2020

SAS Data Curation Primer

Qiong Xu
CUNY Queens College

DATA CURATION NETWORK

SAS Data Curation Primer

Format Overview

Topic	Description
File Extension[1]	.sas7bdat .sas7bcat .sas .xpt
MIME Type	application/x-sas
Structure	The structure of the SAS data set (.sas7bdat) • Data values in a SAS data set are arranged in a matrix/frame structure • Each data set contains a descriptor portion that includes details about a data set The structure of SAS programs[2] (.sas) • A SAS statement ends with a semicolon • A program starts with a keyword such as proc, and end with another keyword such as run/quit
Versions	9.4 (current)
Primary fields or areas of use	Healthcare, biology, agriculture, business intelligence, finance
Source and affiliation	SAS is developed by SAS Institute
Metadata standards	• Standards vary by field of research • The SAS Catalog (.sas7bcat), can store user-defined formats, such as variable level metadata
Key questions for curation review	• Where is the data located? What file format/formats was the data set saved? • What is the data about?

[1] More: SAS documentation

[2] SAS Certification Prep Guide: Base Programming for SAS 9, Fourth Edition. Retrieved from https://www.sas.com/storefront/aux/en/certpgbp/71337_excerpt.pdf.

	• Where did the data come from or who collected the data? • When was the data collected? How was the data collected? • Is there a readme file, data documentation file or other file which describes how to use the data set? • Is there a SAS code or program file for data importing, preparation and/or analysis? • Are there any comments included in the SAS code explaining how to process the code? • Can SAS data files (.sas .sas7bdat, etc.) be loaded into other software in addition to SAS?
Tools for curation review	SAS Universal Viewer SAS University Edition SAS OnDemand for Academics Python, pandas.read_sas or sas7bdat R, sas7bdat, rio or haven package Stat/Transfer
Date Created	March 9, 2020
Created by	Creator: Qiong Xu, Queens College of the City University of New York Mentor: Jenn Darragh, Duke University Contributors: Gin Corden and Susan Borda.
Date updated and summary of changes made	

Suggested Citation: Xu, Qiong. (2020). SAS Data Curation Primer. Data Curation Network http://hdl.handle.net/11299/216586

This work was created as part of the "Specialized Data Curation" Workshop #3 held at Washington University in St.Louis, MO on November 5-6, 2019. These workshops have been generously funded by the Institute of Museum and Library Services # RE-85-18-0040-18.

Table of Contents

Description of Format

.sas is a file extension for SAS programs.

.sas7bdat is a file extension for SAS data sets. SAS data sets (.sas7bdat) store data values and descriptor information.

- Data values are arranged in a matrix/frame structure
  - The rows are called observations/objects and the columns are called variables/ characteristics
  - Variables contain the data values for each observation.
- Descriptor information includes details about a data set
  - Data set name, data set type, and data set label
  - The names and attributes of all the variables
  - The number of observations in the data set
  - The date and time that the data set was created and updated.

Note: Extended attributes, which contain metadata for the SAS files, can be defined and preserved with the DATA step. When it is saved on disk, the data set has a new extension ".sas7bxat".

".sas7bcat" is a file extension for SAS catalogs. SAS catalogs contain multiple entries such as function key definitions, fonts for graphic applications, some of your selections from the Preferences dialog box, and other information from interactive windowing procedures.

.xpt is a file extension for transport file. See Library of Congress information.

SAS Application and Research Data Documentation

To learn better how SAS data set is generated, stored, shared and reused, the principal investigator of this primer conducted several semi-structured interviews. The interview questions were adapted from a prior primer[3]. Three interviewees from two universities in the United States completed their answers via email or face-to-face interview. Two of the three interviewees were faculty and one of them was a doctoral student in statistics. All the interviewees were researchers who used SAS for research and teaching.

1. Is SAS used by many researchers doing data analysis? How is SAS compared to similar statistical programs?

The interviewees indicated that SAS is used by many researchers or scholars who need to deal with large data sets, for example census data, health science data, biology data, agriculture data, etc. Therefore, SAS is most popular among public health scholars, like epidemiologists, biostatisticians, as well as agriculture researchers. Compared to other packages, SAS has the longest history and possibly the largest user group and hence SAS is the most developed package.

SAS originated from North Carolina State University in the 1960s. Back then, design of experiments and clinical trials both were very hot topics, so departments of agriculture, biology, etc. used SAS very often even now. In addition to research fields, SAS programmers also work at many companies, for example insurance companies, for processing claims, data entry centers, finance industry, etc.

There are several factors making SAS a popular and powerful data analysis tool. First, SAS has a module for all different types of analysis, graphics, etc. SAS is powerful in building generalized linear (mixed) and linear (mixed) models, design of experiments, repeated measure analysis, sequential analysis and longitudinal data analysis, as well as data management. Even popular software R does not have such powerful tools in these areas.

Secondly, SAS has large storage to save data that can save computer memory and make data processing smoother. For a large data set, like a 300 GB national data set, only SAS can directly read and analyze it without occupying a huge computer memory. Using other packages, a researcher even may not be able to load the data due to the limit of computer memory. In the meantime, SAS is famous for its information security.

Last but not the least, SAS programming can make researchers analyze data more efficiently. Using SAS programming, researchers don't need to frequently point and click operational menus. SAS programming gives the researchers the opportunity to know exactly what and how SAS is processing.

[3] Deng, Sai; Dull, Joshua; Finn, Jeanine; Khair, Shahira (2019). SPSS Data Curation Primer. Data Curation Network GitHub Repository. https://github.com/DataCurationNetwork/data-primers

2. What kind of data do researchers generate on import into SAS?

Nearly any format of data file can be read into SAS. SAS can import various types of data, Excel files, CSV files, even DBF files. SAS Enterprise Guide can help with importing data and then exporting it as a SAS datafile.

3. Considering SAS data sharing and/or reuse via data repositories, how do researchers document their data in SAS, or what related data files need to be documented from SAS?

SAS can import and export nearly any data format, but other software packages can barely read SAS codes or SAS data files (.sas, .sas7bdat, etc.) directly. To store and share SAS data via a data repository, researchers usually create three files – a metadata file, a code/syntax file, and a data set file. Researchers usually store data into metadata with labels, correct formats, correct variable types, comments, etc. SAS metadata usually contains notes, information for all the variables (i.e., variable names, variable labels, the range of values, data type, etc.) and even some descriptive statistics (i.e., mean, standard deviation, sample size, min, max, missing data, etc.) SAS data set can be opened in the SAS program directly or be read into SAS software by running SAS statements (code/syntax).

4. What kind of SAS data outputs are researchers able or willing to share? (Considering data sharing and/or reuse via data repositories, what kind of SAS data outputs/files are researchers able or willing to share?)

Researchers may be willing to share their data openly depending on the data field. Also, data sharing really depends on the original agreement made by researchers and research agencies who funds the data collection. For data sets containing confidential information such as records of clinical trials, researchers can share data in a controlled way if the risk of identity disclosure still exists. In recent years, more and more researches are required by funders and publishers to share data via repositories as a condition for publishing an article. For instance, many researchers who receive research funds from NIH, National Science Foundation, Federal Money, etc. are usually obligated to disseminate the data, which are shared after the data are de-identified to preserve confidentiality protections.

After researchers collect the data, they are required to remove personally identifiable information or protected health information from the data to minimize the risk of the identification of individual respondents before the data is ready for sharing. In other words, researchers should share data in certain ways if they have an agreement with research funders and publishers.

SAS File Examples

SAS data is comprehensively used in health science, biology, agriculture, business and social sciences.

In addition to research fields, SAS programmers also work at many companies, for example insurance companies for processing claims, data entry centers, finance industry, etc.

Examples below link to OSF, DataLumos and CDC which offer SAS data files for data curation (see Figures 1, 2 and 3).

There might be no a perfect example for SAS data curation due to various limitations, but the Biological Psychology Data, the State Library Administrative Agencies survey (SLAA) and the National Health Interview Survey (NHIS) created good sample SAS data files for data sharing and reusing.

There are a variety of files can be documented for data sharing. In general, the essential documentation should include three types of documents: (1) a data set in more than one format; (2) a SAS code/program file for data importing, preparation or analysis; and (3) an instructional file, data documentation file or linkable publication that provides information about variables, data analysis, and/or how to use the data set (see Figures 1, 2 and 3).

In addition to a SAS data file (.sas7bdat), the data set should be provided with at least one of the following formats to make the data set reusable in other software environments.

- *CSV*
- *TXT*
- *ASCII*
- *MDB/DBF*

The SAS code/program file (.sas) is a syntax file which can be used to import data, prepare data (i.e., rename variable names, recode variables, compute variables, etc.), and repeat and validate data analysis.

A file providing information about how to use the data set can be one of the following documents.

- *Readme file (.pdf)* directs users how to use the data file. It may contain such contents as data use restriction, data collection instrument (i.e., questionnaire) structure, description of data files and their formats, and how to use the data set in different formats.
- *Data documentation file (.pdf)* provides such information as research background and purpose, research methodology, data collection and processing, variable description, how to use the data file, etc.
- *A linkable publication (.pdf)* is the original publication using the data set, which provides detailed information about the data.

In addition to the above three essential data documents, the National Health Interview Survey (NHIS) provides more documents which may be very useful for depositing a large data set. For example,

- *Variable summary (.pdf)* is a metadata or variable dictionary file, providing more details about to the variables.
- *Variable Layout (.pdf)* is a codebook file, providing labels of each numerically categorized variable.
- *Variable frequencies file (.pdf)* provides descriptive statistics of the variables, such as frequency, percent, missing values, etc.

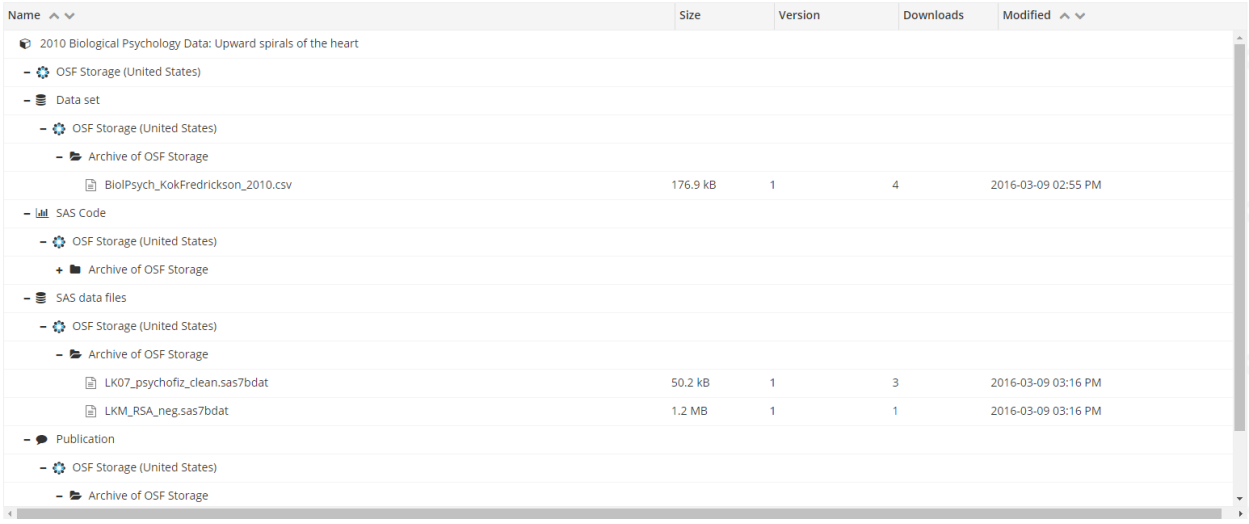

Figure 1: SAS data documentation provided by Open Science Framework.

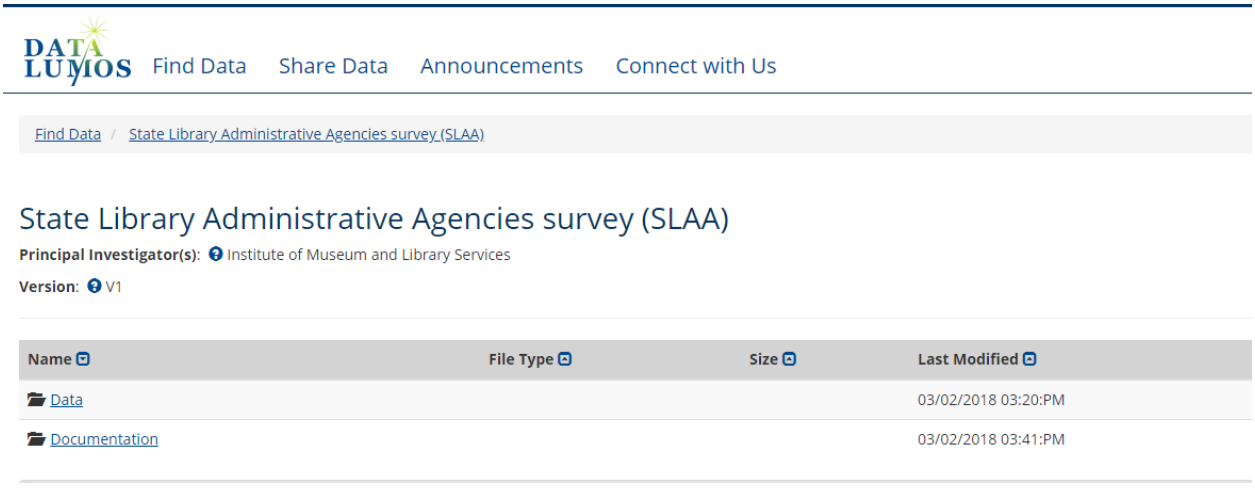

Figure 2: SAS data files provided by DataLumos.

Figure 3: SAS data documentation provided by [National Center for Health Statistics and Centers for Disease Control and Prevention](#).

Key Questions to Ask Yourself

- Where is the data located? What file format/formats was the data set saved?
- What is the data about?
- Where did the data come from or who collected the data?
- When was the data collected? How was the data collected?
- Is there a readme file, data documentation file or other file which describes how to use the data set?
- Is there a SAS code or program file for data importing, preparation and/or analysis?
- Are there any comments included in the SAS code explaining how to process the code?

Key Clarifications to Get from Researcher

It is noticed that SAS data can be documented in different formats to facilitate data sharing and reusing. In addition to SAS data format (.sas7bdat), a SAS data set can be saved as plain text file (i.e., .csv, .txt, .dat, etc.) or database file (.mdb or .dbf).

- A SAS data file (.sas7bdat) can be opened and viewed directly with SAS installed (see Figure 4).
- A text data file (.csv or .txt) and a database file (.mdb or .dbf) can be directly imported into SAS software using SAS import Wizard (see Figure 5).
- A column-delimited ASCII data file (.dat) can be imported into SAS software running SAS statements (.sas) (see Figure 6).

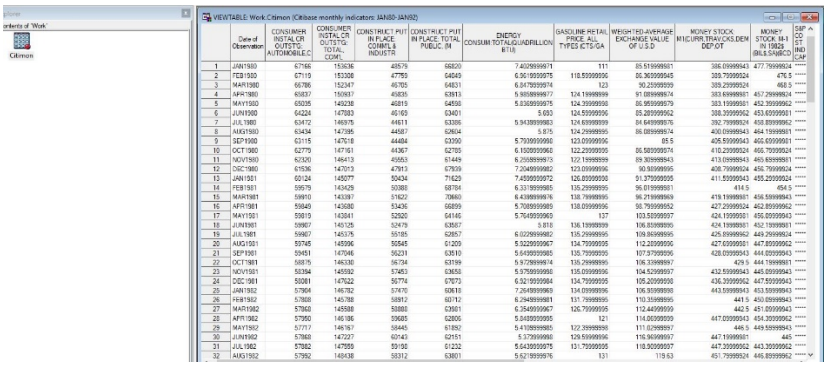

Figure 4: Directly open a SAS data set (.sas7bdat) and view the data with SAS.

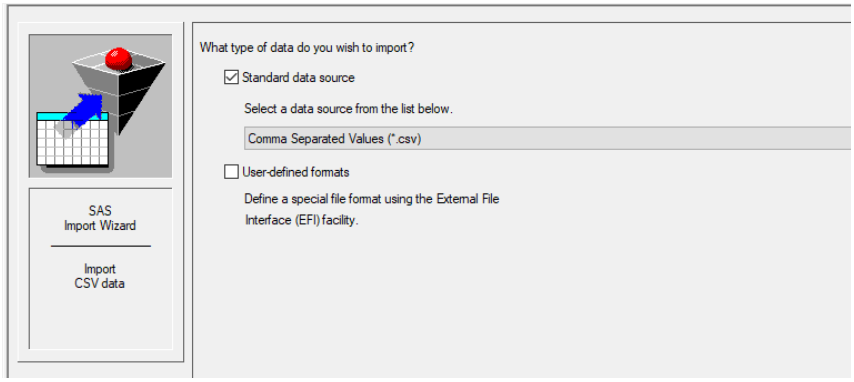

Figure 5: Use SAS Import Wizard to import a data set in CSV format into SAS.

```
*******************************************************************
 APRIL 13, 2018 11:39 AM

 This is an example of a SAS program that creates a SAS
 file from the 2017 NHIS Public Use FAMILYXX.DAT ASCII file

 This is stored in FAMILYXX.SAS
*******************************************************************;

* USER NOTE: PLACE NEXT STATEMENT IN SUBSEQUENT PROGRAMS;
LIBNAME  NHIS     "C:\NHIS2017";

* USER NOTE: PLACE NEXT STATEMENT IN SUBSEQUENT PROGRAMS
            IF YOU ALLOW PROGRAM TO PERMANENTLY STORE FORMATS;
LIBNAME  LIBRARY  "C:\NHIS2017";

FILENAME ASCIIDAT 'C:\NHIS2017\FAMILYXX.dat';

* DEFINE VARIABLE VALUES FOR REPORTS;

*  USE THE STATEMENT "PROC FORMAT LIBRARY=LIBRARY"
     TO PERMANENTLY STORE THE FORMAT DEFINITIONS;

*  USE THE STATEMENT "PROC FORMAT" IF YOU DO NOT WISH
     TO PERMANENTLY STORE THE FORMATS;

PROC FORMAT LIBRARY=LIBRARY;
*PROC FORMAT;

   VALUE $GROUPC
      ' '< - HIGH    = "Range of Values"
   ;
   VALUE GROUPN
      LOW - HIGH    = "Range of Values"
   ;
```

Figure 6: Run SAS statements to extract a data set in an ASCII file (.dat) into SAS.

Based on the interview results and the above examples, we suggest that at least three data documents should be prepared for SAS data curation - (1) a Readme file or a data

documentation, (2) a data set in SAS format (.sas7bdat) and in CSV format (.csv), and (3) a SAS code/program file (.sas) to import, prepare or validate the data set.

Applicable Core Elements of Metadata and Readme Requirements

Core elements of Metadata

The fundamental questions about SAS data indicated that the core elements of SAS metadata include data location, content and purpose of the data, data ownership and collection procedure, etc.

Readme requirements

A "Readme" or a data documentation file includes important metadata information, such as what the data is about, how the data is used, the meaning and context to the piece of data, etc. Rich and accurate metadata information can facilitate the usage of SAS data sets. As a result, a Readme file or a data documentation file can include the following elements.

Title: The title of a readme file should match the name of the data set

Author(s):

- Name(s) of researcher(s) (e.g., PI and all co-PIs) or organization(s)
- Contact information (mailing address, telephone/facsimile numbers, and E-mail address of PIs)

Data file overview and data format:

- Brief introduction about Readme or data documentation file
- Brief description about the format(s) of data set
- Brief introduction about SAS code/program file
- Variable definition/description

Instrument description:

- Brief description about the research instrument

Data collection and processing:

- Description of data collection procedure
- Description of data processing techniques
- Assessment of the data (e.g., instrument problems, quality issues, etc.)
- Software to view/process the data

References:

- List of documents cited in the Readme or the data documentation file.

Notes for Data Users

Since users may need more information about the data set or need to look at other relevant data sets, it is a good practice to provide such information and links to relevant data sets.

Resources for Reviewing Data

There are rich instructional and tutorial resources guiding users to review data. The procedure to review a SAS data set may include such steps as (1) creating/processing a SAS data set; (2) checking the structure of the data set; (3) exploring descriptive statistics of the variables, outliers and missing values, etc.

- Step-by-step instructions for reviewing a SAS data set using downloaded SAS files (i.e., SAS codes, data files, etc.)
- Tips and techniques for looking at SAS data files
- Step-by-step instructions for exploring a SAS data set[4]
  - Summarizing Data with PROC CONTENTS:
  - Viewing a data set using the Viewtable Window or Printing a data set to the Output Window with PROC PRINT
  - Check frequencies and missing values using PROC FREQ
- Reading data into SAS tutorial (video)

Software for Viewing or Analyzing Data

In addition to SAS package (see the "Key clarifications to get from researcher" and "Preservation actions" sections in this primer), a SAS data set (.sas7bdat file) can be imported into the following software with coding or special tools.

R:

- Read SAS data set (.sas7bdat) into RStudio using sas7bdat, rio, or haven package
- Import SAS Transport files into R using the function read.xport ().

Python:

- Read a SAS XPORT or SAS7BDAT file into Python using the method pandas.read_sas.

See Appendix A "Additional Information on Tools" for more information.

Preservation Actions

A SAS data set can be preserved in a data repository with different formats **(**.sas7bdat, csv, ascii, etc.). Before loading the data set to the repository, it's recommended to check for correct

[4] Kent State University Libraries. (2017, May 22). SAS tutorials: Subsetting and splitting data sets. Retrieved from http://libguides.library.kent.edu/SAS/SubsetData.

variables and value labels, and to verify them with Readme through reading/browsing the data set. There are two ways to do so.

- With a SAS software (i.e., SAS 9.4) installed (i.e., in Windows), a SAS data file (.sas7dbat) can be read directly by double clicking the data file, or right clicking the file and selecting "browse with SAS 9.4."
- Run the code (see Figures 7 and 8) to load data files in different formats into a SAS Work library for checking.

```
/*Use DATA step to save a SAS dataset to the temporary Work library.*/
DATA Work.STLA01;
SET 'C:\SLAA\DATA\STLA01.sas7bdat';
run;
```

Figure 7: Load a SAS dataset (.sas7bdat) into SAS.
Data source: http://doi.org/10.3886/E101764V1.

```
* Bethany Kok- February 24, 2016;

/*1. Use the PROC IMPORT statement to import the dataset "BiolPsych_KokFredrickson_2010.csv";
     the DATAFILE argument is required to tell SAS where to find this data file;
     the OUT argument is used to output the dataset with a defined name to the WORK library;
     the DBMS option is used to indicate the type of data file(.csv)imported;
     the replace option will overwrite an existing file.
  2. Use GETNAMES statement to specify whether the IMPORT procedure generates SAS variable names
     from the data values in the first row in the input file.*/

proc import datafile="C:\DATASET\BiolPsych_KokFredrickson_2010.csv" out=ush dbms=csv replace;
     getnames=yes;
run;
```

Figure 8: Import a CSV dataset into SAS.
Data source: https://osf.io/zma9h.

Running the IMPORT procedure listed in the first part of the SAS code file (see Figure 8), a curator can check if the code executes. When running the code, be sure to replace the original data directory with the correct directory on the curator's machine or environment where the data set is saved.

If other SAS procedures, such as renaming variable names, dropping certain columns, step-by-step data analysis, etc., are included, it's recommended for the curator to check if the code for each procedure works. But the curator can decide on if this action is necessary depending on patrons' request.

For the purposes of facilitating data reuse and data preservation with more longevity, it is recommended that in addition to a SAS7BDAT file, a SAS data set should be saved as plain text formats (i.e., .csv, .dat or .txt) in a data repository. We can save a data set into a text format using the SAS "Export Data" Wizard or running a SAS program (see Figures 9 and 10).

```
*Save a SAS data file (.sas7bdat) as a CSV file;
PROC EXPORT DATA= WORK.Lk07_psychofiz_clean
            OUTFILE= "C:\DATASET\Lk07_psychofiz_clean.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;
```

Figure 9: Save a SAS data file (.sas7bdat) as a CSV file.
Data source: https://osf.io/zma9h.

```
*Save a SAS data file (.sas7bdat) as an ASCII file;
data _null_;
SET work.LK07_psychofiz_clean;
FILE  'c:\dataset\LK07_psychofiz_clean.dat';
put (_all_) (:);
list;
run;
```

Figure 10: Save a SAS data file (.sas7bdat) as an ASCII file.
Data source: https://osf.io/zma9h.

Documentation of Curation Process

SAS data sets can be saved in the directories of a local computer or a permanent SAS library. Researchers can save their SAS data sets in different formats. In addition to SAS7BDAT file, researchers can export and save their data in CSV (.csv), ASC II (.dat), etc.

For SAS data curation, such files need to be captured as (1) a data set in SAS format (.sas7bdat), ASCII text format (.dat), and/or CSV format (.csv); (2) metadata documentation that includes such information as title, author, variable description, how to use the data set, etc.; (3) a SAS code/program file (.sas) for data importing, preparation or analysis.

In addition, it is a good practice to document a file of notices to users that provides more information relevant to the data set. For example, where to find more information or data sets relevant to the current data set, etc.

Appendix A Additional Information on Tools

Application/Package	SAS Filetypes works with	SAS Version	Notes
SAS Universal Viewer	Data sets & libraries .xpt	Versions 7+ Version 5	Free

			Windows application Allows viewing, sorting, and filtering, and saving as .csv[5]
SAS University Edition			Free for student/academic use Virtual application, any OS
SAS OnDemand for Academics			Free access to SAS Studio for non-commercial use Browser based
R, rio, sas7bdat, haven	.sas7bdat & .sas7bcat .xpt	Versions 7+ Versions 5 & 8	See especially convert() function in rio, read.sas7bdat () function in sas7bdat, and read_sas () function in haven
Python, pandas.read_sas	.sas7bdat .xpt	Versions 7+ Version ?	
Python, sas7bdat	.sas7bdat	Versions 7+	Includes command line script, sas7bdat_to_csv
Stat/Transfer Data, transport, cport (read only)		Versions 6.08+	Not free
Stat/Transfer	Data, transport, cport (read only)	Versions 6.08+	Not free
Other resources			
SPSS (GET SAS DATA)	Data, transport	Versions 6+	Not free
Stata (import sasxport5/8)	.xpt	Versions 5 & 8	Not free

Appendix B SAS Data File CURATED Checklist

CHECK Step

CURATE Action	Curator Checklist
Check data files and read documentation	• Files open as expected ○ Issues _____

[5] Highlight entire table contents by clicking in upper left cell, right-click, and select "Save As…"

• Review the content of the data files (e.g., open the data file (.sas7bdat) or run the code for data IMPORT procedure).
• Verify all metadata provided by the author and review the available documentation.
	• Code runs as expected
	  ○ Produces minor errors
	  ○ Does not run and/or produces many errors
	  ○ Did not try to run code
	• Metadata quality is rich, accurate, and complete
	  ○ Metadata has issues _____
	• Documentation Type (circle)
	Readme / Codebook /
	Data Dictionary / Other:_____
	  ○ Missing/None
	  ○ Needs work

UNDERSTAND Step

CURATE Action	Curator Checklist

Understand the data (or try to)
• Check for quality assurance and usability issues such as missing data, ambiguous headings, code execution failure, and data presentation concerns
• Try to detect and extract any "hidden documentation" inherent to the data files that may facilitate reuse.
• Determine if the documentation of the data is sufficient for a user with similar qualifications to the author's to understand and reuse the data. *If not, recommend or create additional documentation (e.g., a readme.txt template).*
	• SAS code is included
	  ○ Yes
	  ○ No
	• The SAS code's comments clearly describe each step/procedure
	  ○ Yes
	  ○ No
	• Clear variable names
	  ○ Documentation describes variable names
	  ○ SAS code renames the variable names
	  ○ Comments describe variable names
	  ○ Missing/None
	  ○ Needs work
	• Clear SAS steps/procedures
	  ○ Each SAS step/procedure is properly ended
	  ○ Comments describe code actions
	  ○ Documentation describes code actions
	• Review Documentation (in previous step, CHECK) for completeness and clarity

REQUEST Step

CURATE Action	Curator Checklist
Request missing information or changes • Generate a list of questions for the data author to fix any errors or issues.	Narrative describing the concerns, issues, and needed improvements to the data submission. • Inquiry sent to researcher • Response received • Additional follow up communication needed

AUGMENT Step

CURATE Action	Curator Checklist

Augment the submission
• Enhance metadata to best facilitate discoverability.
• Create and apply metadata for the data record, including descriptive keywords.
• When appropriate, structure and present metadata in domain-specific schemas to facilitate interoperability with other systems.
	• Discoverability sufficient
	  ○ Recommend (circle one) full-text index / file rename / file reorder / file descriptions / zip files into one archive Other _____
	• Keywords Sufficient
	  ○ Suggestions _____
	• Linkages Sufficient
	  ○ Link to report/paper
	  ○ Link to related data sets
	  ○ Link to source data
	  ○ Link to other _____

TRANSFORM Step

CURATE Action	Curator Checklist

Transform file formats
• Identify specialized file formats and their restrictions (e.g., Is the software freely available? Link to it or archive it alongside the data).
• Transform files into open, non-proprietary file formats that 2 broaden the potential audience for reuse and ensure that preservation actions might be taken by the repository in later steps. Retain original files if data transfer is not perfect.
	• Preferred file formats in use
	  ○ Recommend conversion from _____ to _____
	  ○ Retain original formats
	• Software needed is readily available
	  ○ Unclear version of software
	  ○ Unclear software used
	• Visualization of data easily accessible
	  ○ Recommend graphical representation _____
	  ○ Recommend web-accessible surrogate _____

EVALUATE Step

CURATE Action	Curator Checklist

Evaluate and rate the overall data record for FAIRness[6].
• Score the data set and recommend ways to increase the FAIRness of the data and become “DCN approved."
	• Findable
	  ○ Metadata exceeds author/ title/ date.
	  ○ Unique PID (DOI, Handle, PURL, etc.).
	  ○ Discoverable via web search engines
	• Accessible-
	  ○ Retrievable via a standard protocol (e.g., HTTP).
	  ○ Free, open (e.g., download link).
	• Interoperable
	  ○ Metadata formatted in a standard schema (e.g., Dublin Core).
	  ○ Metadata provided in machine-readable format (OAI feed).
	• Reusable
	  ○ Data include sufficient metadata about the data characteristics to reuse
	  ○ Contact info displayed if the direct assistance of the author needed.
	  ○ Clear indicators of who created, owns, and stewards the data.
	  ○ Data are released with clear data usage terms (e.g., a CC License).

DOCUMENT Step

CURATE Action	Curator Checklist
Document throughout curation activities. • Record all necessary information capturing who did what to the data set and when	• Accessioning & deposit records (Names, dates, contact information, submission agreements, etc.) • Repository collection metadata • Provenance logs • Service workflow • Preservation packaging • Any additional requirements at your institution

[6] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18