Using Extended Attributes in Data Analysis Software-

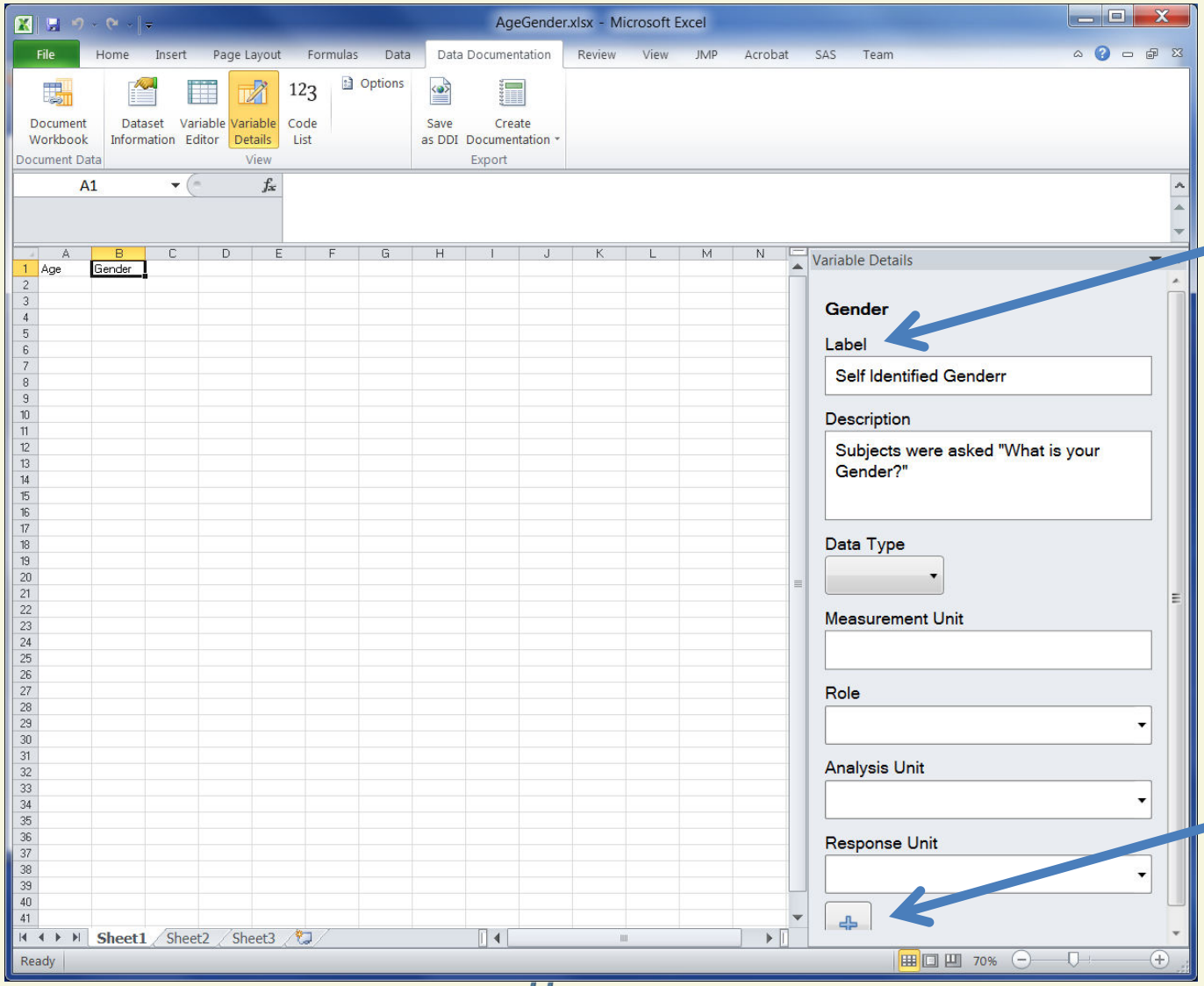
Controlled Vocabularies, Tools and DDI

Larry Hoyle
Institute for Policy & Social Research
University of Kansas
LarryHoyle@ku.edu

Custom, or Extended, Attributes

- Excel with Colectica Plugin, R, SAS, SPSS, Stata all have something in common
- User defined attributes for variables
  - (and sometimes datasets)

A Quick Tour – Excel with Colectica Plugin



Predefined Attributes

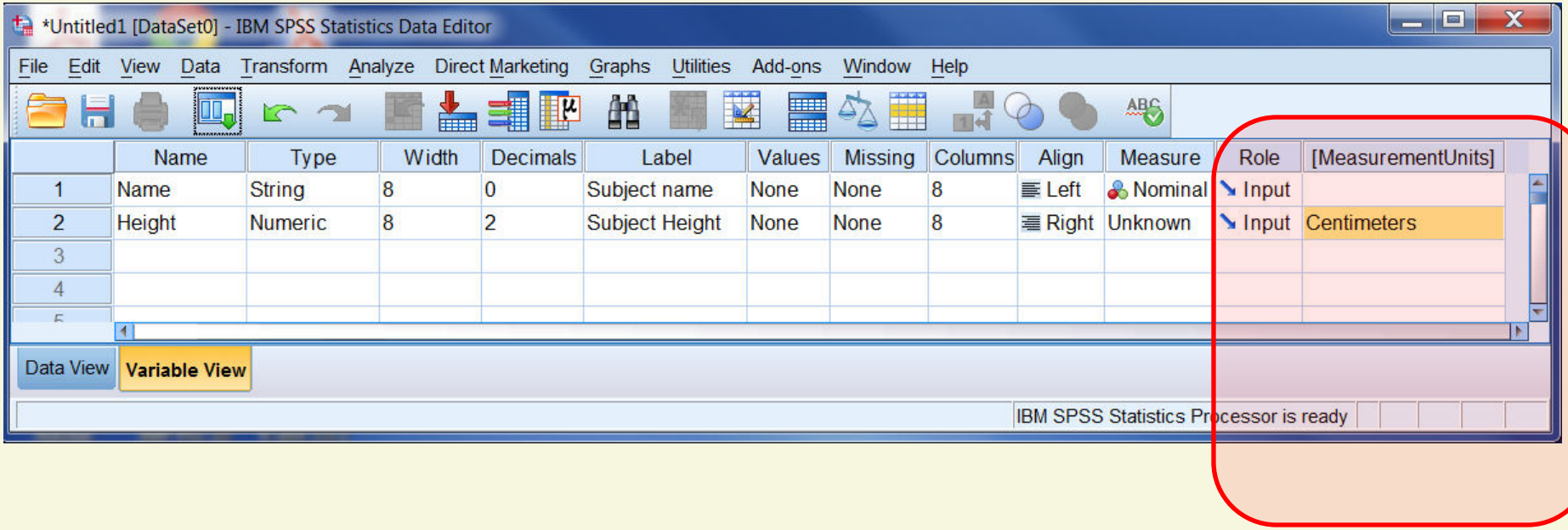
User Defined Attributes

A Quick Tour – R

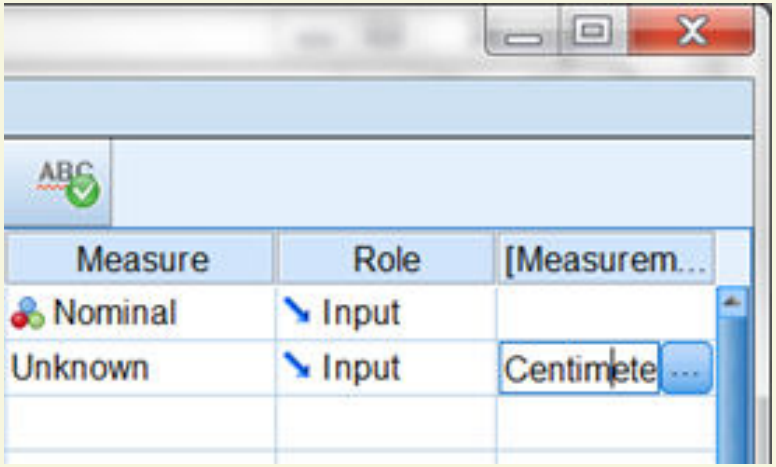
R Console

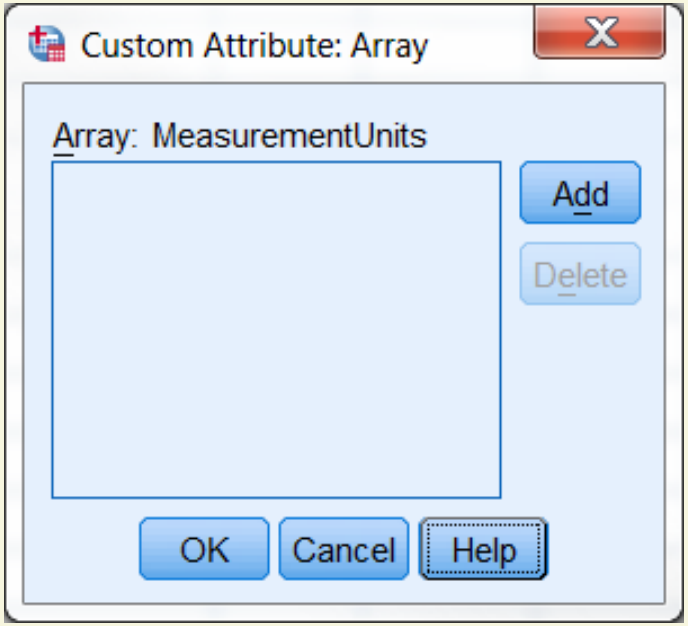
```
> fee<-c(1,2)
> name<-c('Adams', 'Zorro')
> rData<-data.frame(name,fee)
> attr(rData$fee, "MeasureMentUnits") <- "Fee is currency in Euros"
> attributes(rData$fee)
$MeasureMentUnits
[1] "Fee is currency in Euros"
```

A Quick Tour – SPSS



A Quick Tour – SPSS





A Quick Tour – Stata

```
. char define Height[MeasurementUnits] "Centimeters"
. char define _dta[Universe] "Persons aged 65 and over"
.  notes Height: First note on Height

. char list
 _dta[Universe]:            Persons aged 65 and over
 Height[note1]:             First note on Height
 Height[note0]:             1
 Height[MeasurementUnits]:  Centimeters
```

A Quick Tour – SAS

```
proc datasets lib=work nolist ;
  modify sales;
    xattr set ds Concept="purpose"
                 Description="Testing Extended Attributes";
    xattr set var
          purchase ( Role="target"
                     LevelOfMeasurement="nominal"
                     Description="A text description of the type
of item purchased")
          age (      Role="reject"
                     Minimum="0" MeasurementUnits="years")
          income ( Role="input“
                     LevelOfMeasurement="interval" );
```

Metadata in the Research Workflow, Goals

- Capture metadata when generated
  - Recorded by those who understand the work
  - Recorded while still in mind
- Minimize overhead to research
  - Burdensome procedures won't get done
    - (or won't get done well)
- Part of regular workflow

- See for example Iverson (2009) or Long (2009)

The Dilemma of Structure

- Structure makes data and metadata easier to use
  - Searching
  - Machine actionability
- Adding structure can be additional work for creators
  - Knowledge of structure scheme
  - Navigating large structures
- **Tools can help**

Multiple Controlled Vocabularies

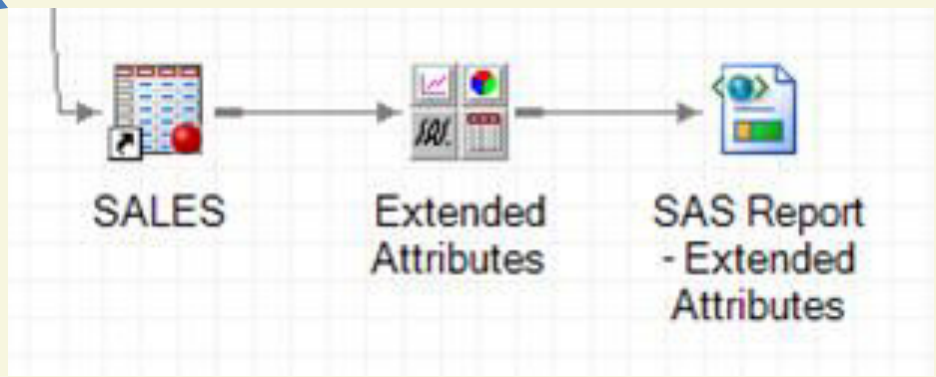
- Agreed upon attribute names
- Agreed upon values of some attributes

- E.g. Attribute: “MeasurementUnits”
  - Values from International System of Units cf. http://physics.nist.gov/cuu/Units/current.html

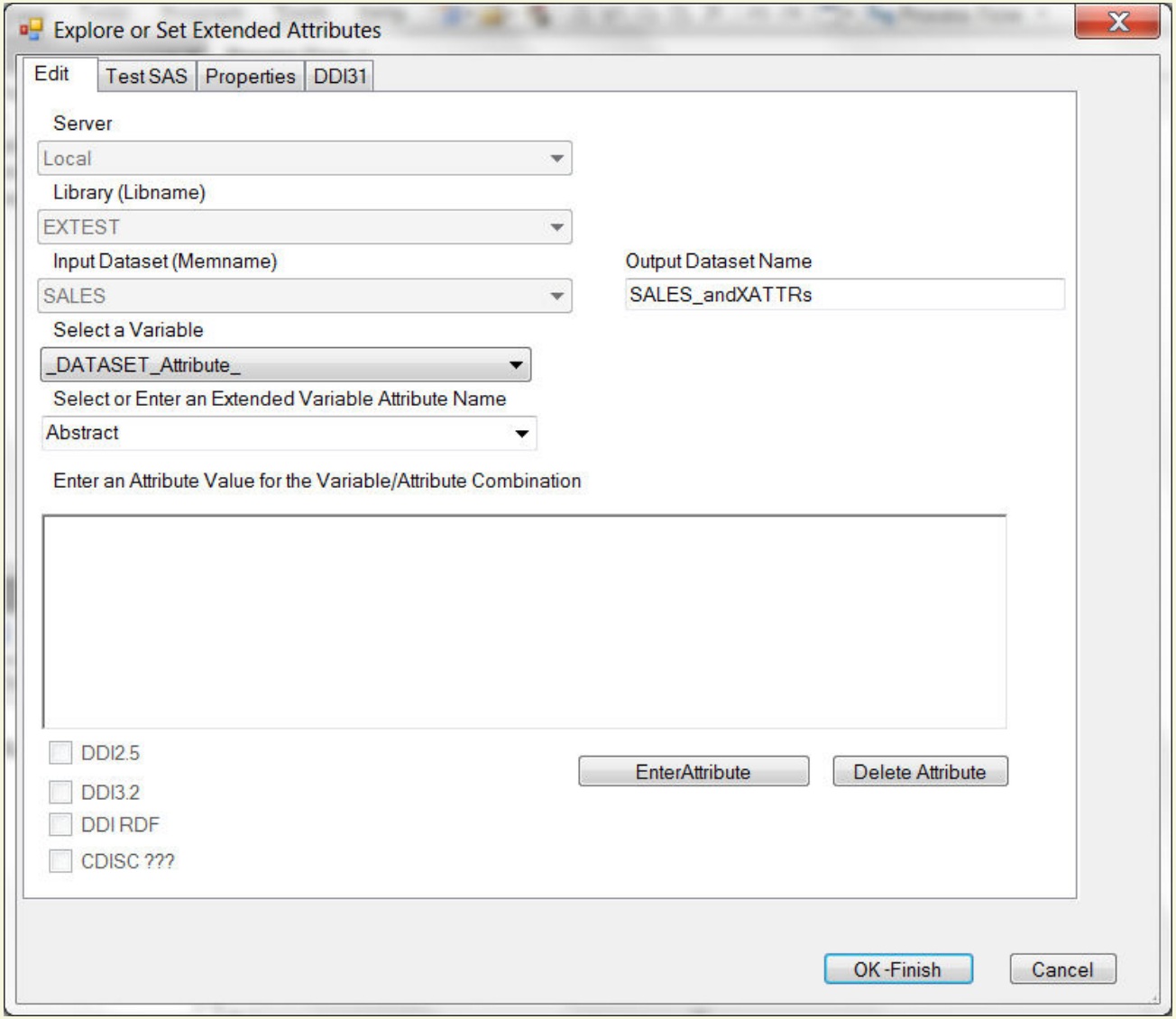
A Prototype Tool

Custom Add-in for SAS Enterprise Guide

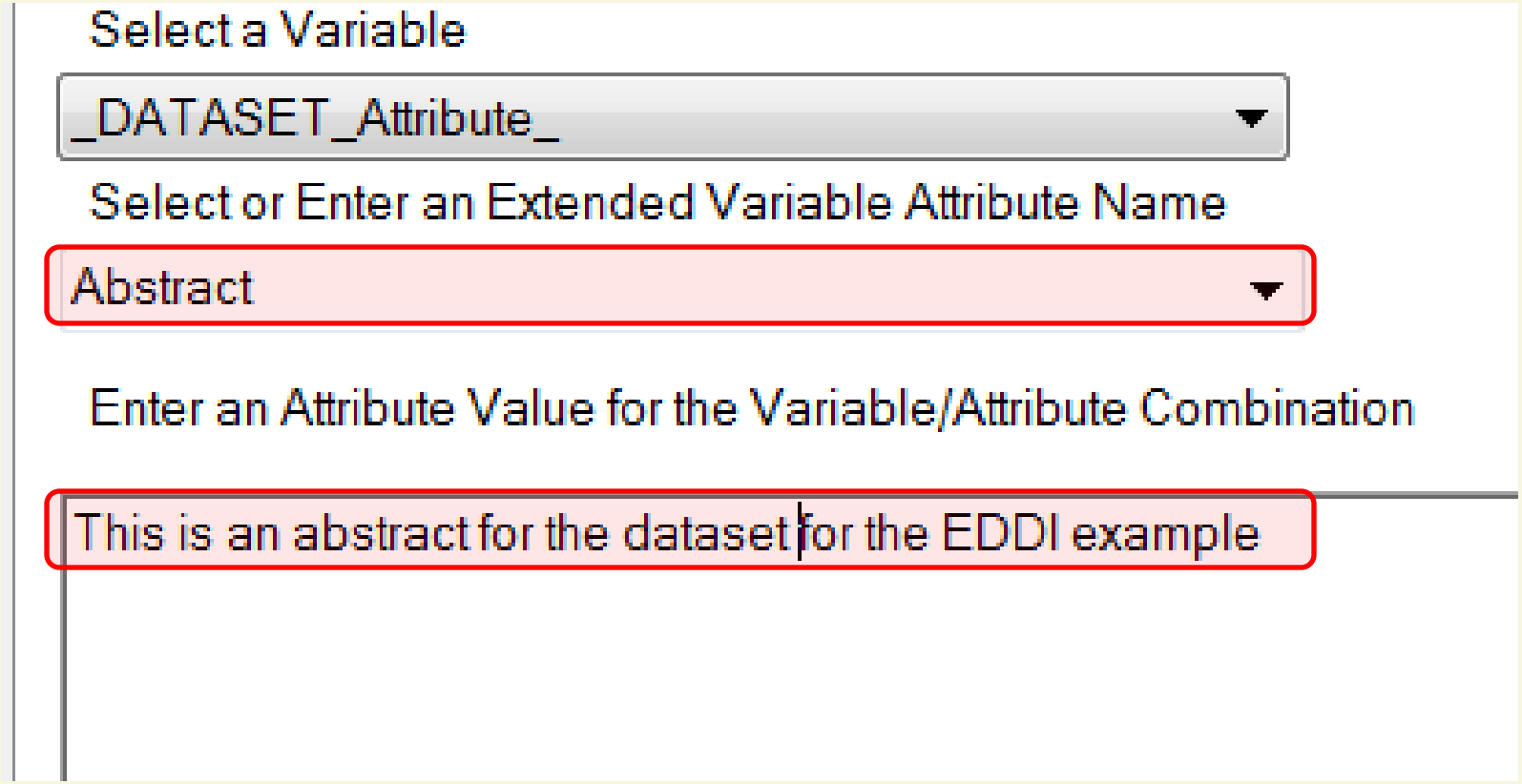
Becomes a node in a process flow diagram



First Pass at a User Interface

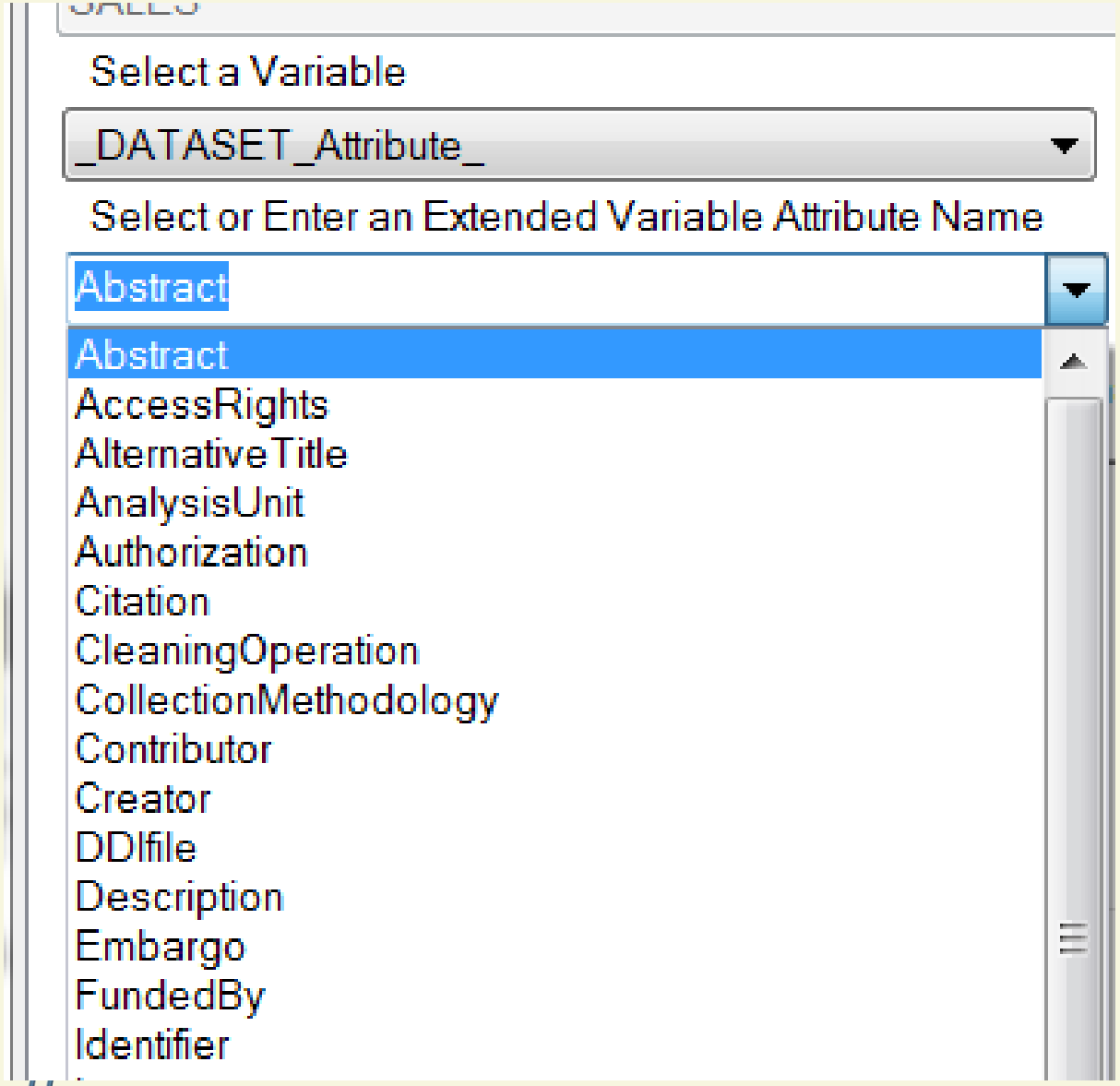


Enter, Change, or Delete Dataset Attributes



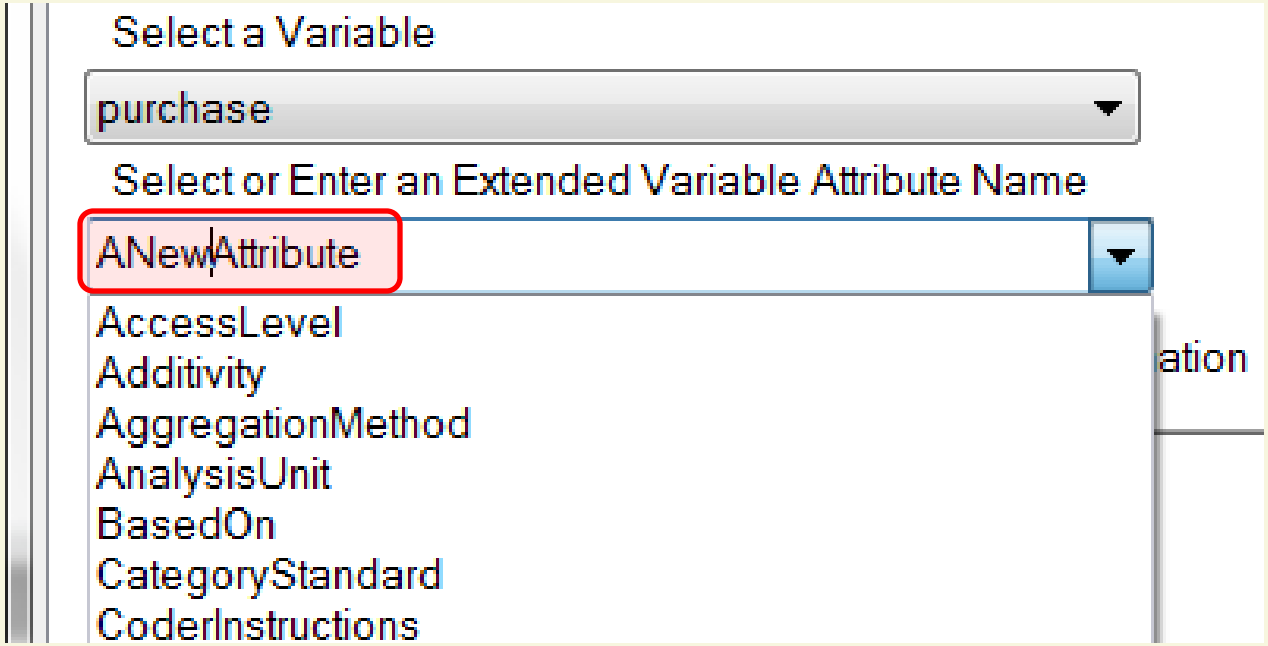
Attributes Prompted From a List

This list can be based on DDI or some other source (or both)

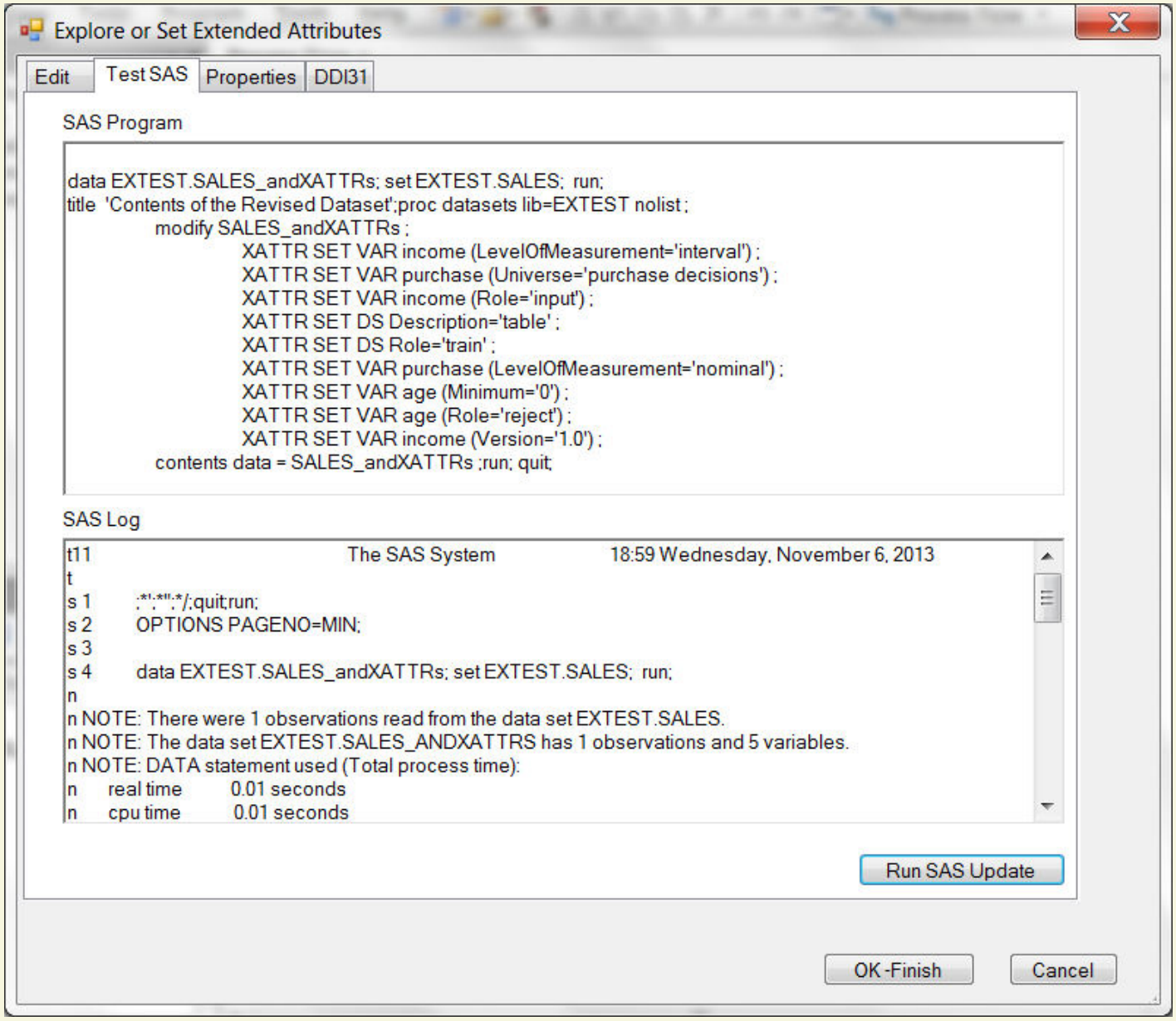


Or User Defined Attributes

A user can also enter a new attribute name.



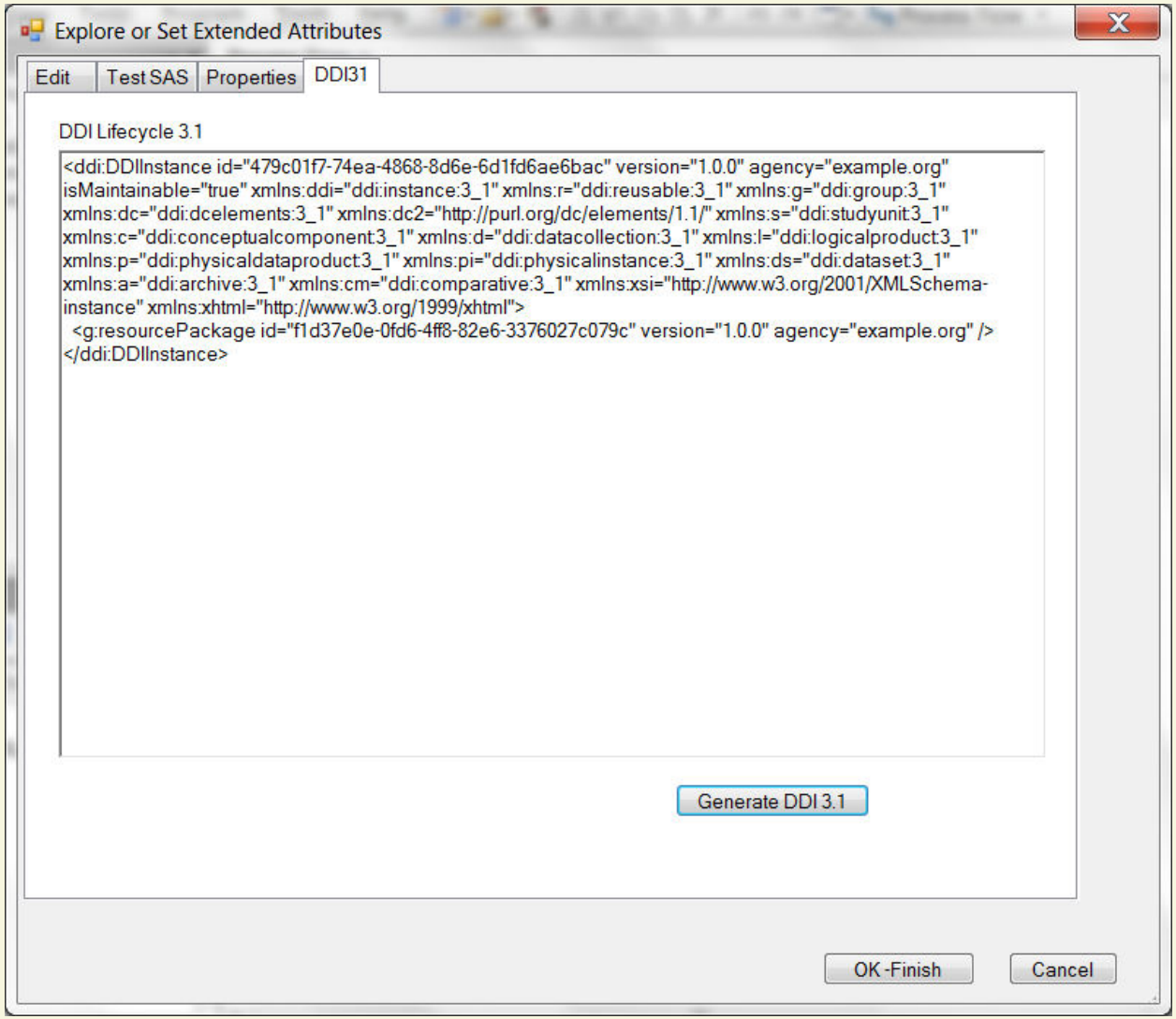
Runs SAS Procedure on Demand



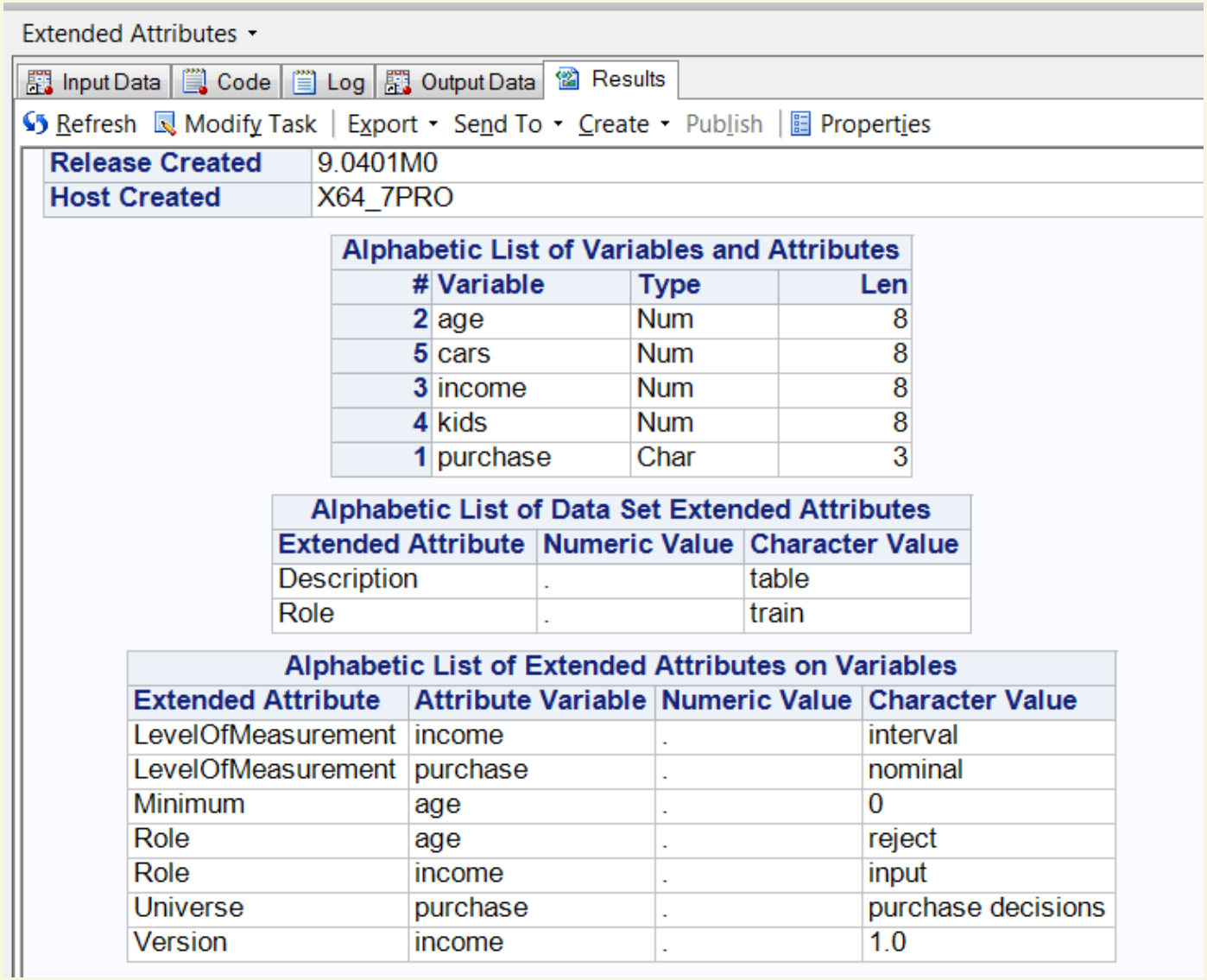
Generate DDI

Can contain Extended Attributes

Can also contain all other metadata extracted from the dataset and current SAS environment (e.g. categories and codes)



Finishes by Running SAS Program



Release Created	9.0401M0
Host Created	X64_7PRO

Alphabetic List of Variables and Attributes

#	Variable	Type	Len
2	age	Num	8
5	cars	Num	8
3	income	Num	8
4	kids	Num	8
1	purchase	Char	3

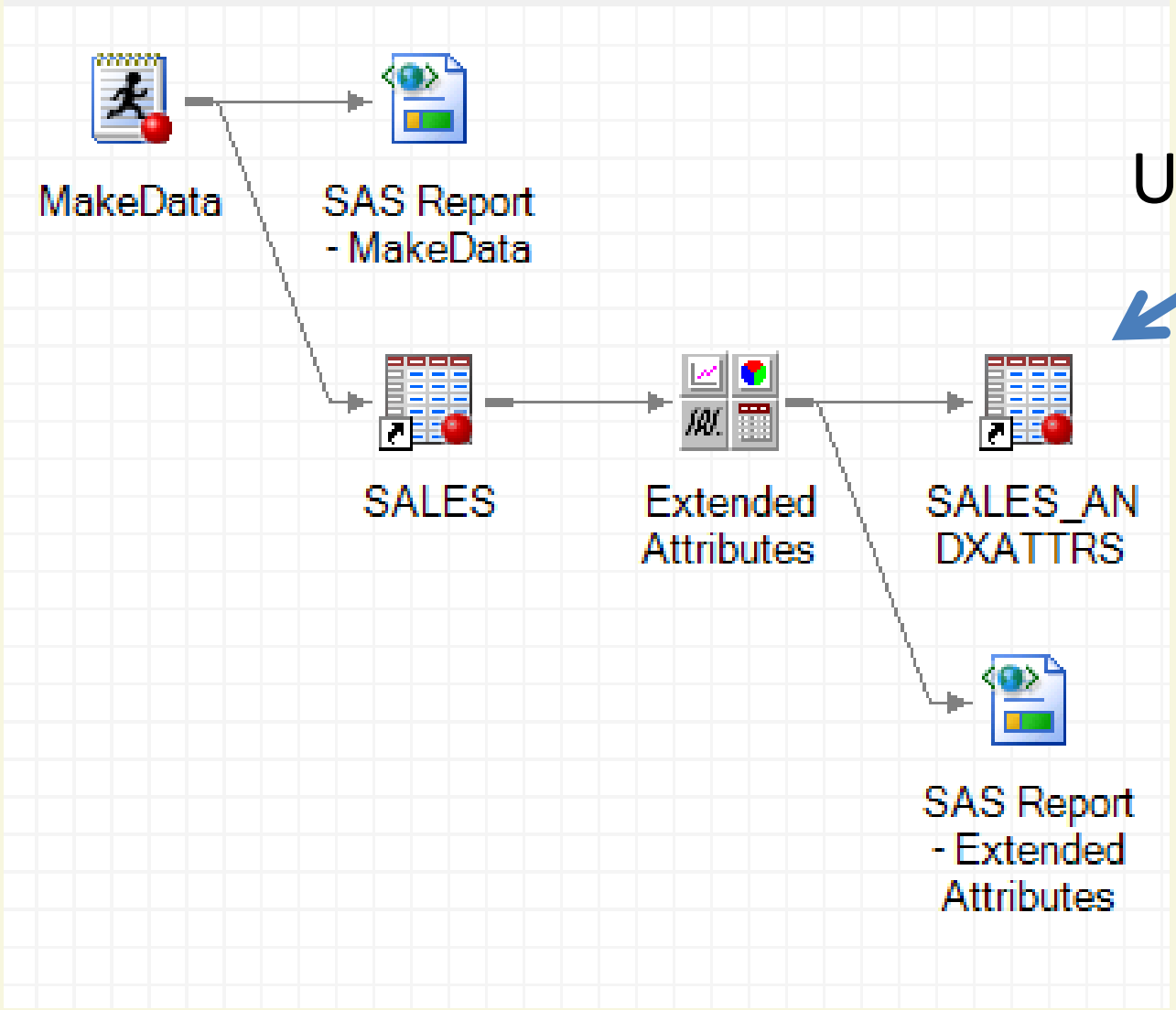
Alphabetic List of Data Set Extended Attributes

Extended Attribute	Numeric Value	Character Value
Description	.	table
Role	.	train

Alphabetic List of Extended Attributes on Variables

Extended Attribute	Attribute Variable	Numeric Value	Character Value
LevelOfMeasurement	income	.	interval
LevelOfMeasurement	purchase	.	nominal
Minimum	age	.	0
Role	age	.	reject
Role	income	.	input
Universe	purchase	.	purchase decisions
Version	income	.	1.0

Updated Process Flow



Updated Dataset

Questions?

Full Paper and sample code are available at:
http://hdl.handle.net/1808/12488

(after 03-Dec- 2013)

Larry Hoyle

LarryHoyle@ku.edu