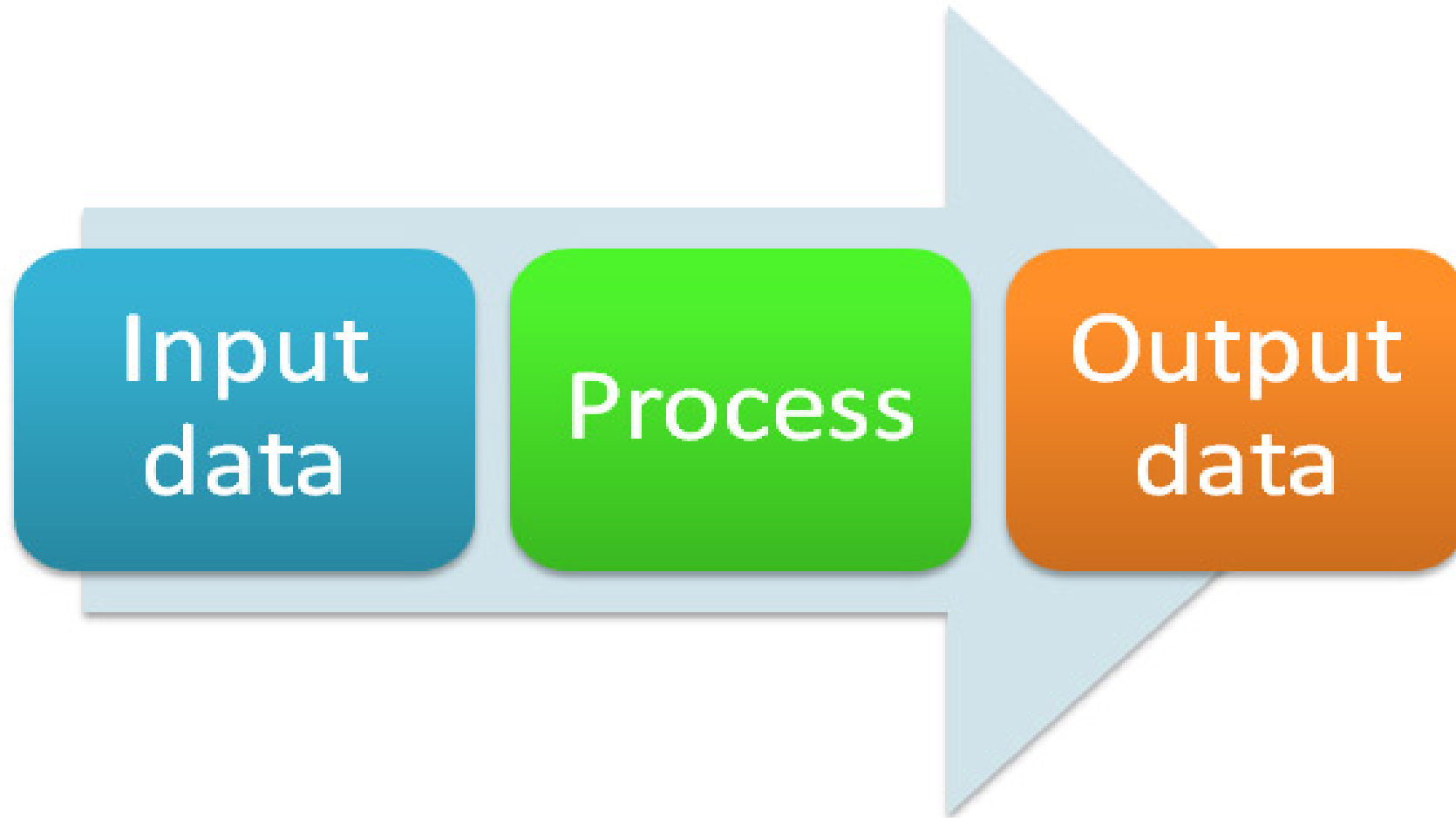


Data Management – Processing

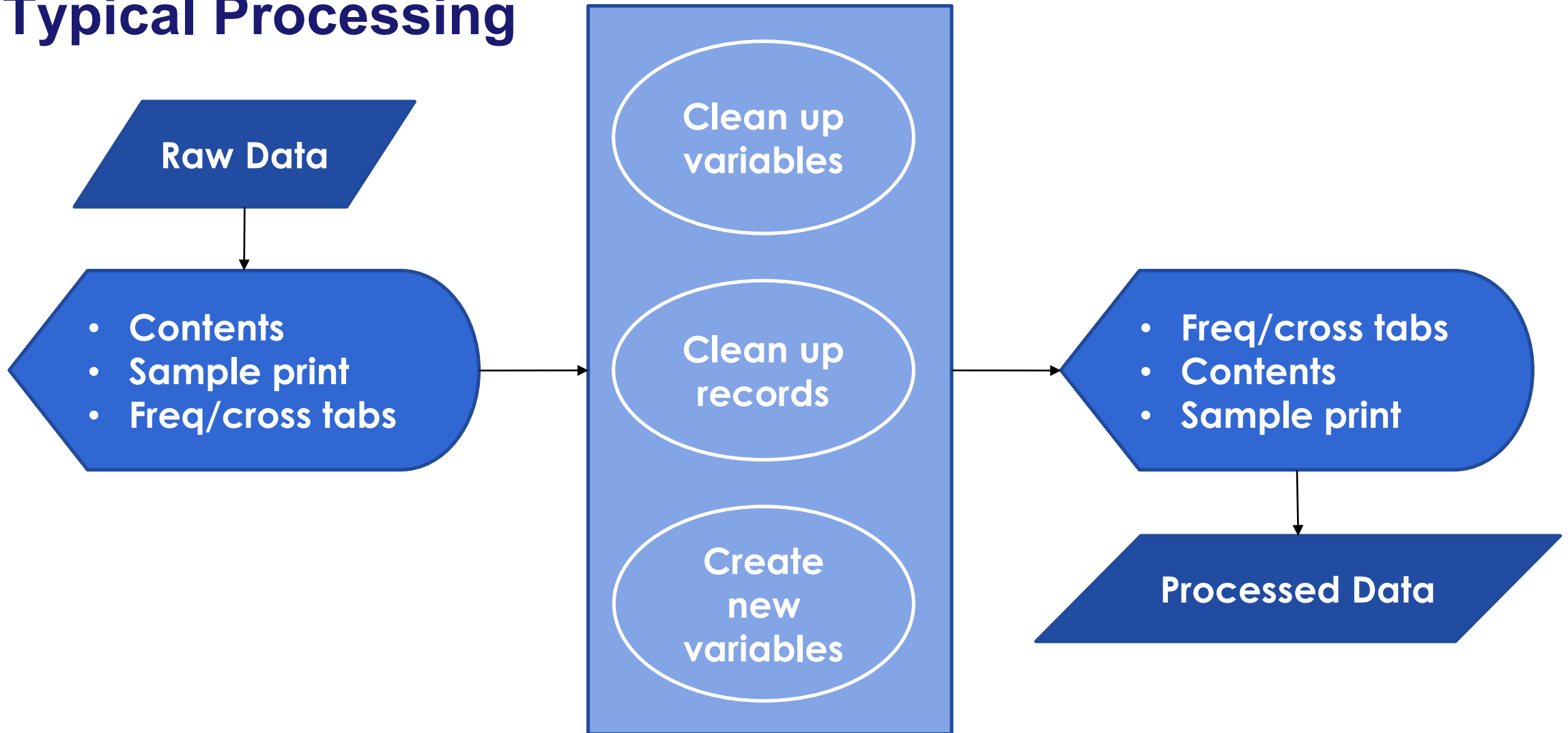
Claire Osgood
November 2017



Processing



Typical Processing



Typical Processing

Clean up variables

- Label
- Rename
- Drop empty variables
- Standardize values within fields
- Convert dates and times to system format
- Numbers as text:
 - Convert to numeric; OR
 - Add leading zeros
- Deal with invalid, missing/unknown, 0/blank

Race

{
 "Hispanic"
 "hispanic"
 "hisp"
 "H"
 } = "Hispanic"

Numeric Code

1	01
10	02
11	10
12	11
2	12
21	21
22	22

State Abbreviation

—	—
--	AR
AR	LA
LA	OK
NA	TX
OK	
TX	
Unk	

Typical Processing

Clean up
records

- Check for unique identifier
 - If none, what combination of variables uniquely identifies the record?
- Look for duplicates and deal with them
- Delete empty records
- Exclude other records as appropriate
- Sort or index

Typical Processing

Create
new
variables

Address

Street number + Pre Dir + Street Name + Post Dir

Address

Age Group

1

2

...

17

18

19

20

21

22

...

30

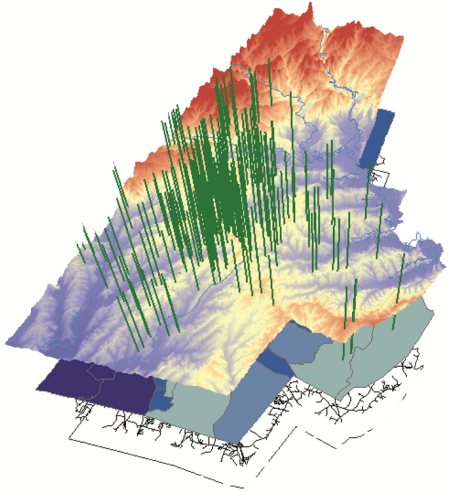
1

2

3

- Indicators commonly used
- Calculated variables;
combined/concatenated values;
separating values
- Coded variables
- Standard variables

Other Processing



GIS

- Geocoding
 - Create and output a file for geocoding
 - Separately, geocode
 - Second program to incorporate geocoded information
- GIS process to add spatial data



Merge or combine with other datasets

Tips

- You are going to have to account for adding/removing records, merges with other datasets, etc. Set up programs to have those counts readily available.
- Template programs for common processes
 - Cumulative Inclusion Enrollment Form for grants
 - Standard exclusions or common calculations
- Common/standard variables
 - Race or age groups
 - Indicators