Integrating SAS® with ERwin® to Capture Design and Business Metadata

Jim Ferrari, Juan Olvera, Valencia College, Orlando, FL, USA

ABSTRACT

Do you have difficulty documenting architectural or design ideas? Are naming standards and consistency throughout the data warehouse difficult to maintain? Do related fields have unrelated names in different areas of the data warehouse? A modeling tool like ERwin® can assist with some of these issues. But how do you integrate a modeling tool like ERwin® with your SAS® BI environment?

Learn data architect and modeling techniques that will help refine your ideas. Learn how to integrate the metadata from ERwin® (or similar tool) into your SAS® environment. The metadata can be extracted into tables for reporting purposes (data dictionary, requirements, etc) or imported using a metadata bridge to populate the SAS® metadata directly.

INTRODUCTION

3rd party tools are available to assist with design and data modeling. ERwin® is a tool that captures data modeling metadata for logical and physical models. Data modeling is very helpful when architecting a data warehouse and ERwin® can integrate with SAS® BI.

DESIGN AND BUSINESS METADATA CAPTURED BY ERWIN

Prior to construction in the SAS® environment the data warehouse should be created logically. Stephan Covey (Author of "THE 7 HABITS OF HIGHLY EFFECTIVE PEOPLE") says in Habit 2 to begin with the end in mind. (https://www.stephencovey.com/7habits/7habits-habit2.php)

"Habit 2 is based on imagination--the ability to envision in your mind what you cannot at present see with your eyes. It is based on the principle that all things are created twice. There is a mental (first) creation, and a physical (second) creation. The physical creation follows the mental, just as a building follows a blueprint."

ERwin® provides an environment for the first creation, where Fact tables and Dimension tables can be modeled. A powerful design object available in ERwin® is the Domain; it is the parent object to a column in a table, which can inherent characteristics. This is invaluable when establishing consistent naming conventions across the entire data warehouse, both display and technical names.

In the higher education industry Term and Semester are very common variables. It is used in many fact and dimension tables and should have standardized naming. The availability of Domains and modeling before development makes this much easier.

Other useful metadata that ERwin® collects are variable definitions, notes, mapping to source field(s), and the SAS® library location of tables. This information can be extracted and populated in SAS® tables for reporting data dictionaries, source field mappings, calculations and even requirements.

ERwin® also has enough technical metadata that it can be imported into SAS® to create the data target. An ETL process can then be developed to populate the data target that was created by the ERwin® metadata.

BRIDGING ERWIN METADATA TO SAS

SAS® Metadata Bridges are available for SAS® Data Integration Studio 4.2 (DIS).

SAS® supports various Metadata Bridges for third-party vendor tools as listed at the following site:

http://support.sas.com/software/bridges/

SAS Technical Support provided the mapping crosswalk from ERwin® metadata to the SAS® metadata.

The ERwin® Model metadata can be imported into SAS® Metadata to create data targets. The ETL to populate the data targets can then be developed independently. This allows consistency from the design to the implementation phase of the ETL process. The imported metadata can also be compared against existing metadata in SAS®. Additionally, an impact or reverse impact analysis can be performed to check for differences in the SAS® data target versus the actual ERwin® Model metadata.

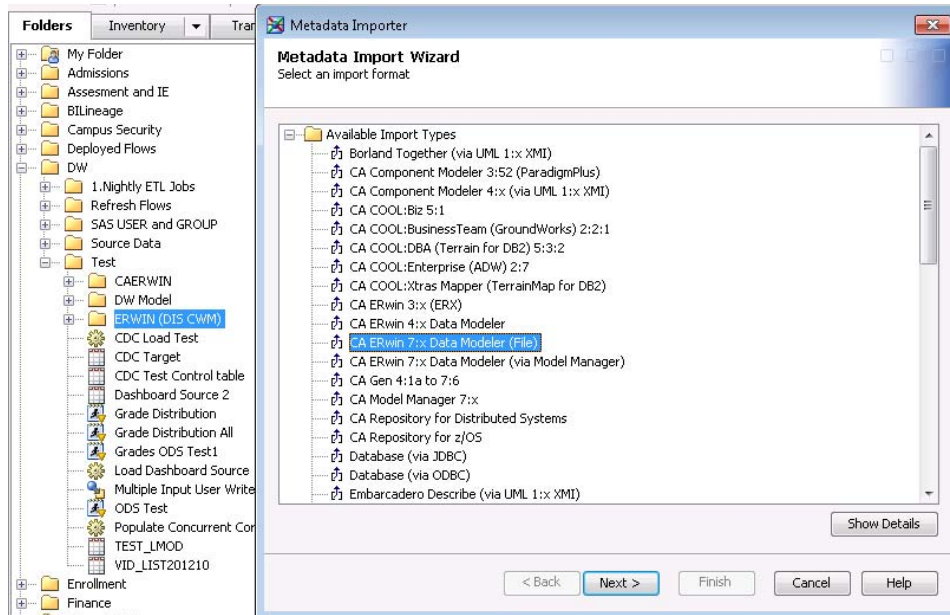Integrating SAS® with ERwin® to Capture Design and Business Metadata, continued

STEPS TO IMPORT CA ERWIN MODEL METADATA

1. Import metadata file using Import Metadata wizard tool.

    In DIS, right click on the target folder and select "Import" then "Metadata".
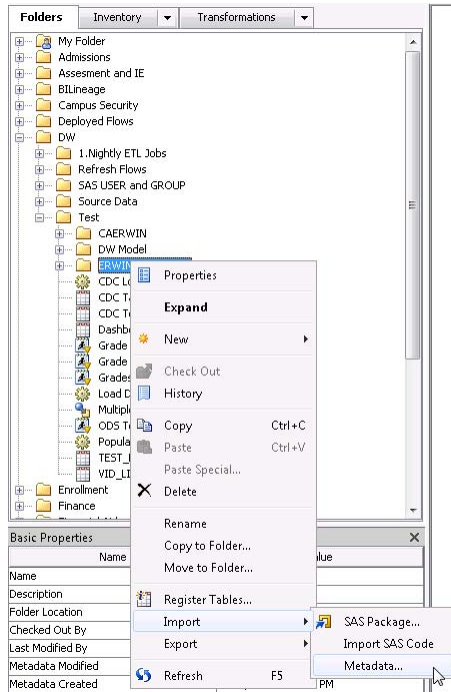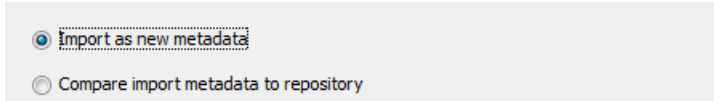

    From the Metadata Importer wizard, select the appropriate Metadata Bridge and it will guide you through the steps.


2. Import/Compare to existing metadata.

    The SAS® metadata import wizard will prompt for two choices when importing the file from ERwin®:

Integrating SAS® with ERwin® to Capture Design and Business Metadata, continued

Select "Import as new metadata" if the tables in the ERwin® model don't exist in the SAS® environment, otherwise you will want to select the "Compare import metadata to repository" option.

The comparison will return the following wizard.
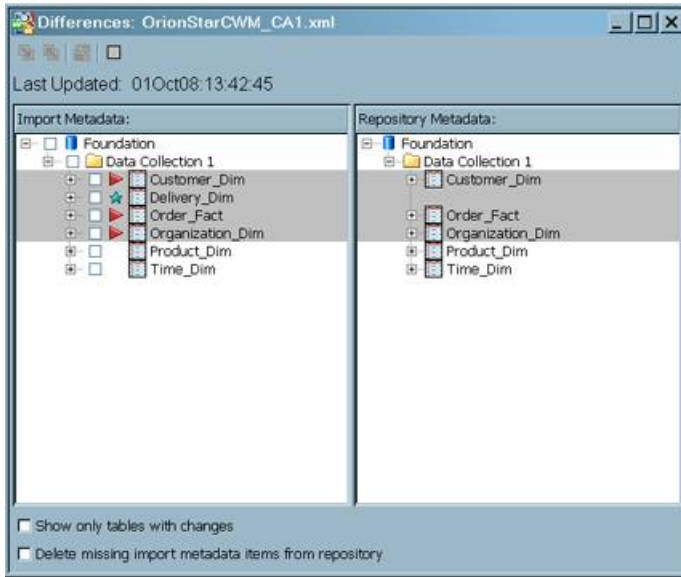
3. View changes and run impact analysis.

    An impact analysis can be performed by clicking on the item changed and use the reverse/impact analysis icons to apply the changes. Developers can then create ETL jobs to populate data from the source transaction system to the imported data target.
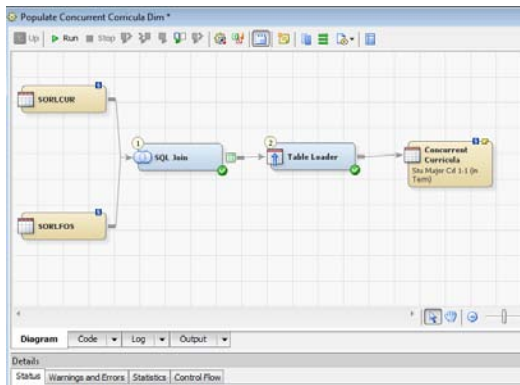
## USING ERWIN METADATA FOR REPORTING IN A SAS ENVIRONMENT

ERwin® has ODBC capability, but it is a PC based application. Most SAS® BI solutions are hosted on a server and don't have access to individual PCs, therefore extracting the metadata from ERwin® cannot be performed the same as other server based source systems. In order to extract data from ERwin® and populate it in SAS® tables, Enterprise Guide® (EG) and the ODBC connection needed to be utilized.

Once the data is extracted into SAS® tables it can be surfaced by any of the many SAS® reporting tools. The data can be used to create a data dictionary to be reference by end users or by developers for more technical information.
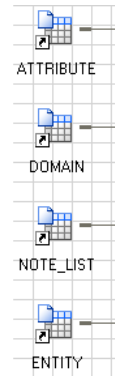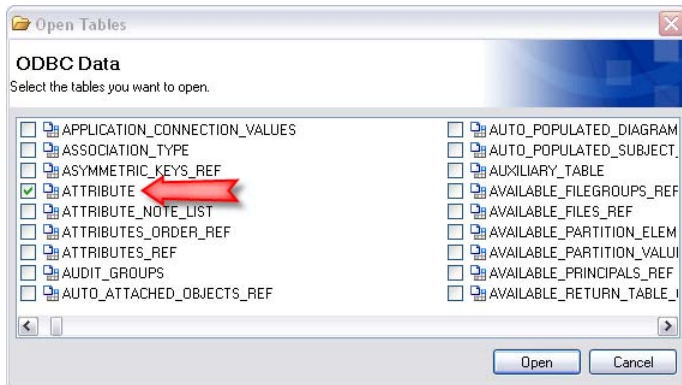
Integrating SAS® with ERwin® to Capture Design and Business Metadata, continued

## STEPS FOR LOADING ERWIN METADATA INTO SAS TABLES

1. You will need to have a target library, where you want to store the extracted ERwin® metadata, defined with the appropriate security settings for your identity.

2. Open the ERwin® model from which you wish to extract the metadata.

3. Open an ODBC connection in EG using the Machine Data Source of "ERwin_r8_Current" or "ERwin_Current" depending on the version of ERwin®.
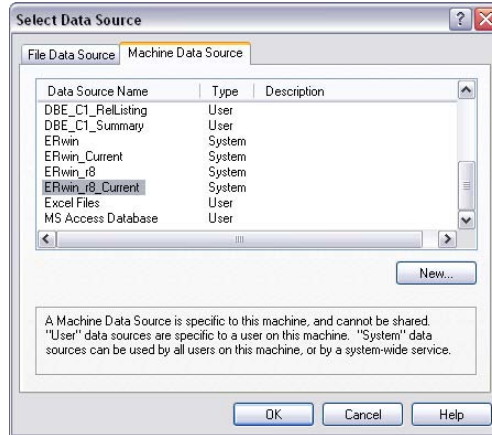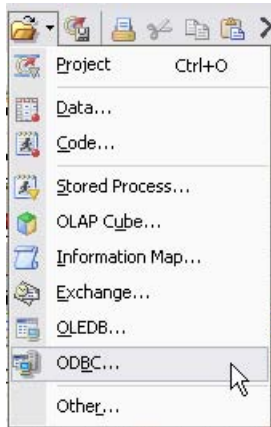
4. Check the box next to the ERwin® metadata table(s) that you wish to extract into your SAS® environment and click Open. This will add pointers to the tables into the EG project.


5. Create a simple 'libname' statement as the first step in the project to define where the ERwin® metadata will reside in the SAS® environment.
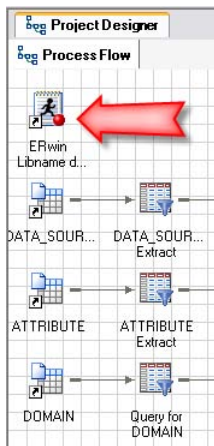

```
libname ER '/data2_dev/ERwin';
```

4

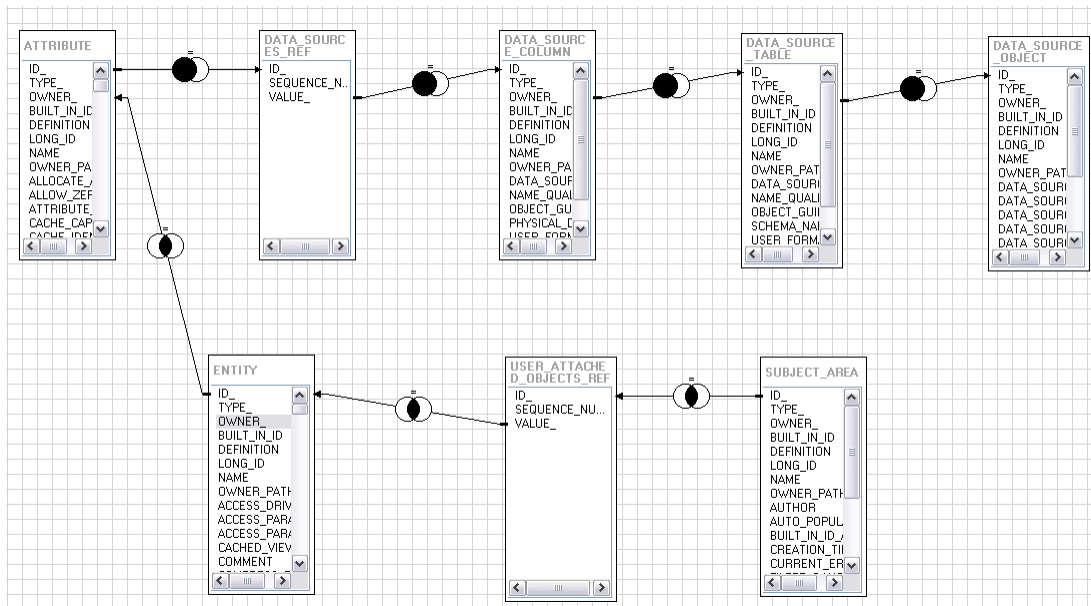Integrating SAS® with ERwin® to Capture Design and Business Metadata, continued

6. Using a Query task, select all columns from the ERwin® table and write them to a SAS® table with the same name in the Library that you created. Some ERwin® column (variable) names contain an "@" which will need to be changed to SAS® friendly character or removed.
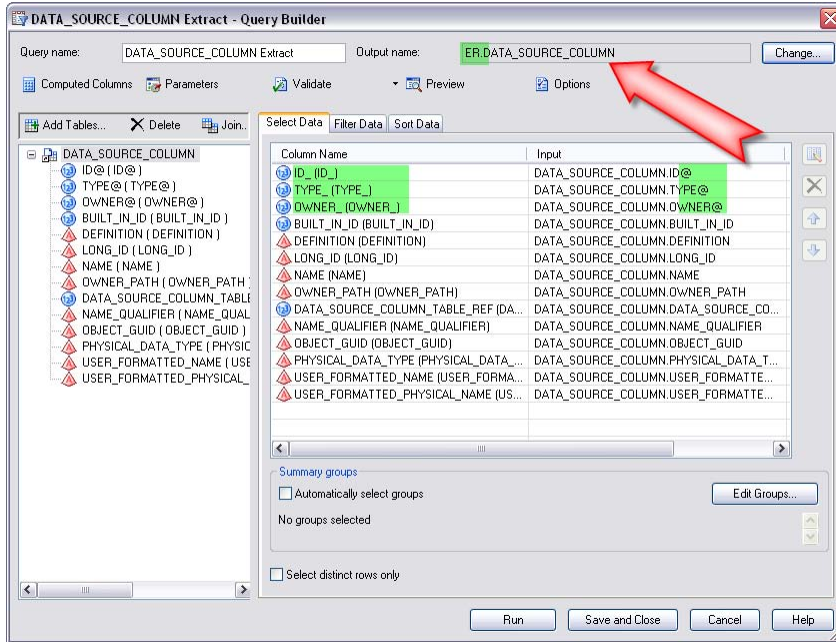
7. This process will need to be performed for each table that you want to extract from the ERwin® metadata model. You may want to include all tables and extracts in a single job that you can run whenever an update of the ERwin® metadata is needed.

REPORTING ON ERWIN METADATA

To report on ERwin® metadata tables you need to understand how they relate to each other. Here is one example of the relationships among the tables to create a repot that shows the source system fields for each column in the model.

5

Integrating SAS® with ERwin® to Capture Design and Business Metadata, continued

```
%_eg_conditional_dropds(WORK.DATA_SOURCE_MAPPING);

PROC SQL;
  CREATE TABLE WORK.DATA_SOURCE_MAPPING AS SELECT SUBJECT_AREA.USER_FORMATTED_NAME FORMAT=$100. label="Subject Area" AS SUBJECT_AREA_NAME,
        ENTITY.USER_FORMATTED_NAME FORMAT=$100. label="Entity" AS ENTITY,
        ATTRIBUTE.USER_FORMATTED_NAME FORMAT=$100. label="Attribute" AS ATTRIBUTE_NAME,
        ENTITY.NAME_QUALIFIER,
        ENTITY.USER_FORMATTED_PHYSICAL_NAME,
        ATTRIBUTE.USER_FORMATTED_PHYSICAL_NAME FORMAT=$100. label="Column" AS COLUMN_NAME,
        ATTRIBUTE.DATA_SOURCE_COMMENT,
        DATA_SOURCE_OBJECT.USER_FORMATTED_NAME FORMAT=$100. label="Source System" AS SOURCE_SYSTEM,
        DATA_SOURCE_TABLE.USER_FORMATTED_NAME label="Source Table" AS SOURCE_TABLE_NAME,
        DATA_SOURCE_COLUMN.USER_FORMATTED_PHYSICAL_NAME FORMAT=$100. label="Source Column" AS SOURCE_COLUMN_NAME
  FROM ERWIN.SUBJECT_AREA AS SUBJECT_AREA
      INNER JOIN ERWIN.USER_ATTACHED_OBJECTS_REF AS USER_ATTACHED_OBJECTS_REF ON (SUBJECT_AREA.ID_ = USER_ATTACHED_OBJECTS_REF.ID_)
        INNER JOIN ERWIN.ENTITY AS ENTITY ON (USER_ATTACHED_OBJECTS_REF.VALUE_ = ENTITY.ID_)
          INNER JOIN ERWIN.ATTRIBUTE AS ATTRIBUTE ON (ENTITY.ID_ = ATTRIBUTE.OWNER_)
            LEFT JOIN ERWIN.DATA_SOURCES_REF AS DATA_SOURCES_REF ON (ATTRIBUTE.ID_ = DATA_SOURCES_REF.ID_)
              LEFT JOIN ERWIN.DATA_SOURCE_COLUMN AS DATA_SOURCE_COLUMN ON (DATA_SOURCES_REF.VALUE_ = DATA_SOURCE_COLUMN.ID_)
                LEFT JOIN ERWIN.DATA_SOURCE_TABLE AS DATA_SOURCE_TABLE ON (DATA_SOURCE_COLUMN.OWNER_ = DATA_SOURCE_TABLE.ID_)
                  LEFT JOIN ERWIN.DATA_SOURCE_OBJECT AS DATA_SOURCE_OBJECT ON (DATA_SOURCE_TABLE.OWNER_ = DATA_SOURCE_OBJECT.ID_)
  WHERE SUBJECT_AREA.ID_ <> 14
  ORDER BY SUBJECT_AREA.USER_FORMATTED_NAME, ENTITY.USER_FORMATTED_PHYSICAL_NAME, ATTRIBUTE.COLUMN_ORDER, DATA_SOURCE_OBJECT.NAME, DATA_SOURCES_REF.SEQUENCE_NUMBER_;
QUIT;
```

Reports on all of the metadata that is entered into ERwin® can be generated. There are many different interesting reports that can be created; one would be a Data Dictionary report.

CONCLUSION

Using modeling tools can be very beneficial but accessing the information they contain can be challenging. Using these techniques can provide access to the information maintained in the modeling tool.

REFERENCES

SAS® Institute Inc. 2009. SAS Data Integration Studio 4.21 User's Guide. 51-61. Carey, NC, USA: SAS® Institute Inc.

Covey, Stephan. The seven habits of highly effective people : restoring the character ethic. March 2012.
https://www.stephencovey.com/7habits/7habits-habit2.php

ACKNOWLEDGMENTS

Thank you to Anja Fischer from SAS® Technical Support – Network Technologies for helping us to map to SAS® Metadata from Erwin®.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Jim Ferrari
Enterprise: Valencia College
E-mail: jrferrari@valenciacollege.edu
Web: http://valenciacollege.edu/oit/dw/

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.