

Using SAS® to Prepare Postsecondary Data Partnership (PDP) Data Submission Files

Bobbie J. Frye, Achieving the Dream

ABSTRACT

Transforming the way a higher education institution measures and reports student progress so that every student can thrive can be bolstered by joining the Postsecondary Data Partnership (PDP) and submitting comprehensive student data to the National Student Clearinghouse. The initial data collection for the PDP is the most time consuming, includes up to five years of data, but the process outlined in this document can be replicated for future data collections. In this paper, several pieces of SAS® code were extracted and isolated for illustration purposes. The SAS® code can be utilized by beginners or intermediate programmers to prepare files for the PDP. While the PDP equips institutions with accessible reports and visualizations, SAS® software provides a flexible roadmap to successful data submissions.

INTRODUCTION

Transforming the way a higher education institution measures and reports student progress so that every student can thrive can be accomplished with joining the Postsecondary Data Partnership (PDP) and submitting comprehensive student data to the National Student Clearinghouse. The PDP empowers institutions with more comprehensive data, easier analysis, and better visual representations to help institutions understand, improve, and communicate student momentum, outcomes, and equity.

In 2021, Achieving the Dream launched a PDP Onboarding Course designed to support the initial data submission process. The PDP Onboarding Course is designed to support both two-year and four-year higher education institutions in successfully onboarding to the PDP. During the live sessions of the course, participants asked for clarification on how to assemble the files before uploading to the PDP Service. The discussion led to a conversation on conceptual and technical aspects of PDP file preparations, and the process for integrating student cohorts into their respective files. Preparing successful data submissions for the PDP is a challenge for institutions. This paper demonstrates a SAS® solution to the file preparation process for the PDP Data Submission cohort and course files.

To facilitate the preparation of data files, the National Student Clearinghouse (NSC) provides documentation for file formatting and field length requirements. To support the process, PDP Technical Briefs have been authored by Achieving the Dream SAS® programmers to provide technical support to those institutions onboarding to the PDP. The purpose of this paper was to highlight the power of SAS® software in facilitating the file creation process. Specifically, the use of macro variables in the SAS® environment adds efficiency to the file creation process, and macro variables utilized throughout file creation aid in accomplishing file and field formatting requirements. The flexibility of macros equipped the programmer with a file creation process that is replicated across multiple years or terms. Additionally, to harness the power of SAS® macros in file creation processing, macros were combined with LIBNAME statements to point to file locations, used to add additional required fields, and used in file naming conventions.

In this paper, several pieces of SAS® code were extracted and isolated for illustration purposes. The initial data collection is the most time consuming, including up to five years

of data, but the process outlined in this document can be replicated for future data collections. The SAS® code can be utilized by beginners or intermediate programmers to prepare files for the PDP. While the PDP equips institutions with accessible reports and visualizations, SAS® software provides a flexible roadmap to successful data submissions.

CONCEPTUAL FRAMEWORK: ASSEMBLING A PDP IMPLEMENTATION TEAM AND THE FILE PREPARATION PROCESS

The File Preparation Process

The first step to successfully creating your data files is to thoroughly read the PDP Data Submission Guide, paying particular attention to the different coding requirements as these may be different than how these are designated in your Student Information System (SIS). Getting to know the data fields will also help you to identify where the data “lives” within your SIS, and involving staff from IT, Institutional Research, and Enrollment and Registrar staff can help not only with finding what database tables will need to be accessed to collect the data, but also what reports already exist that include the data, as well as what queries may need to be written to collect the data not contained on existing reports.

You may also need to involve Academic and Advising staff, as you may need guidance on identifying Math and English gateway courses and determining Math and English college readiness. Having these conversations and knowing how to determine college readiness as well as defining gateway courses will allow you to appropriately identify these students during the data collection process.

Once you have identified where data elements are located, prepare the data, assemble the cohort, and course Files. Select the first “start” term such as Fall 2016. You may need to consult with your PDP partner to align with reporting needs for a group of institutions. Create a term cohort and course file as a test to upload to the clearinghouse, setting up the columns exactly as they are titled and laid out in the [PDP Data Submission Guide](#). Upload the test files, address needed corrections, and refine the process before creating the other terms. The process used to create the test files can be replicated for subsequent terms. Again, pay particular attention to the codes and formats for each data field!

Building the Cohort and Subsequent Course File

A good practice in the PDP data submission process is to prepare test cohort and course files that you will use for validation and quality checks. Conceptually, the cohort file is a single record file, one row per student, and a student has membership in only one cohort file. In other words, students remain in the cohort with which they were first enrolled in the institution. The second type of file is a multiple-record course file that consists of student’s course records for the term being reported. The associated information that identifies a student and ‘tags’ them as entering the institution with student id, first term and cohort year is unique to the student and this information should be retained for complete matching and record selections in both the cohort and course files moving forward.

Visualizing and Mapping the Process

Figures 1 and 2 demonstrate the beginning of the PDP reporting cycle for an institution with a fall 2016 start term. The figures below reflect a cohort file for the fall 2016 term in this example and the associated fall cohort 2016 students captured in the course records file. In this example, the students are enrolled in three courses each and this number of courses is used only for illustrative purposes.

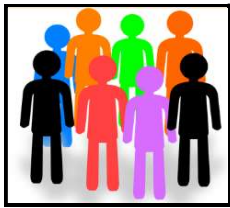


Figure 1. Cohort 1- Fall 2016

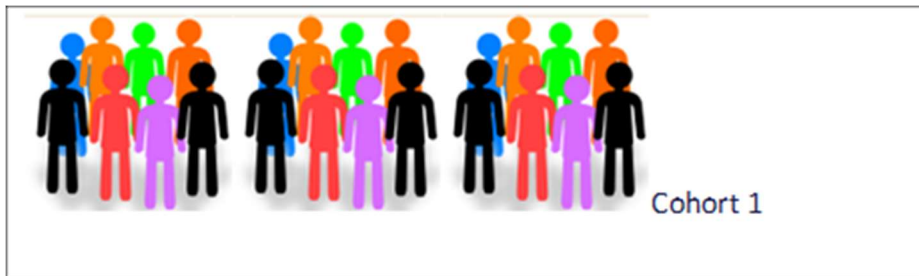


Figure 2. Fall Course Records Include Students Enrolled in Cohort 1

Figures 3 and 4 below demonstrate the reporting requirements for the next term. In this example, a spring 2017 cohort file is generated with the students first enrolling in the institution during the spring 2017 term. Since the course file includes students in cohorts reported to date, the course file should capture course records for both cohorts 1 and 2. Again, for illustrative purposes, the students are enrolled in three courses each for both cohorts 1 and 2.



Figure 3. Cohort 2- Spring 2017

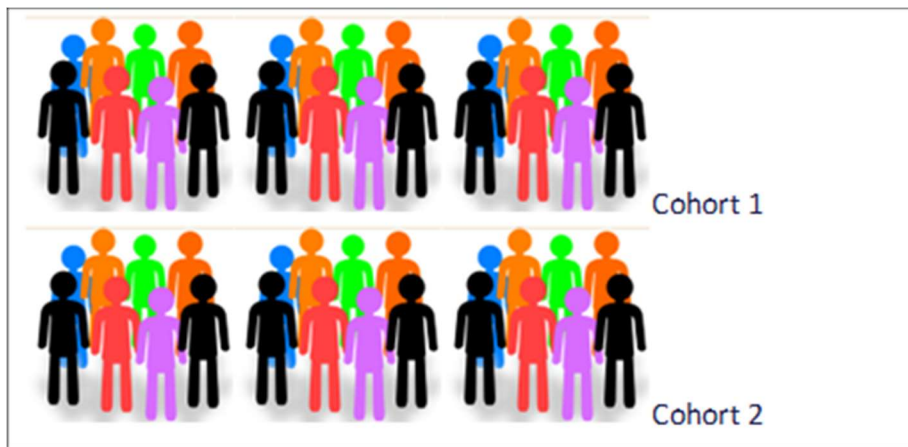


Figure 4. Spring 2017 Course Records Include Students Enrolled in Cohorts 1 and 2

Continuing with the reporting process for the historical data submission, the institution reports students attending the institution for the first time in the summer 2017 term. Now that the institution has three cohorts of students with which they are reporting, the course file includes course records and associated data fields for cohorts 1, 2 and 3. Again, this example in figure 3 shows all three cohorts are registered in three courses each, for illustrative purposes. It is more likely that student enrollments vary, and students may not appear in subsequent course files or stop-out and return to the institution later. The Clearinghouse will continue to follow the students for up to 8 years of enrollment and compute metrics for each student.

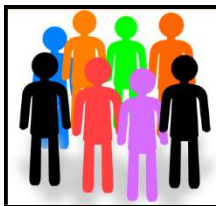


Figure 5. Cohort 3- Summer 2017



Figure 6. Summer 2017 Course Records Include Students Enrolled in Cohorts 1, 2, and 3

This data file preparation process continues to build multiple cohort and course files until 3-5 (5 recommended for good trend data) years of the historical data files are prepared and ready to upload. For example, the next cohort file created would be fall 2017, and the course files would include students enrolled during the fall 2017 term in cohorts 1,2,3 and 4. The historical data files should include the cohort and course information for all cohorts up to the most recent term available and according to the file reporting requirements.

SAS MACROS FACILITATE DATA SUBMISSIONS

The National Student Clearinghouse Postsecondary Data Partnership data collection requires two types of files. The first type of file is a single-record student cohort file that consists of all new students at the institution for the term being reported. The second type of file is a multiple-record course file that consists of student's course records for the term being reported. All cohort students to date, should be captured in the course files and complete listing of data fields and formats can be found in the [PDP Data Submission Guide](#). To facilitate the data collections, the programmer utilized macro variables in the SAS environment to format required fields and to provide identifying information in the file uploads. The macros are loaded during the file creation process to 1) provide proper formatting and complete field names and 2) to accomplish the creation of files for the PDP with headers and footers.

MACROS IN THE PATH LOCATION AND LIBNAME STATEMENTS INCLUDE THE FOLLOWING:

- LIBNAME statements provide file addresses and name the folder locations that will be utilized during file processing. Macro variables facilitate the preparation of the files for the PDP data collection. During file preparation, the programmer provides information in

the SAS environment regarding what files will be accessed during processing and where the files are located. In this example, LIBNAME statements identify the location of SAS files to be accessed and can also be combined with macro variables.

- The first macro listed below assigns a path location for your PDP files. The path is a main folder that resides on your local computer or network. The macro variable name in this example is assigned by the statement: `%LET VarName=EntirePathLocation;` and is subsequently employed in LIBNAME statements to add additional information about where the files are located. Notice the use of double quotes to embed the macro PathRoot in LIBNAME statements. Notice in this example the macro name is PathRoot and the path location is provided so that the program can locate the PDP files to be utilized. The LIBNAME statement can then be used to provide additional path information and point to the sub-folder locations where the specific files are located in the folder named PathRoot on physical drive F.

```
%LET PathRoot=F:\AchievingtheDream\ATD_XXXX\PDP; /*Change to your path*/

LIBNAME Cohort "&PathRoot\CohortFiles";
LIBNAME Course "&PathRoot\CourseFiles";
LIBNAME ALL "&PathRoot\SASDatasets\00_AllCohorts";
LIBNAME PDP "&PathRoot\Uploads";
```

Figure 7. Using Macros to provide Path and Folder Locations

MACROS IN THE FILE CREATION PROCESS INCLUDE THE FOLLOWING:

- The cohort files included new students in each term being reported. While the institution's existing SAS datasets included a data field that identified the students as new students in a particular term, not all of the information was included in the datasets. The programmer completed the process with required data fields by utilizing macro variables in SAS software. Notice the use of macros to name the academic term, using the institution's term naming convention for ease of reference, the academic year for the respective cohort, and the name of the term.
- Additionally, in September 2019, the Clearinghouse added a requirement that term begin and term end dates be added to the cohort files ([PDP Data Submission Guide](#)). The additional information was accomplished by consulting the academic calendar and using macro variables to accommodate the changes. Information that could not be located in the student information system and was added to the files is shown below, respectively. If the information is required by the Clearinghouse, missing information will create an error when files are uploaded to the clearinghouse. File preparers will need to adhere to field requirements and ensure that required fields are provided for each student in the file.

```

%LET yt01=2016FA; /*Academic term using term naming convention*/
%LET ch01=2016-17; /*Academic year*/
%LET chtm01=Fall; /*PDP term - Fall, Winter Spring or Summer*/
%LET TermBeginDate=20160821; /*Term Begin Date in this format, yyyymmdd*/
%LET TermEndDate=20161220; /*Term End Date in this format, yyyymmdd*/

```

Figure 8. Using Macros to Provide Term Information

- To add the required fields to the cohort and course files, the programmer added data fields to the files by associating the field name with the macro field in “quotes” as shown below. Associating a field name with a SAS macro requires the following naming convention: *FieldName*="*¯oname*;" Data fields that were not populated in the institution’s files were subsequently added using SAS coding. The process required collaboration with the institution’s registrar to ensure that the fields were accurately coded. The files were processed in the SAS environment using appropriate field formatting and prepared for uploading to the Clearinghouse.

```

CohortTerm="&chtm01";

TermBegin="&TermBeginDate";

TermEnd="&TermEndDate";

Cohort="&ch01";

```

Figure 9. Using Macros to Create Data Fields

MACROS IN THE FILE FORMATTING PROCESS INCLUDE THE FOLLOWING:

- The Clearinghouse process requires file naming conventions and files uploaded in a specific format. In each file, two header rows are required: the first row identifies the institution reporting files and the second row contains the field names associated with the files. The footer row requires T1 in the first position of the row followed by the total number of rows reported, including the footer row. Not adhering to the file structure, will cause a structural file error and Clearinghouse files are not accepted for processing [PDP Data Submission Guide](#). The specific file structure can be accomplished by using SAS macros to identify the reporting institution’s PDP account information and corresponding put statements for each segment of the Clearinghouse file.

```

/*Enter Student Clearinghouse Information*/

%LET SBD=20221112;      /* CLEARINGHOUSE File Completion Date */
%LET SC=XXXXXX;        /* OPEID# */
%LET SA=XXXXXXXX;      /*Service Account # for Header Row*/
%LET BC=00;            /*Branch Code # for Header Row*/
%LET ID=XXXX;          /*Institutional Short Name for Header Row*/
%LET SFTPMailbox=XXXXXXpdp; /*PDP Mailbox Address OPEID, BC, & pdp*/

```

Figure 10. Using Macros to Enter Student Clearinghouse Account Information

- The SAS code below provides the location to export the final file and names the file using the macro variables as defined. The set statement reads the SAS datafile-final&yt01- that was created during file processing. The file statement names the file with the appropriate txt extension and establishes the location for the course file, i.e., the uploads folder appropriately named PDP in the LIBNAME statement (Figure 1.)

```

/*Export to Text File*/

data test2;
set /*input sas file*/
  Final&yt01 end=IsLast ;
file "&pathroot\Uploads\'&SFTPMailbox'_course&yt01&SC .txt" DSD dlm=",";

```

Figure 11. Using Macros to Enter Student Clearinghouse Text File Information

- The first put statement creates the first of two header rows, on the first line of text (_N_=1). Notice the use of macro variables to accomplish the Clearinghouse requirements for the first header row in the course file as shown below: Commas are used to indicate blank spaces in the header row and may be used by Clearinghouse ([PDP Data Submission Guide, 2022](#)) in the future. The entire row of text is listed in double quotes.

```

If _N_=1 then do;
put
  @1 "DCE02, &SA, &SC, &BC, , , , , &SBD, , &ID, , , , "

```


Table 1. Using Macros to Create First Header Row

- The second put statement creates the second column header row as outlined in the Clearinghouse documentation. A third put statement is used to reference the actual institutional field names in the data file. The advantage of using the two put statements here is that there is no need to rename your institution's field names in your files which vary across data systems. The fields are associated with the specific Clearinghouse documentation requirements using the put statement as shown below. The first entire set of text is enclosed in double quotes. The next put statement references the actual field names in the data file and this text is not enclosed in double quotes.

```
put
```

```
"CH1,Cohort,Cohort Term,Academic Year,Term,Institution ID Type,Institution ID,SSN,ITIN,Student ID,First Name,Middle Name,Last Name,Suffix, Current Street 1,Current Street 2,Current City,Current State,Current Zip/Postal Code,Current Country,Date of Birth,Student Phone Number,Pell Recipient,Student Email,CompleteDevMath,CompleteDevEnglish,TransferIntent,Degree Type Sought,Semester/Session GPA,Overall GPA,Course Prefix,Course Number,Section ID,Course Name,Course Description,Course CIP,Course Type,MathOrEnglishGateway,Co-requisite Course,Course Begin Date,Course End Date,Grade,Number of Credits Attempted,Number of Credits Earned,Delivery Method,Core Course,Core Course Type,Core Competency Completed>Total Combined Earned and Transferred Credits,Purpose of Course Exchange,Certification Endorsed Curriculum/Program,Certificate Endorsing Industry,Grade Effective Date,DGI Institution ID Type,DGI Institution ID,DGI Student ID"
```

```
;
```

```
end;
```

```
put CH1 Cohort CohortTerm AcademicYear Term InstitutionIDType InstitutionIDSSN ITIN ID FirstName MiddleName LastName ADCStuNameSuffix Address1 Address2 CityState Zip Country Birthdate ADCStuPersonalPhone Pell ADCStuEmailAddress CompleteDevMathCompleteDevEnglish Intent DegreeType SemesterGPA2 CumulativeGPA2 CoursePrefix CourseNumber SectionID CourseName ADCClsShortTitle CourseCIP CourseType MathorEnglishGateway CorequisiteCourse CourseBeginDate CourseEndDate Grade NumCreditsAttempt NumCreditsComplete Method CoreCourse CoreCourseType CoreCompetencyCompleted TotalCombCredits PurposeCourseExchange CertCurriculumProgram CertEndorsingIndustry GradeEffectiveDate DGIInstitutionIDType DGIInstitutionID DGISTudentID
```

```
;
```

Figure 12. Using Macros to Create Column Header Row and Detail Rows

- Finally, the footer row is accomplished by the final put statement and proper formatting is applied to the last row of the file. Using the sample below the footer row is properly formatted, listed on the last row of the text file, and ready for Clearinghouse submission without manual computations.

```
/*Sample Output Line*/  
  
IF IsLast=1 then do;  
  CT=_n_+3;  
  put @1 "T1"  
  @3 ', '  
  
  @4 CT " "  
  
  ;
```

Figure 13. Using Macros to Create Footer Row

CONCLUSION

Transforming the way a higher education institution measures and reports student progress so that every student can thrive can be accomplished with joining the Postsecondary Data Partnership (PDP) and submitting comprehensive student data to the National Student Clearinghouse. The PDP empowers institutions with more comprehensive data, easier analysis, and better visual representations to help institutions understand, improve, and communicate student momentum, outcomes, and equity.

In 2021, Achieving the Dream launched a PDP Onboarding Course designed to support the initial data submission process. The PDP Onboarding Course is designed to support both two-year and four-year institutions in successfully onboarding to the PDP. During the live sessions of the course, attendees asked for clarification on how to assemble the files before uploading to the PDP Service. The discussion led to a conversation on technical aspects of PDP file preparations, and the process for integrating student cohorts into their respective files. This paper demonstrated a SAS solution to the file preparation process for cohort and course files of the PDP Data Submission.

To facilitate the preparation of data files, the National Student Clearinghouse (NSC) provides documentation for file formatting and field length requirements. PDP Technical Briefs were developed by Achieving the Dream SAS® programmers to provide support to those onboarding to the PDP. The purpose of this paper was to highlight the power of SAS® software in facilitating the file creation process. Specifically, the use of macro variables in the SAS® environment added efficiency to the file creation process, and macro variables utilized throughout file creation aided in correct file and field formatting requirements. The flexibility of macros equipped the programmer with a file creation process that was replicated across multiple years or terms. Additionally, to harness the power of SAS® macros in file creation processing, macros were combined with LIBNAME statements to point to file locations, used to add additional required fields, and used in file naming conventions.

In this paper, several pieces of SAS® code were extracted and isolated for illustration purposes. The initial data collection is the most time consuming, including up to five years of data, but the process outlined in this document can be replicated for future data collections. The SAS® code can be utilized by beginners or intermediate programmers to prepare files for the PDP. The PDP equips institutions with accessible reports and visualizations and SAS® software provides a flexible roadmap to successful data submissions.

RECOMMENDATIONS FOR SUCCESSFUL ONBOARDING

- **Consult the [PDP Knowledge Base](#).** The first step to successfully creating your data files is to thoroughly read the [PDP-Data-Submission-Guide](#), paying particular attention to the different coding requirements.
- **Save Official Cohort Files-**A useful tip is to create the cohort file once and save it as the official cohort file for that specific cohort term with student id, cohort term, cohort year which includes all data elements reported for the student cohort for easy retrieval.
- **Document and keep records of where you are in the file submissions process-** Keep an excel spreadsheet that documents your progress in the file submission process as you upload files and correct or modify submissions. The historical data collection process, including approvals, error corrections, etc. could take several weeks and day-to-day tasks can interrupt the workflow. However, set aside time to work on the data submissions each day, if possible, since the nature of this work requires a focused and persistent effort to complete. Establish an internal timeline and stay on task to ensure an efficient, robust, and complete process.

- **Upload Test Files Before Creating All of the Files:** Create a term cohort and course file as a test to upload to the clearinghouse, setting up the columns exactly as they are titled and laid out in the PDP Submission Guide. Upload the test files, address needed corrections, and refine the process before creating the other terms. The process used to create the test files can be replicated for subsequent terms. Again, pay particular attention to the codes and formats for each data field!
- **Submit Manageable File Sizes:** Since larger files can lead to longer processing and editing time, it is recommended to submit multiple smaller files. Start with the first cohort term and/or first year and move forward with the subsequent course file submissions in sequential order.
- **Ensure Data Quality:** Data quality checks will enhance the quality of the file submissions, reduce errors, and shorten the time to upload.

REFERENCES

Achieving the Dream, "Postsecondary Data Partnership." Accessed September 25, 2022. <https://achievingthedream.org/initiatives/postsecondary-data-partnership/>

National Student Clearinghouse, "PDP Knowledge Base." Accessed September 25, 2022. <https://help.studentclearinghouse.org/pdp/>

National Student Clearinghouse, "PDP Data Submission Guide." Accessed September 25, 2022. <https://help.studentclearinghouse.org/pdp/wp-content/uploads/PDP-Data-Submission-Guide-v2.2-JHM-8-4-22.pdf>

SHEEO, "Using SAS: Complete ATD Examples in PDP Toolkit." Accessed September 25, 2022. <https://sheeo.org/postsecondary-data-partnership-toolkit/>

CONTACT INFORMATION

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

Dr. Bobbie Frye
 Achieving the Dream
 240-450-3862 Ext. 362
bfrye@achievingthedream.org