

# C2 Metadata

---

SDTL, SPSS, R, AND APPS

# Overview

---

- SDTL
- SPSS
- R
- Apps to detect and view transforms

---

SDTL

Structured Data Transform Language

A solid orange horizontal bar at the bottom of the slide.

# SDTL

---

- Structured Data Transform Language
- A model to describe things people do to transform data

# SDTL: Assign a variable label

---

- Stata
  - `label variable price "the price of the car in 1978"`
- SPSS
  - `VARIABLE LABELS price "the price of the car in 1978"`
- SDTL
  - Command: `"setVariableLabel"`
  - Variable: `"price"`
  - Label: `"the price of the car in 1978"`

# Recode Values (Stata)

---

- **recode** -- Recode categorical variables

```
recode var1 var2 (1 2 = 2) (3/max = 5)
```

```
recode a b (1 = 2), prefix("modified_")
```

```
recode x y (1 = 2), generate(modified_x modified_y)
```

```
recode x y (1 = 2 "Labelfor2") (3 = 5 "Labelfor5")
```

```
recode total (0/140=0 F) (141/180=1 D) (181/210=2 C) (211/234=3 B)  
(235/300=4 A), gen(grade)
```

```
recode a (1 . .a 5/6 = 7) (nonmissing = 8) (missing = 9) (* = 2)
```

# Recode Values (SPSS)

---

RECODE **var1 var2 var3** (-1=7) (-2=8).

RECODE **AGE** (*MISSING=9*) (*18 THRU HI=1*) (*0 THRU 18=0*) INTO **VOTER**.

# Recode in SDTL

---

- Command: `"recode"`
- `recodedVariables: "age"`
- Rules [0]:
  - `fromValue: -1`  
`to: 7`
- Rules [1]:
  - `fromValue: -2`  
`to: 8`



# SDTL Development

---

- Using COGS, an open source production framework to generate JSON schema, XML schema, and rich documentation
- <http://c2metadata.gitlab.io/sdtl-docs/>

# SDTL Documentation

---

[Docs](#) » [All Composite Types](#) » [Recode](#)

[View page source](#)

## Recode

Describes recoding values in one or more variables according to a specified mapping.

### Contents

- [Recode](#)
  - [Item Type Hierarchy](#)
  - [Relationships](#)
  - [Properties](#)
    - [RecodedVariables](#)
    - [RecodedVariableRange](#)
    - [Rules](#)

### Item Type Hierarchy

- [TransformBase](#)
  - [Recode](#)

# SDTL Development

---

- Still in active development
- [gitlab.com/c2metadata/sdtl-cogs](https://gitlab.com/c2metadata/sdtl-cogs)

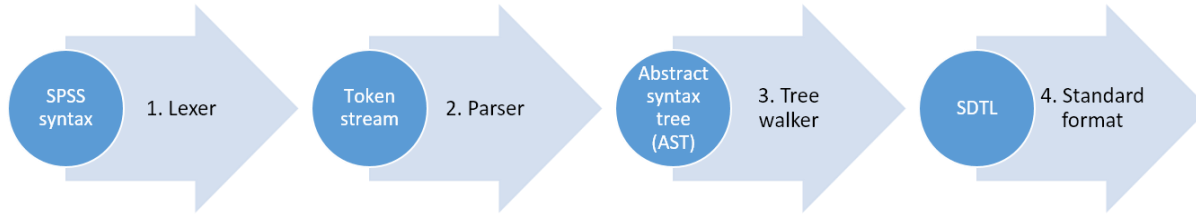
---

SPSS

A solid orange horizontal bar spanning the width of the slide at the bottom.

# Detecting Transforms in SPSS Syntax

---



# Detecting Transforms in SPSS Syntax

---

- Inputs: SPSS source code
- Output: SDTL JSON



# R Package for Transforming Data

---

- R is a more sophisticated, less domain specific language than SPSS and others
- Instead of detecting transforms: produce an R package to perform transforms
- Wrap commonly used libraries like hmisc and dplyr
- Perform transforms easily. Output SDTL as a byproduct
  
- Work begins later this year



---

# Apps to Detect Transforms

# SPSS: Command line

---

```
C:\svn\spss-sdtl-converter\src\C2Metadata.SpssToSdtl.Cli (master)  
λ dotnet run -- c:\svn\spss-samples-public\basic\recode.sps -o d:\out\recode.json  
Recognized command count: 4
```

# SDTL Reader

---

- Desktop application for Windows, macOS, and Linux
- Open SPSS syntax files (\*.sps)
- Open SDTL JSON files

## Summary

- ▶ M3\_PHONE\_STEP 1\_GENERAL CLEANING
- ▶ M3\_PHONE\_STEP 2\_INITIAL FREQS
- ▶ M3\_PHONE\_STEP 3\_RECODE MISSING VALUES
- ▶ M3\_PHONE\_STEP 4\_CHECKING VALUE RANGES
- ▶ M3\_PHONE\_STEP 5\_CHECK SKIP PATTERNS
- ▶ M3\_PHONE\_STEP 6\_CREATING NEW VARIABLES FOR
- ▶ M3\_PHONE\_STEP 7\_RELABEL VARIABLE LABELS
- ▶ M3\_PHONE\_STEP 8\_DEFINE MISSING VALUES
- ▶ M3\_PHONE\_STEP 9\_UPDATE VALUE LABELS
- ▶ M3\_PHONE\_STEP 10\_RENAME UWSC VARIABLE NAM
- ▶ M3\_PHONE\_STEP 11\_MERGE ROSTER WITH PHONE
- ▶ M3\_PHONE\_STEP 12\_GENERAL CLEANING
- ▶ M3\_PHONE\_STEP 13\_SCALES AND CONSTRUCTED V
- ▶ M3\_PHONE\_STEP 14\_FORMATS
- ▶ M3\_PHONE\_STEP 15\_ADD OCC-IND CODES

## Summary

## Functions

## Variables

Total	401
recode	132
execute	117
compute	53
rename	28
load	17
title	15
subtitle	15
save	15
Invalid	9

## Summary

- ▶ M3\_PHONE\_STEP 1\_GENERAL CLEANING
- ▶ M3\_PHONE\_STEP 2\_INITIAL FREQS
- ▲ M3\_PHONE\_STEP 3\_RECODE MISSING VALUES

title

subtitle

load

recode

recode

recode

recode

recode

recode

recode

recode

recode

recode

recode

recode

execute

recode

recode

recode

recode

recode

Recoded Variables A49 -&gt; A49

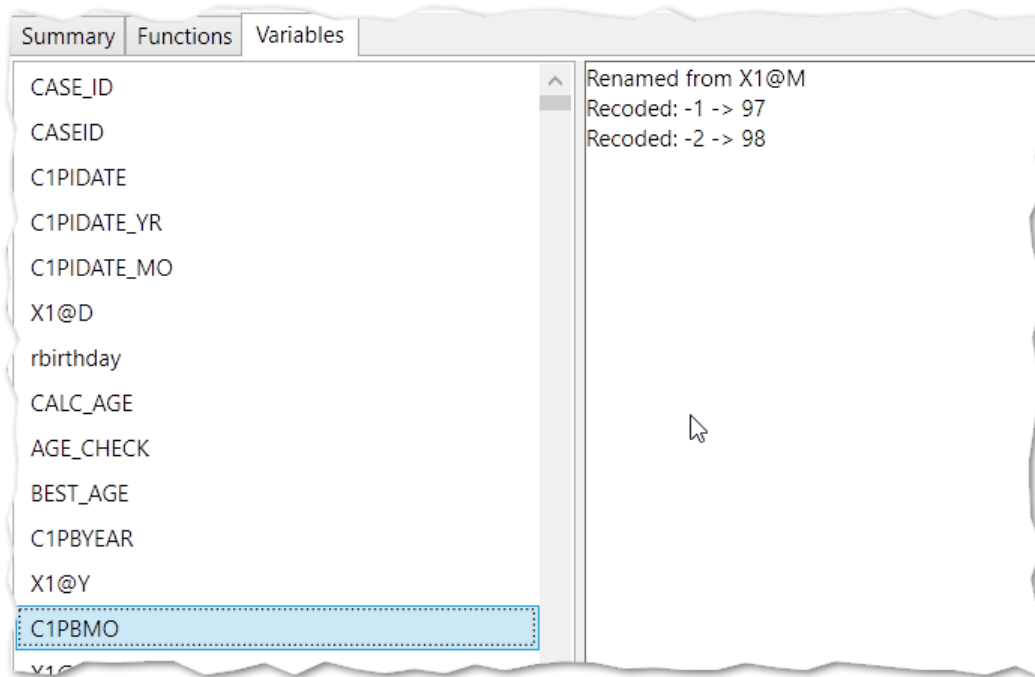
Recode Rules -1 -&gt; 97

-2 -&gt; 98

-3 -&gt; 96

Source File C:\svn\spss-samples-private\midus\M3\M3\_PHONE\_STEP 3\_R

Source RECODE A49  
(-1=97) (-2=98) (-3=96).



# Web Service and Docker Container

---

- Inputs: SPSS Syntax
- Outputs: SDTL JSON
- Install and run:

```
$ docker pull registry.gitlab.com/c2metadata/sdtl-reader:latest  
$ docker run -d -p 8080:80 registry.gitlab.com/c2metadata/sdtl-reader
```

```
$ curl -X POST "http://localhost:8080/api/spsstosdtl" -d "COMPUTE x = 1"
```

# Follow Along

---

[gitlab.com/c2metadata](https://gitlab.com/c2metadata)