

Using Extended Attributes in Data Analysis Software-

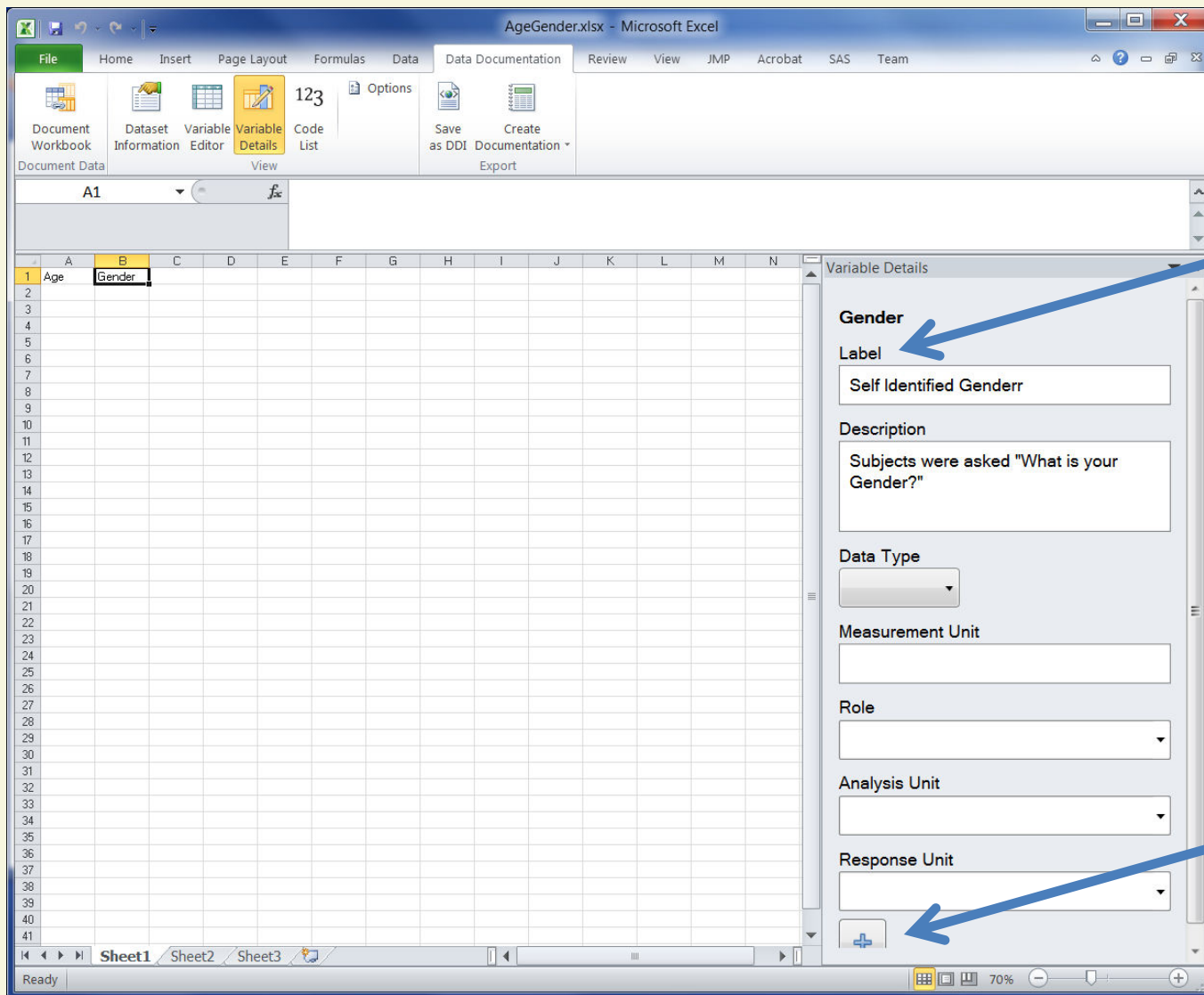
Controlled Vocabularies, Tools and DDI

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

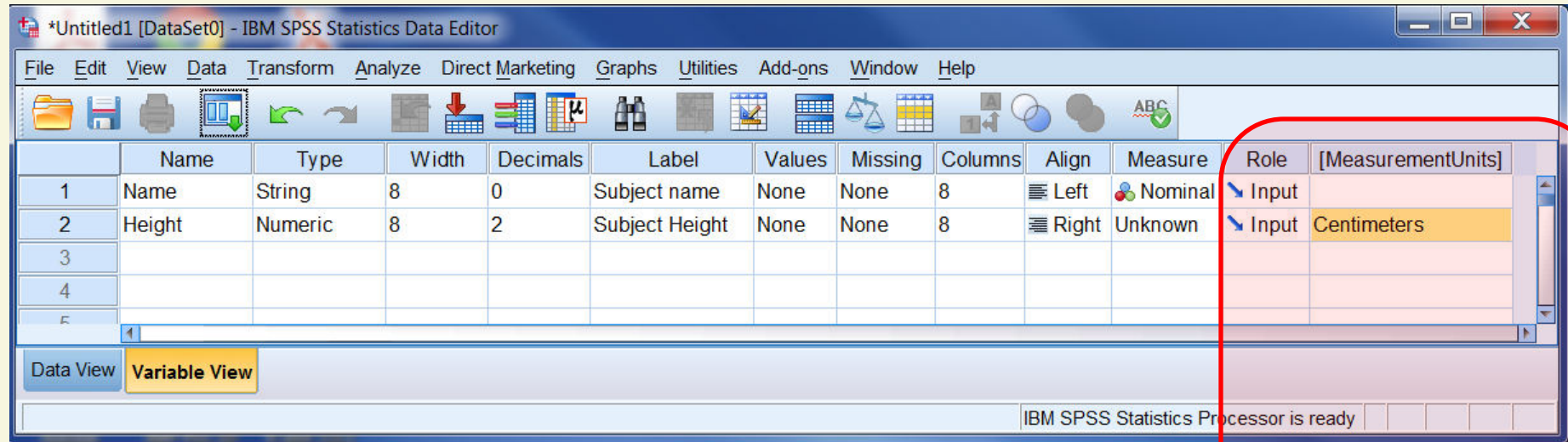
User Defined
Attributes

A Quick Tour – R

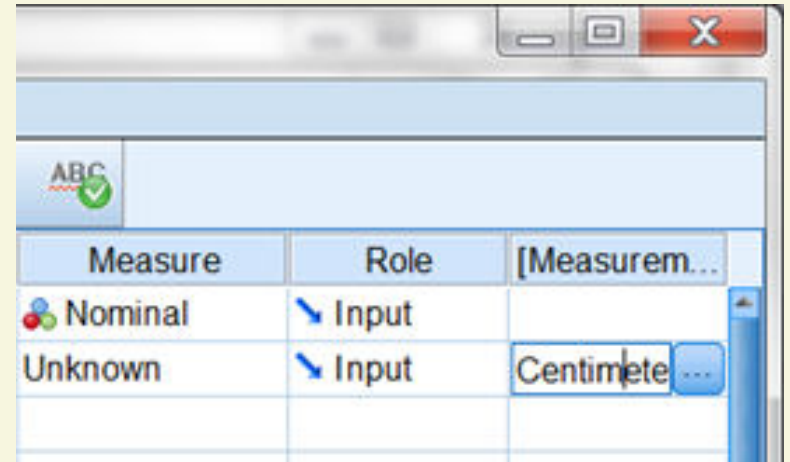
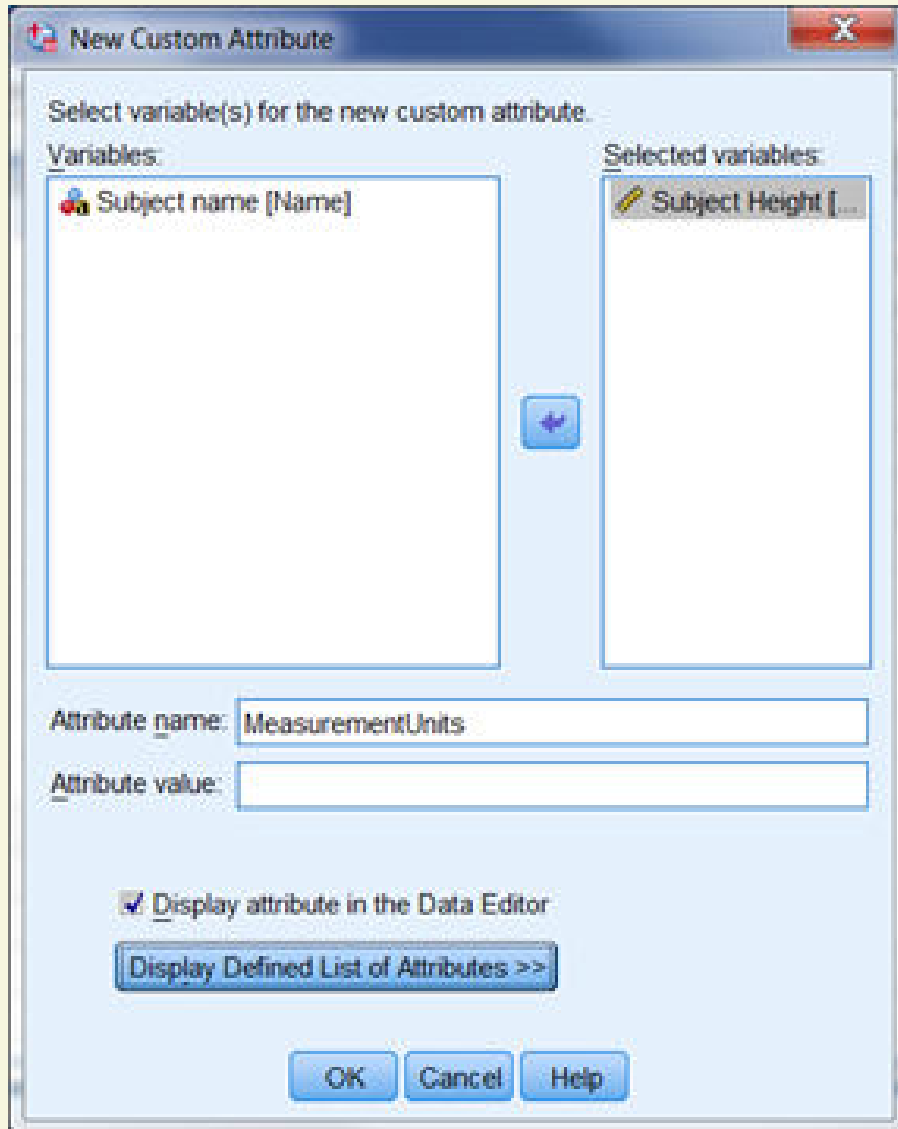
 R Console

```
> 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
```

<i>_dta[Universe]:</i>	<i>Persons aged 65 and over</i>
<i>Height[note1]:</i>	<i>First note on Height</i>
<i>Height[note0]:</i>	<i>1</i>
<i>Height[MeasurementUnits]:</i>	<i>Centimeters</i>

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

The screenshot shows a SAS dialog box titled "Explore or Set Extended Attributes". It has four tabs: "Edit", "Test SAS", "Properties", and "DDI31", with "Edit" currently selected. The dialog contains several input fields and buttons:

- Server:** A dropdown menu with "Local" selected.
- Library (Libname):** A dropdown menu with "EXTEST" selected.
- Input Dataset (Memname):** A dropdown menu with "SALES" selected.
- Output Dataset Name:** A text field containing "SALES_andXATTRs".
- Select a Variable:** A dropdown menu with "_DATASET_Attribute_" selected.
- Select or Enter an Extended Variable Attribute Name:** A dropdown menu with "Abstract" selected.
- Enter an Attribute Value for the Variable/Attribute Combination:** A large, empty text area.
- Checkboxes:** Four checkboxes are located at the bottom left:
 - ☐ DDI2.5
 - ☐ DDI3.2
 - ☐ DDI RDF
 - ☐ CDISC ???
- Buttons:** At the bottom right, there are three buttons: "EnterAttribute", "Delete Attribute", and "OK -Finish".

Enter, Change, or Delete Dataset Attributes

Select a Variable

_DATASET_Attribute_ ▼

Select or Enter an Extended Variable Attribute Name

Abstract ▼

Enter an Attribute Value for the Variable/Attribute Combination

This is an abstract for the dataset for the EDDI example

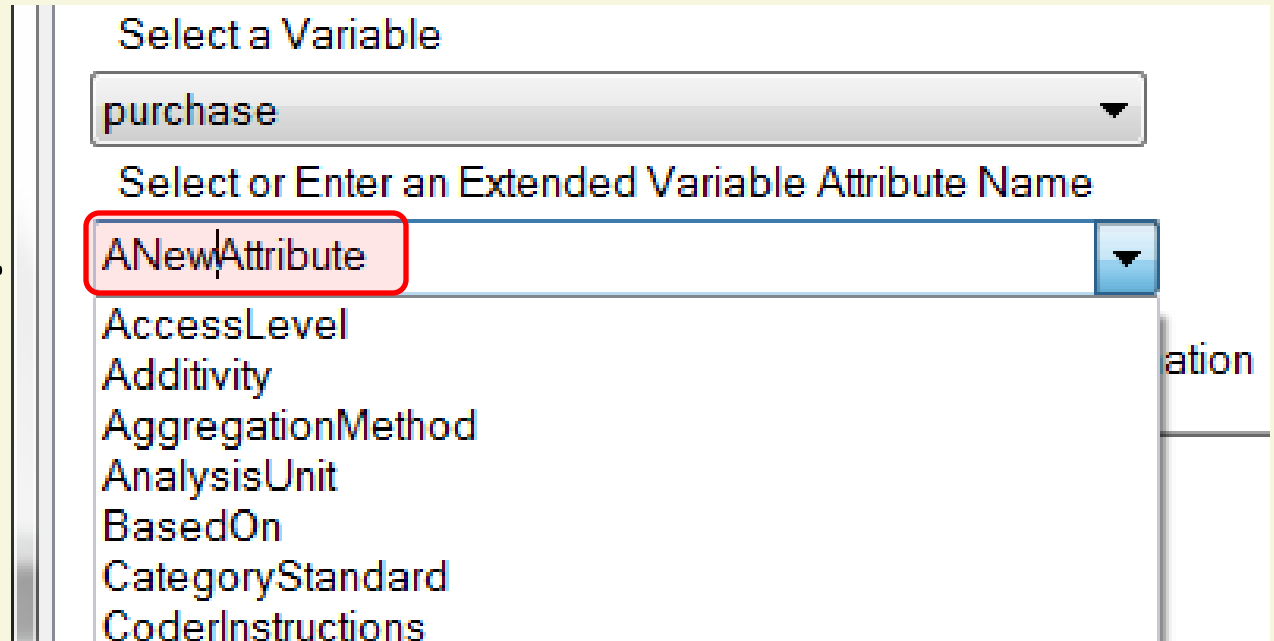
Attributes Prompted From a List

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

The screenshot shows a software interface with a light gray background. At the top, there is a label "Select a Variable" in blue text. Below it is a dropdown menu with the text "_DATASET_Attribute_" and a downward arrow. Underneath this is another label "Select or Enter an Extended Variable Attribute Name" in blue text. Below that is a list box containing a scrollable list of attribute names. The first item, "Abstract", is highlighted with a blue background. The list includes: Abstract, AccessRights, AlternativeTitle, AnalysisUnit, Authorization, Citation, CleaningOperation, CollectionMethodology, Contributor, Creator, DDfile, Description, Embargo, FundedBy, and Identifier. To the right of the list box is a vertical scrollbar. At the very top of the window, a portion of another dropdown menu is visible, showing the word "VALUES".

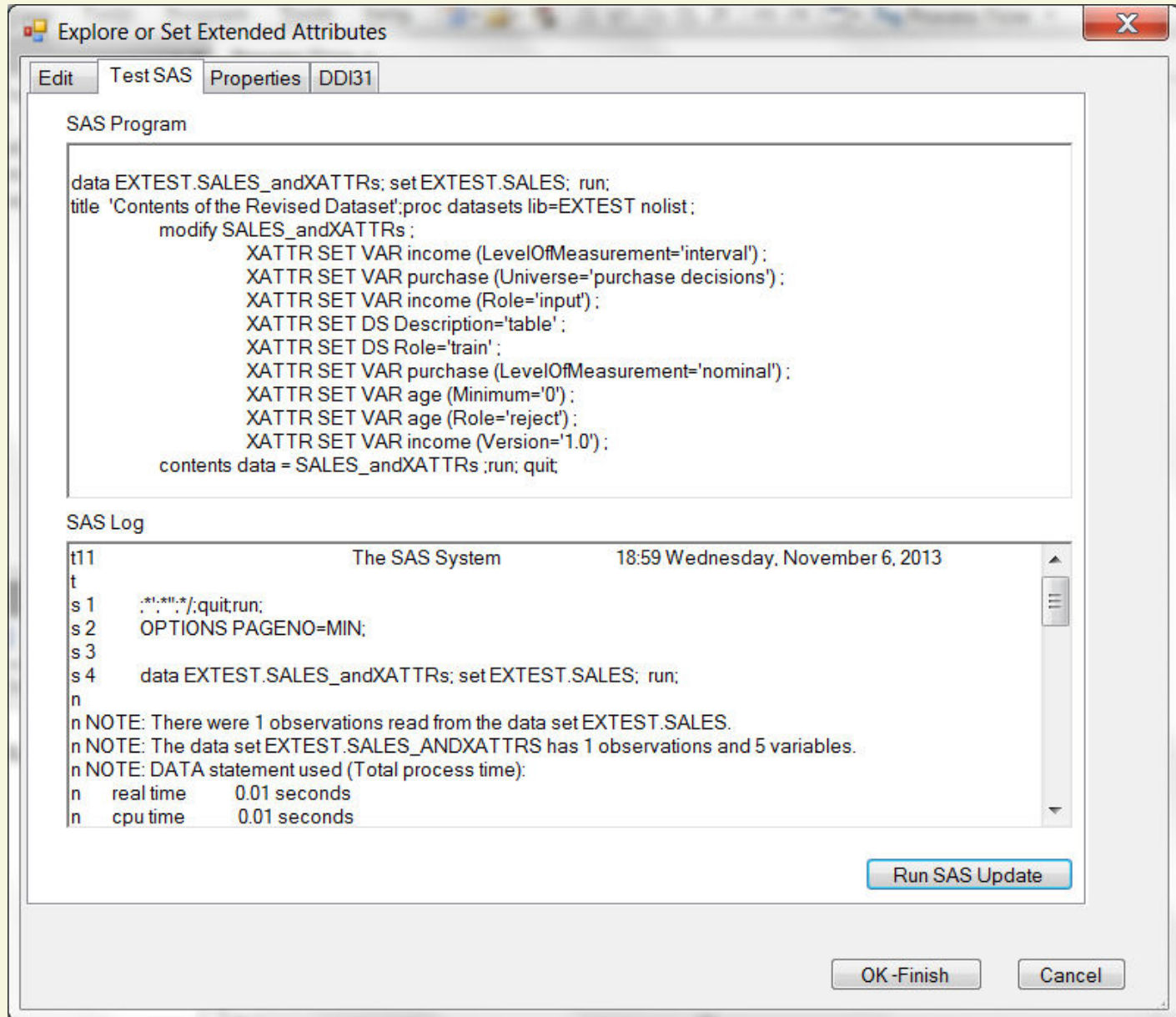
Or User Defined Attributes

A user can also enter a new attribute name.



The screenshot shows a software interface with two dropdown menus. The first dropdown, labeled "Select a Variable", has "purchase" selected. The second dropdown, labeled "Select or Enter an Extended Variable Attribute Name", has "ANewAttribute" entered and highlighted with a red box. A list of predefined attributes is visible below the second dropdown: AccessLevel, Additivity, AggregationMethod, AnalysisUnit, BasedOn, CategoryStandard, and CoderInstructions. The word "ation" is partially visible on the right side of the interface.

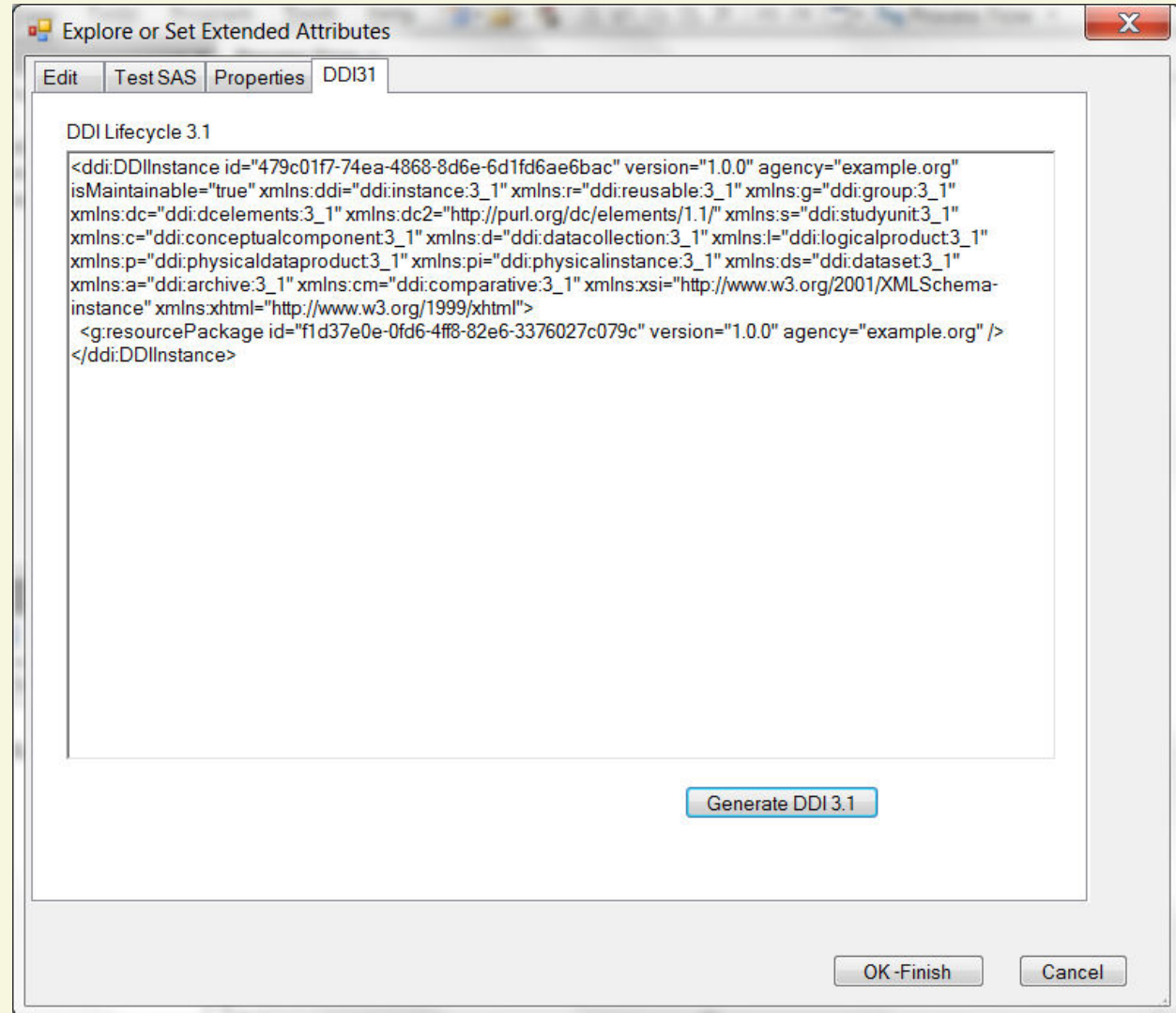
Runs SAS Procedure on Demand



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

Extended Attributes ▾

Input Data Code Log Output Data Results

Refresh Modify Task Export ▾ Send To ▾ Create ▾ Publish Properties

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

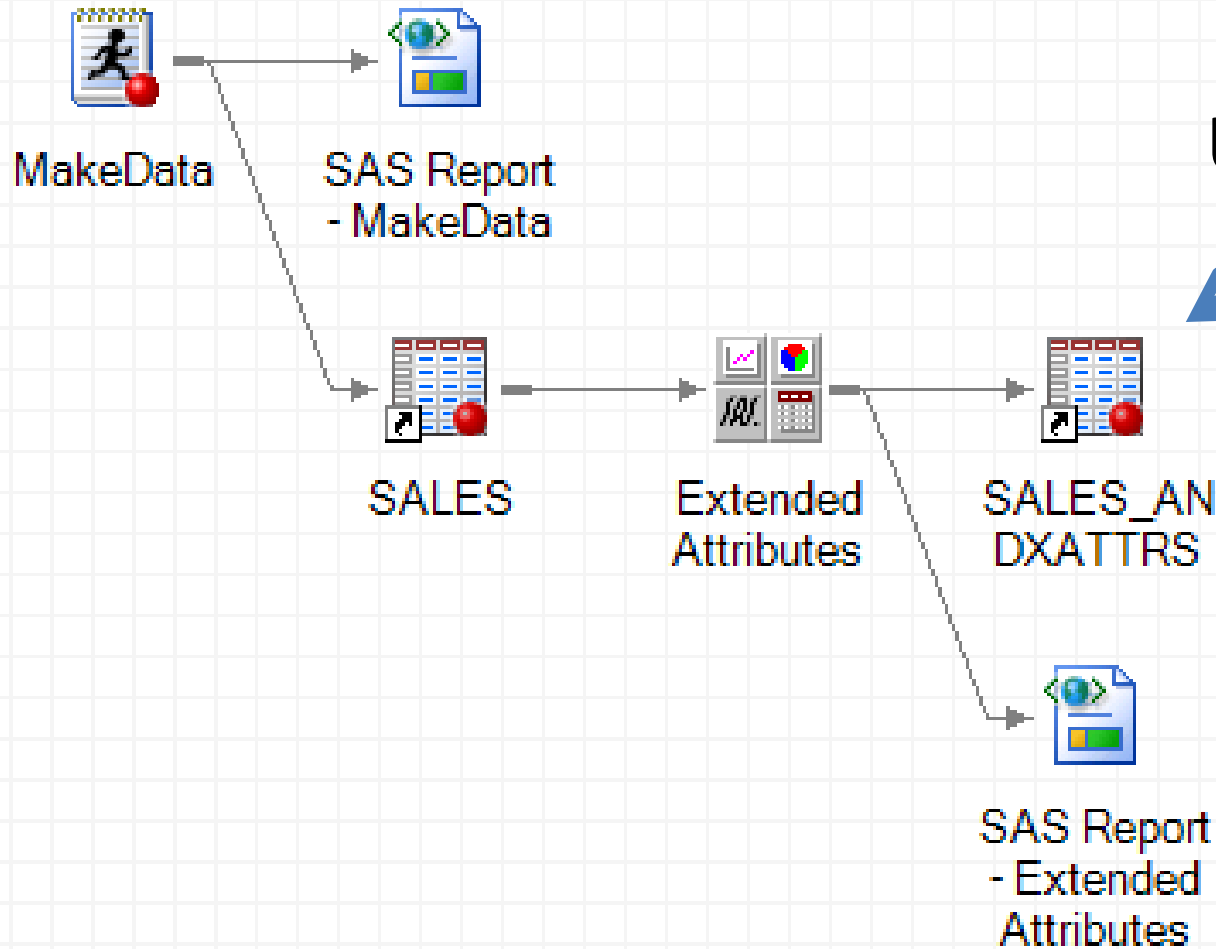
Alphabetic List of Data Set Extended Attributes

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