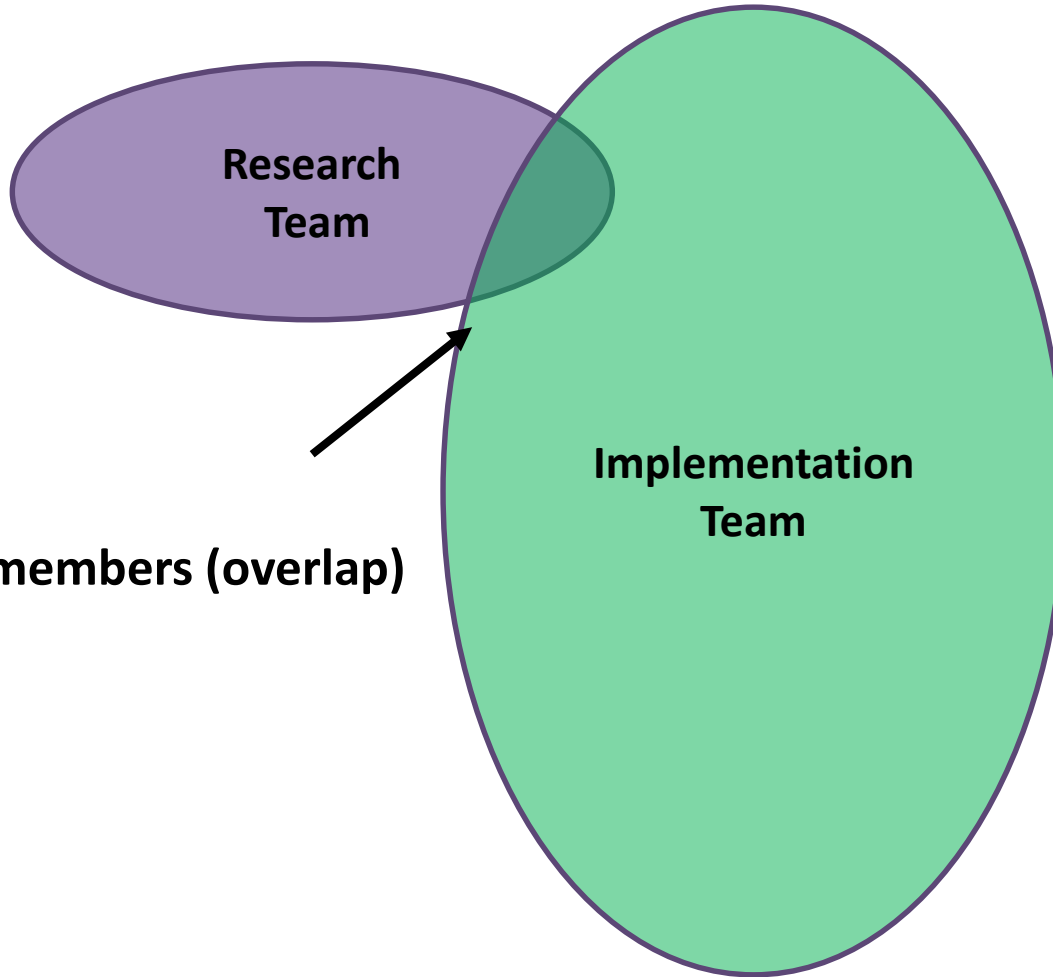
- ≈ 1,000 industries
- ≈ 8,000 products
  - Broad products
  - Detail products
- Sample data (i.e., sampling weights)
- Imputation cells
  - Cell collapsing rules
  - Minimum number of donors
- Restrictions on value (> $1,000)

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 (> $1,000)~~

Implementation Team

- ≈ 1,000 industries
- ≈ 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 (> $1,000)~~

Implementation Team

- ≈ 1,000 industries
- ≈ 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 …**

Vignettes

1. Processing time

2. Imputation Cell Collapsing/Minimal Donor

3. Imputation for Detail Products

Vignette 1: Processing Time

- Programmer concerns
  - Prohibitive processing time
  - Complex coding
- Addressed by
  - 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
  - Collaborative development of specifications

Vignette 2: Imputation Cells

- Needed
  - Imputation cell definitions
  - Collapsing rules
  - Minimum donors

- Limited research

- Education Process

Simple Example

- Cell Collapsing
  - Ideal = Industry By Tax Status By Unit Type
  - Collapse 1 = Industry By Tax Status
    - DROP Unit Type
  - Base = Industry
    - DROP Unit Type and Tax Status

- Minimum cell count = 5
  - Base = 1

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Compute Donor Counts
Ideal Cells

Industry	Tax Exempt Status	Unit Type	# of Establishments
AAAAAA	Taxable	SU	2
	Taxable	MU	2
	Exempt	SU	2
	Exempt	MU	2
BBBBBB	Taxable	SU	23
	Taxable	MU	35
	Exempt	SU	2
	Exempt	MU	4
CCCCCC	Taxable	SU	10
	Taxable	MU	3
	Exempt	SU	200
	Exempt	MU	2

United States™ Census Bureau | U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

27

Determine Usage of Ideal Cells

Industry	Tax Exempt Status	Unit Type	# of Establishments
AAAAAA	Taxable	SU	2
	Taxable	MU	2
	Exempt	SU	2
	Exempt	MU	2
BBBBBB	Taxable	SU	23
	Taxable	MU	35
	Exempt	SU	2
	Exempt	MU	4
CCCCCC	Taxable	SU	10
	Taxable	MU	3
	Exempt	SU	200
	Exempt	MU	2

Less than 5 Establishments $\Rightarrow$ Collapse

Compute Donor Counts for Collapse 1 Cells

Industry	Tax Exempt Status	# of Establishments
AAAAAA	Taxable	4
	Exempt	4
BBBBBB	Taxable	58
	Exempt	6
CCCCCC	Taxable	13
	Exempt	202

STILL Less than 5 Establishments ⟹ Collapse

All Counts ≥ 5 Establishments ⟹ Use Collapse 1 cells

Compute Donor Counts for Base Cells

Industry	# of Establishments
AAAAAA	8
BBBBBB	
CCCCCC	

All Counts ≥ 1
Establishments $\Rightarrow$
Use base cells

Final Cells for Hot Deck

Industry	Tax Exempt Status	Unit Type	Hot Deck Cell
AAAAAA	ALL	ALL	Base
BBBBBB	Taxable	SU	Ideal
	Taxable	MU	Ideal
	Exempt	ALL	Collapse 1
CCCCCC	Taxable	ALL	Collapse 1
	Exempt	ALL	Collapse 1

Final Cells for Hot Deck

Industry	Tax Exempt Status	Unit Type	Hot Deck Cell
AAAAAA	ALL	ALL	Base
BBBBBB	Taxable	SU	Ideal
	Taxable	MU	Ideal
	Exempt	ALL	Collapse 1
CCCCCC	Taxable	ALL	Collapse 1
			Collapse 1

The Contention Point

Industry	Tax Exempt Status	Unit Type	# of Establishments
CCCCCC	Taxable	SU	10
	Taxable	MU	3
	Exempt	SU	200
	Exempt	MU	2

Subject Matter Experts and Programmers Contention
- NO need to collapse for SU unit type
- "NOT FAIR"

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Alternate Proposal

Industry	Tax Exempt Status	Unit Type	Imputation Cell	# of Establishments
CCCCCC	Taxable	SU	Ideal	10
	Taxable	MU	Collapse 1	13
	Exempt	SU	Ideal	200
	Exempt	MU	Collapse 1	202

Methodology Concerns (Severe)

- Imputation cells no longer disjoint
  - Affects variance estimation

- Hurts probability of selecting MU donors

Impasse?
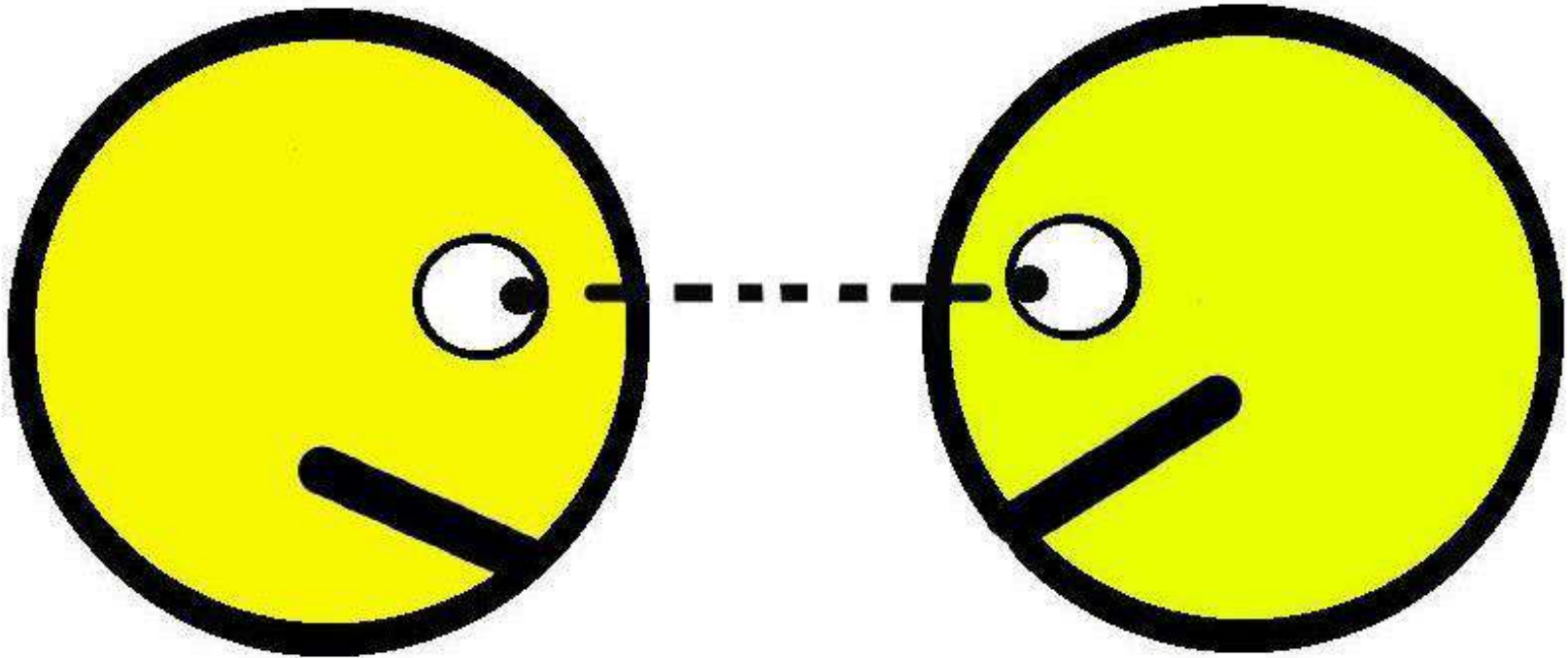
FOR ALTERNATIVE
Subject matter experts
Programmers/IT specialists
Methodologists (2.5)

AGAINST ALTERNATIVE
Methodologists (2.5)

Compromise/Decision

- Alternative Proposal

- Minimum number of donors = 1

- Lessons learned:
  - Put in measures to avoid unacceptable compromises
  - Include programmers and subject matter experts at the research stage

Vignette 3: Detail Products

- Not considered during research stage

- Limited historic data for research
  - Businesses more likely to report broad products than detail products
  - Different types of details by industry

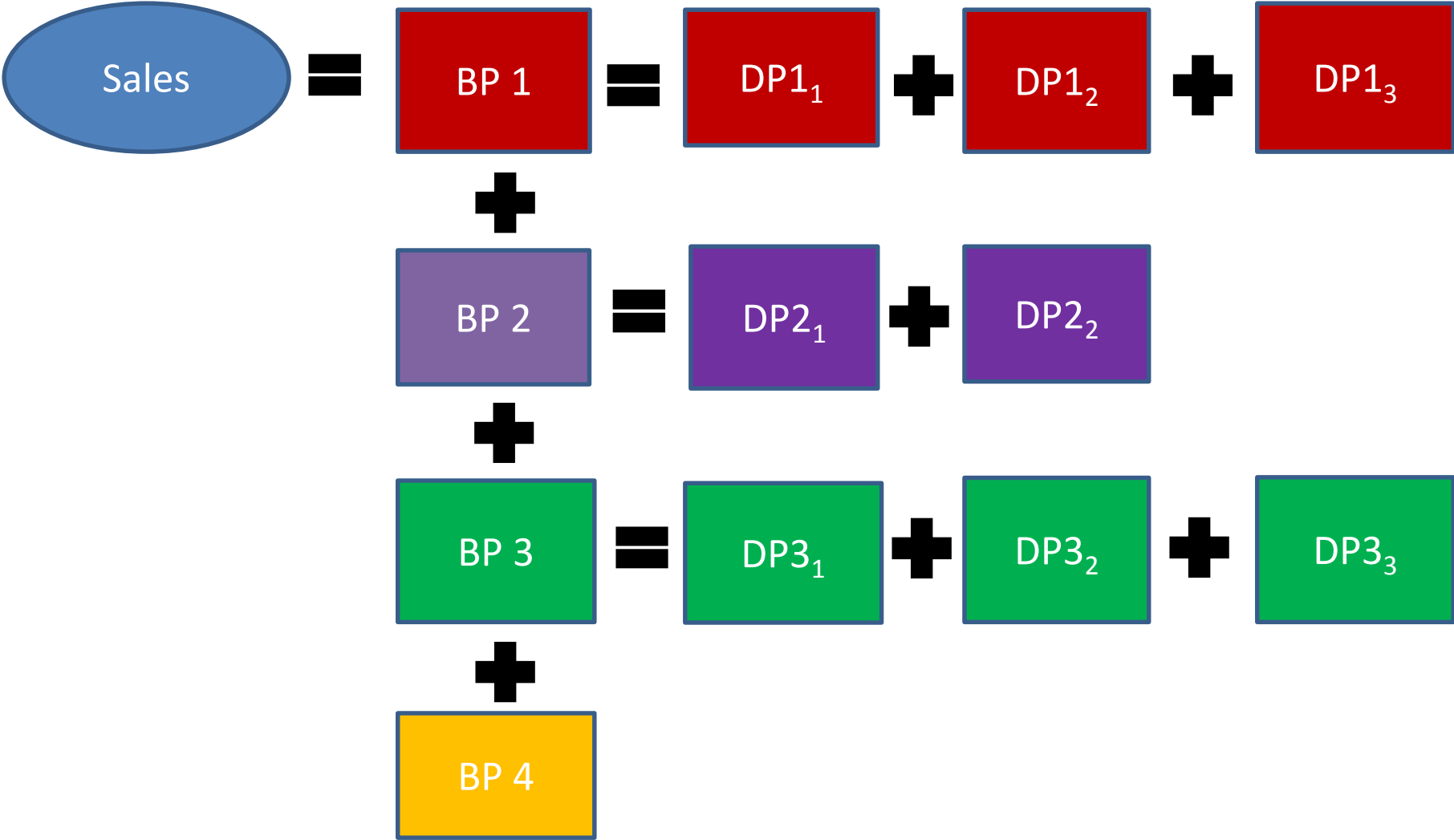- Subject matter experts wanted to maximize use of valid reported data in imputation

Imputation of Detail Products

- Group establishments into types

- Use category average (ratio) imputation
  - Statistical model frequently used for business data
  - NOT part of the research for product data

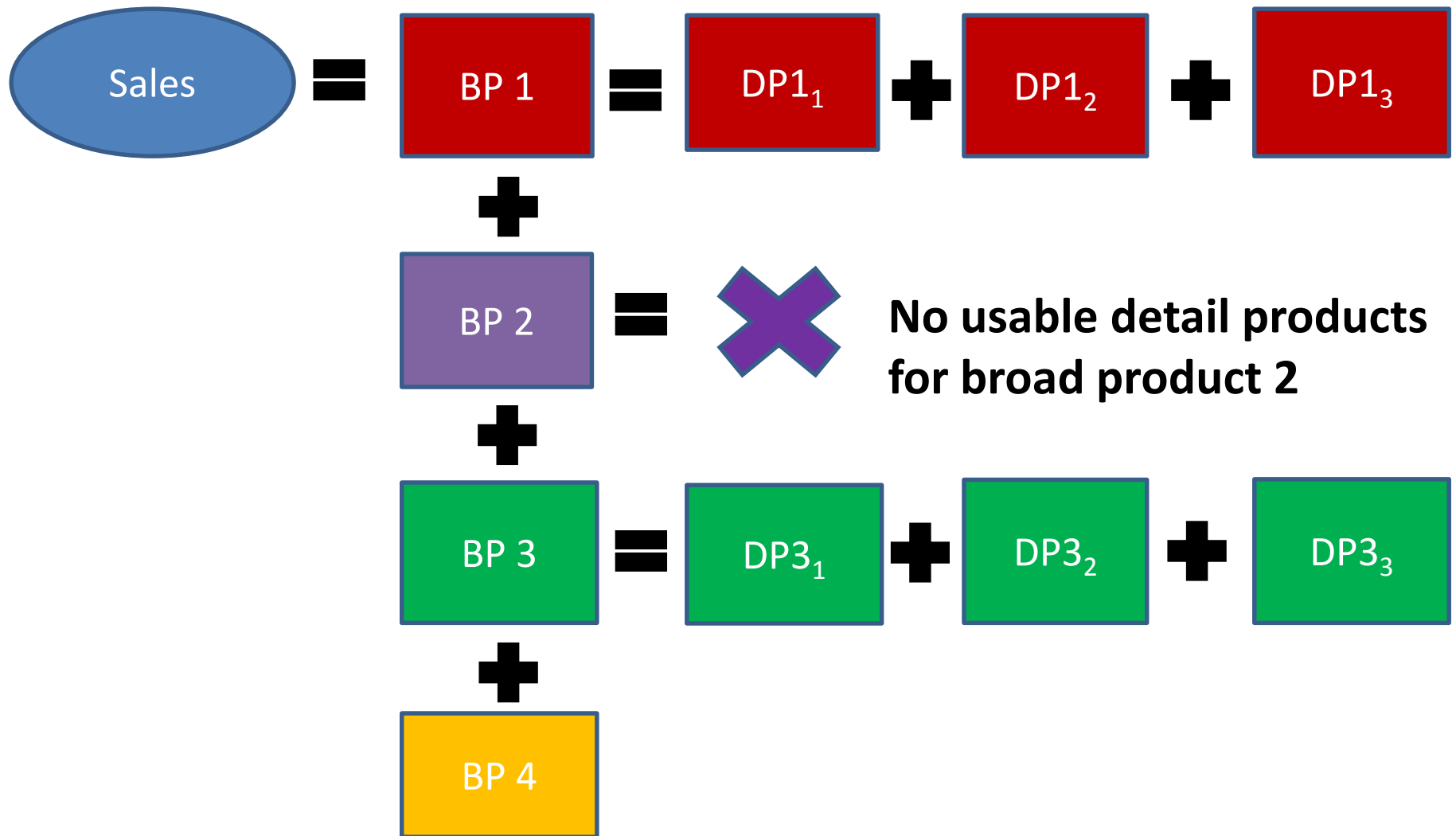- "Hot deck" imputation – all products (broad & detail)

Establishment Groups

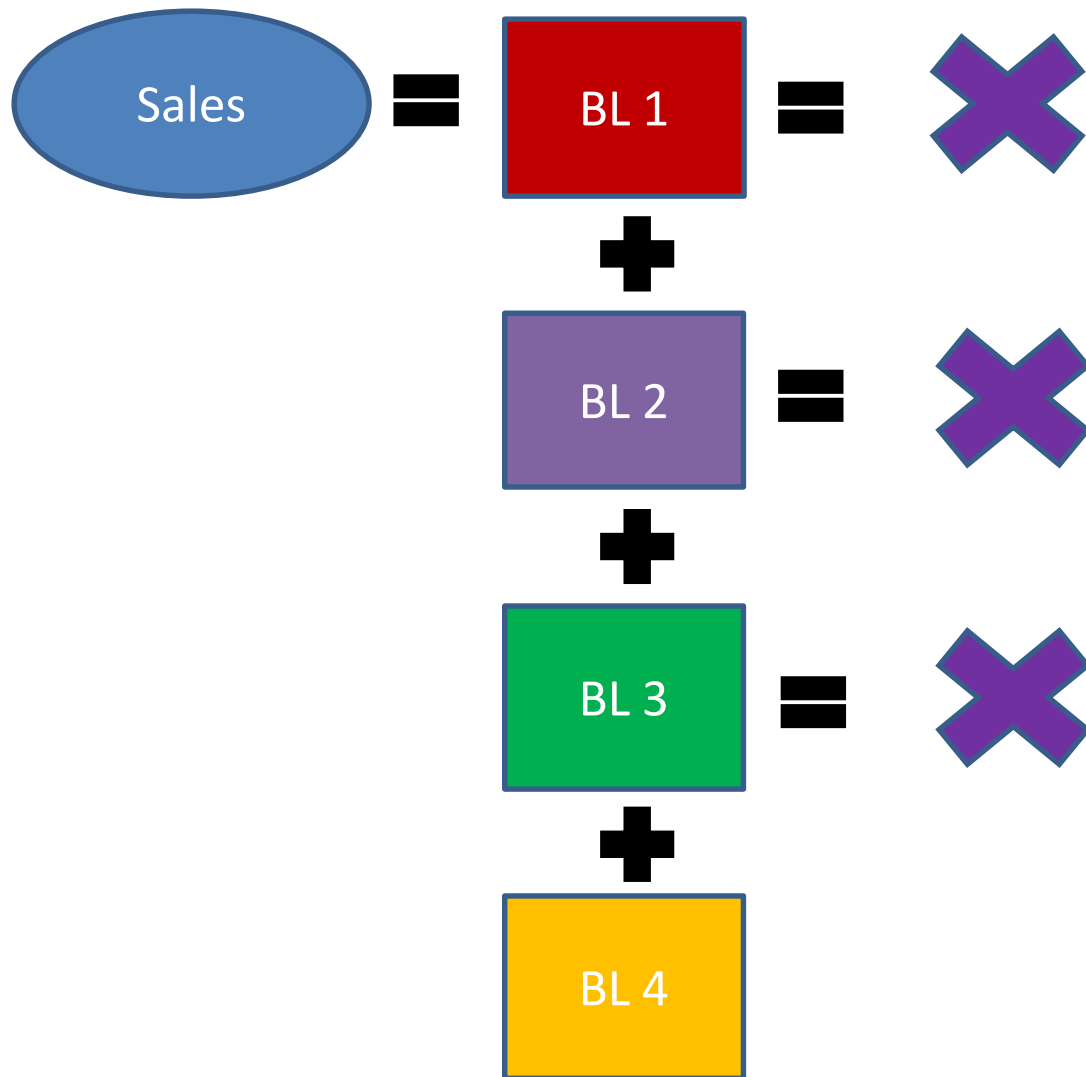| Donors	Broad products usable
Complete	All required Detail products balance to Broad products
Partial	Some usable Detail products
Minimal	No usable Detail products
Recipients	**Missing products**
Full	Need Broad products and Detail products
Partial	Need some (designated) Detail products
Minimal	Need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

Complete Donor Example

$$\text{Sales} = \text{BP 1} = \text{DP1}_1 + \text{DP1}_2 + \text{DP1}_3$$

$$+$$

$$\text{BP 2} = \text{DP2}_1 + \text{DP2}_2$$

$$+$$

$$\text{BP 3} = \text{DP3}_1 + \text{DP3}_2 + \text{DP3}_3$$

$$+$$

$$\text{BP 4}$$

Partial Donor Example

$$\text{Sales} = BP\ 1 = DP1_1 + DP1_2 + DP1_3$$

$$+$$

$$BP\ 2 = \times \quad \textbf{No usable detail products for broad product 2}$$

$$+$$

$$BP\ 3 = DP3_1 + DP3_2 + DP3_3$$

$$+$$

$$BP\ 4$$

Mimimal Donor Example

Sales = BL 1 = ✖

BL 1 + BL 2 = ✖

No usable detail products for any broad product

BL 2 + BL 3 = ✖

BL 3 + BL 4

"Completing" Partial Donors

Missing detailed products for BL 2 and 3

Inputs

Sales — BL 1 — DP1₁ — DP1₂ — DP1₃

BL 2 — ✗

BL 3 — ✗

BL 4

Category Averages for BL 2 and 3

BL 2 — DP2₁ 85% — DP2₂ 10% — DP2₃ 4.5% — DP2₄ 0.5%

BL 3 — DP3₁ 99% — DP3₂ 0.7% — DP3₃ 0.3%

Average proportion that each detail product contributes to a broad product

Output

Sales — BL 1 — DP1₁ — DP1₂ — DP1₃

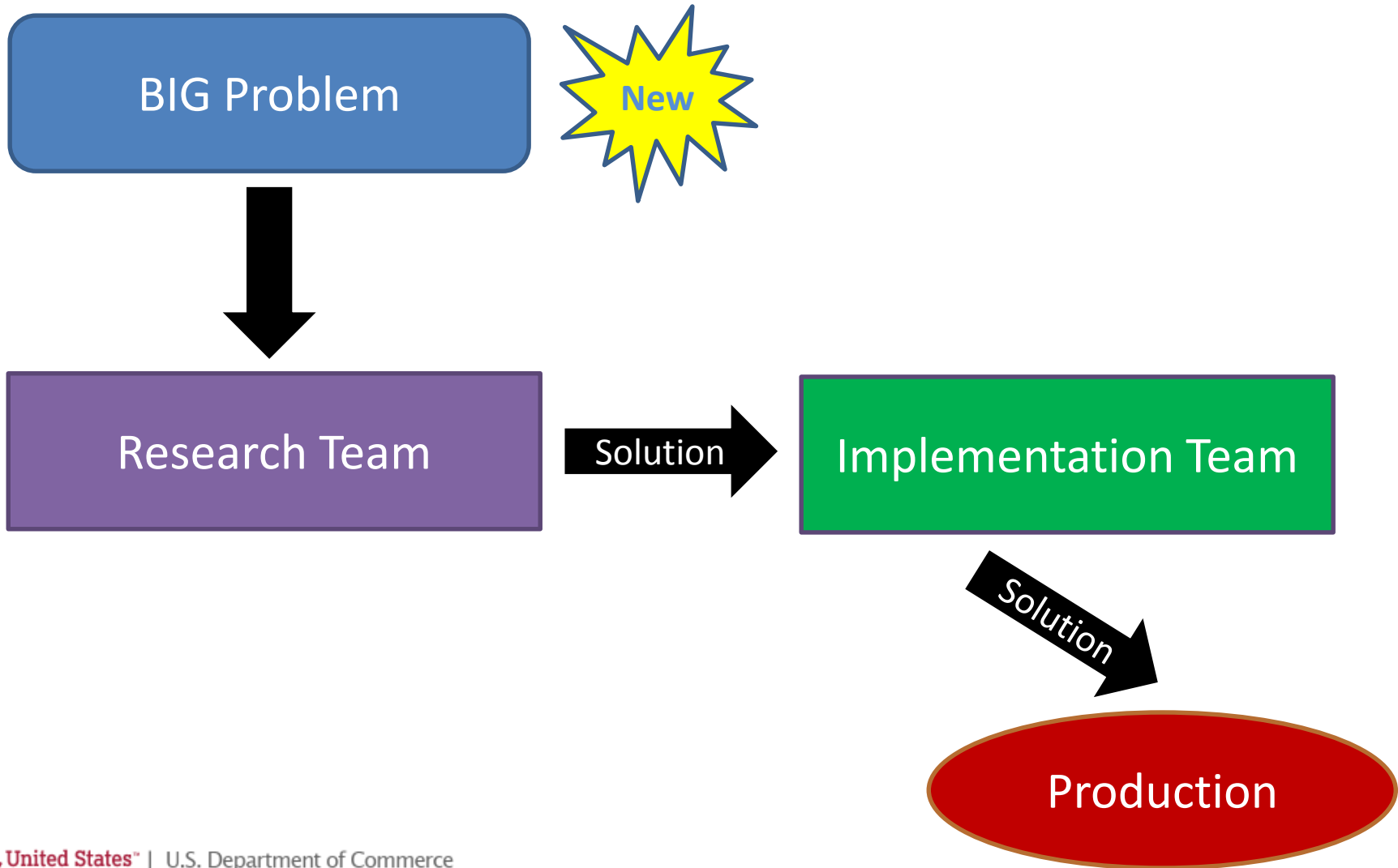BL 2 — DP2₁ — DP2₂ — DP2₃ — DP2₄

BL 3 — DP3₁ — DP3₂

BL 4 $236
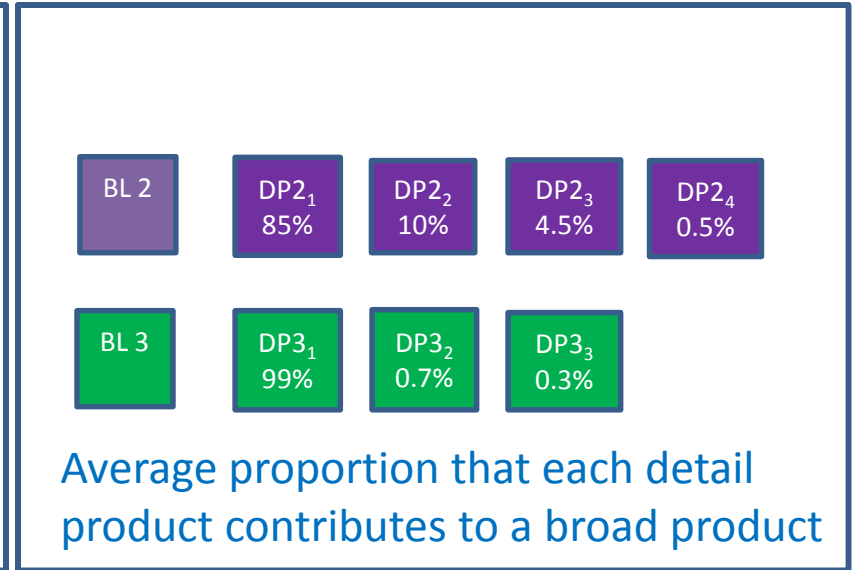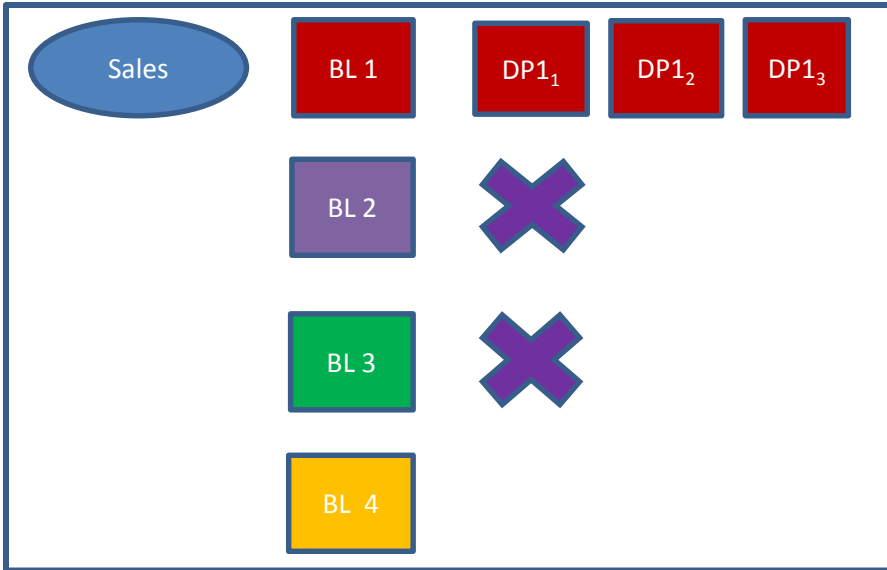
Completed record is now a donor

Establishment Groups

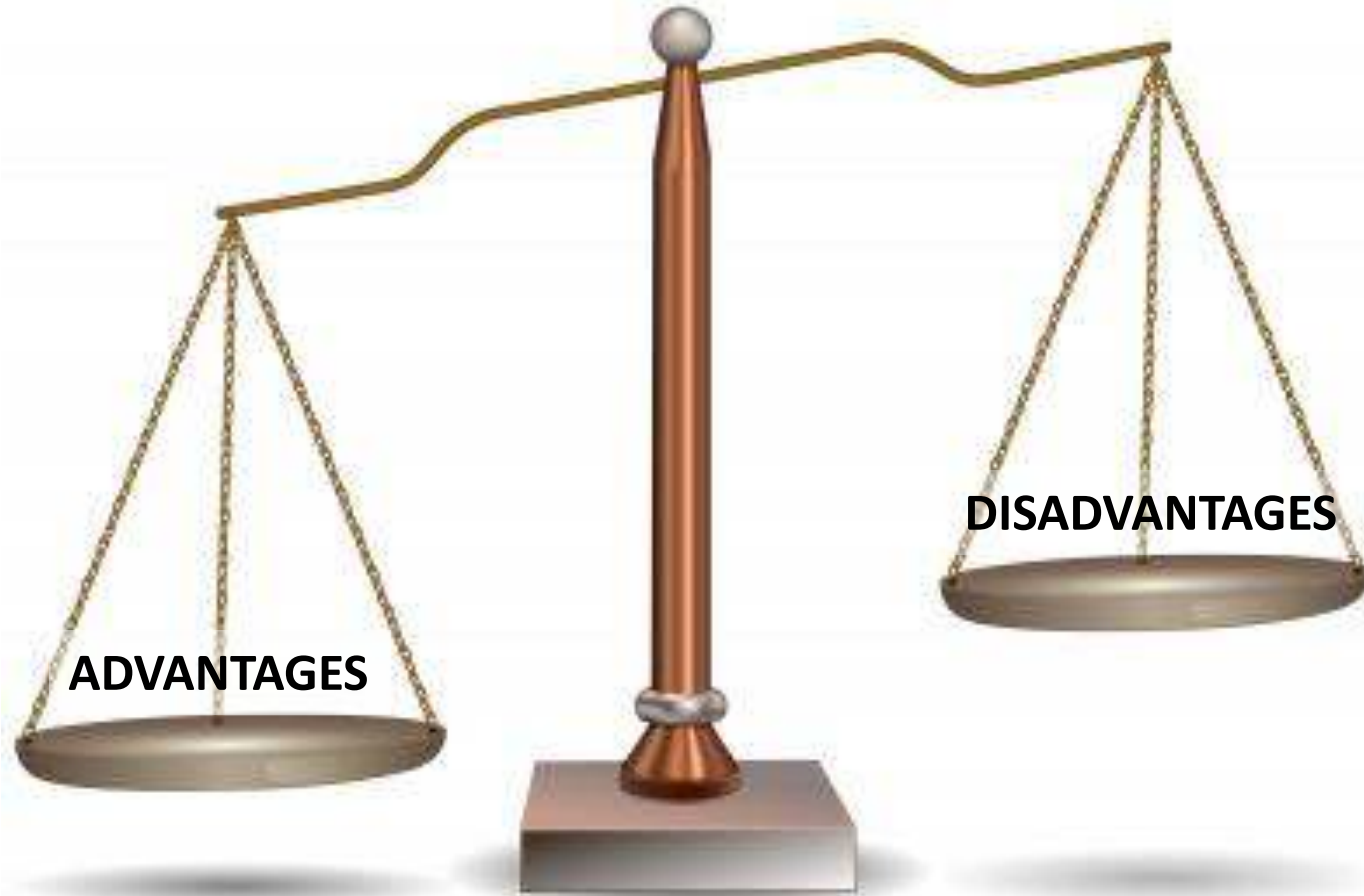| Donors	Broad products usable
Complete	All required Detail products add to Broad products
Partial	Some usable Detail products
Minimal	No usable Detail products

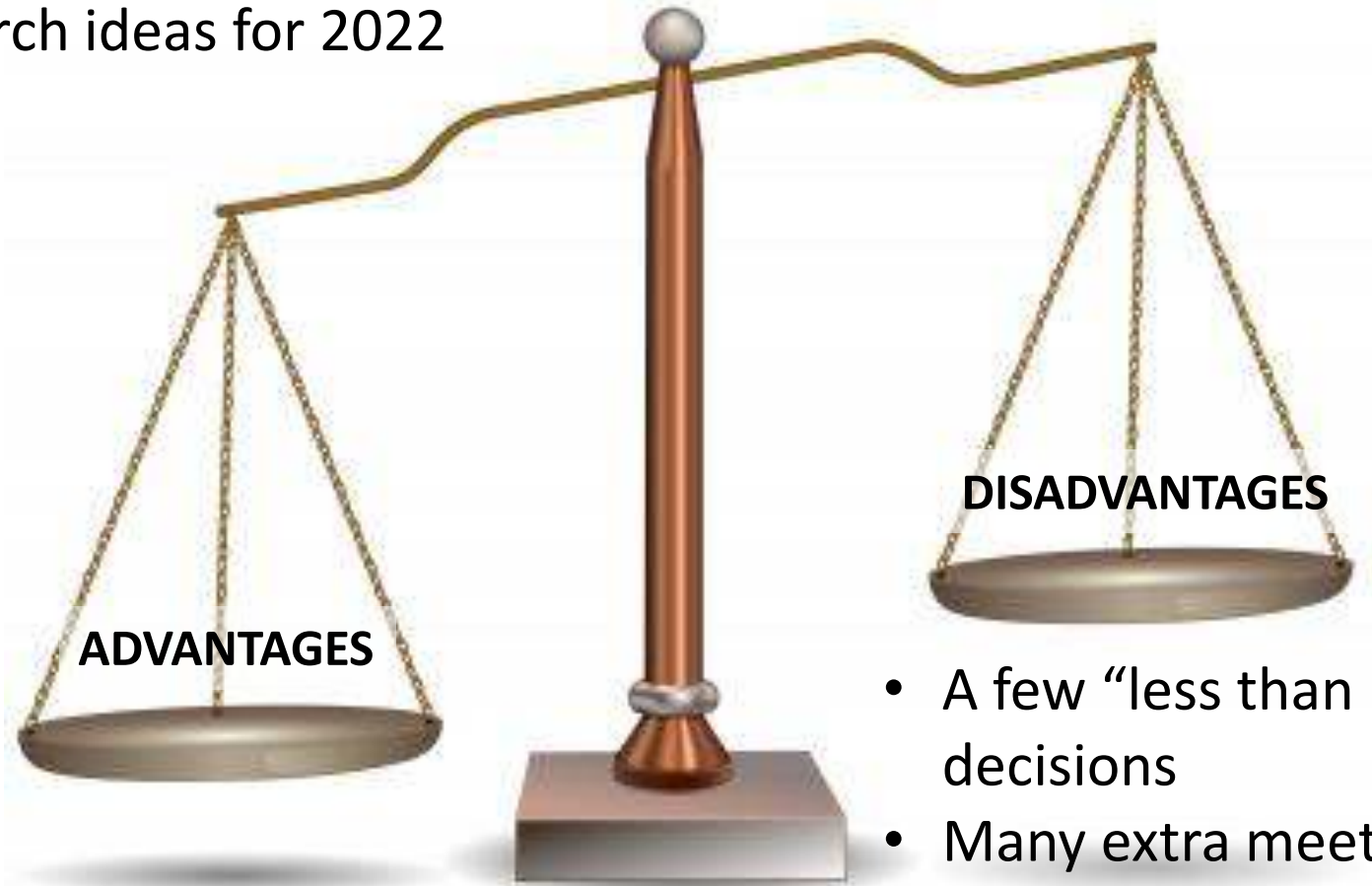| Recipients	Missing products
Full	Need Broad products and Detail pro...
Partial	Need some (designated) Detail products
Minimal	Need all designated Detail products
Ineligible	All products usable, but not "typical"; excluded from donor pool

These units are both donors and recipients!

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Going Back to Original Picture

BIG Problem

New

Research Team

Solution

Implementation Team

Solution

Production

Summary

ADVANTAGES

DISADVANTAGES

Summary

- Workable solutions
- Buy-in on methods
- Shared understanding
- Research ideas for 2022

DISADVANTAGES

ADVANTAGES

- A few "less than perfect" decisions
- Many extra meetings

Thank you

Contacts:

Katherine.J.Thompson@census.gov

Laura.Bechtel@census.gov