AACRAO

Report on Barriers to Postsecondary Data Partnership (PDP) Data Submission

By Ken Sharp, PhD
Senior Consultant, AACRAO

Introduction/Goal

The National Student Clearinghouse (NSC) Postsecondary Partnership (PDP) is a national tool that focuses on helping institutions monitor, report and improve student outcomes. As of 2021, roughly 50% of the institutions that signed up for the PDP have not submitted their initial data. As a result of this low success rate AACRAO undertook a study to help understand the barriers that prevent institutions from submitting their data to the PDP.

At the outset, it was assumed the primary issues were competing priorities and institutional lack of personnel with the technical skills necessary to perform extracts. Additionally, the PDP data preparation and submission processes are relatively complex and provide a number of instances in which institutions may fail to move forward.

A broad outline of these steps include:
- institution reviews data elements to identify potential missing fields, establish definitions and/or identify and resolve internal data-standards issues
- institution maps data from their sources and creates the required (and optional) data files
- institution submits data files
- data goes through series of structural and field validations
- institution fixes any identified issues and resubmits, as necessary
- once file is accepted, institution certifies the data
- data goes through series of quality checks
- institution fixes identified data-quality issues and resubmits, as necessary
- when data passes quality checks, the Analysis Ready file and dashboards are generated

Each step is complex and may result in technical issues.   Issues resulting from differences in the interpretation of the data are also possible.

Approach

The most direct approach to determine barriers institutions may encounter when trying to submit data to the PDP is to interview institutions that successfully submitted data and those that were unsuccessful. This approach was more difficult than anticipated.

AACRAO initially requested a listing of institutional contacts from NSC. After contacting their legal department, NSC concluded they were unable to release a listing of institutional contacts. However, they could release a listing of participating institutions.

As a first step, AACRAO cross-referenced the PDP institutional list with the list of attendees at the annual conference. Out of 103 invitations sent, only three participants attended the two scheduled roundtables. None of these attendees was in a position to provide salient input.

AACRAO and the Association for Institutional Research (AIR) both received grants asking for similar information. The two groups agreed to collaborate on data-collection efforts.

In an effort to identify additional institutions, AACRAO requested NSC send out a consent email to all NSC PDP participating institutions asking them to opt in to being contacted by AACRAO and/or AIR for data-collection purposes. Ninety institutions responded. This process identified candidate institutions to interview, but it may not be representative of the 50% of institutions that did not provide submissions.

Interviews were conducted in three different groups.
- AIR, as part of their related initiatives, invited institutions to participate in focus groups at their 2022 AIR Forum (June, 2022) to discuss the PDP and PDP-related issues. AACRAO was also invited to participate in these sessions. Four focus groups were held with 2-9 participants per group. Feedback from 15 individuals representing 12 institutions was collected.

- Third- party Interviews were conducted. Four individuals from three third parties were interviewed. Third parties in this context are organizations engaging with institutions under contract (or grant) to use the PDP as a reporting tool. Interviews and discussions were held with Achieving the Dream (ATD), Complete College America (CCA) and the National Association of Student Personnel Administrators (NASPA). DIscussions were held based on the assumption that third parties could represent an aggregate of issues faced by institutions.

- Individual Interviews were conducted throughout 2022. Fourteen individuals representing 11 institutions were interviewed.

In total, interviews were conducted with 29 individuals representing 23 institutions.

In order to solicit the broadest range of responses relevant to each institution's experiences, those interviewed were asked to reflect on their experiences rather than answering specific questions. Probing questions were asked for clarification on topics when necessary. No recordings of responses were made; responses were documented as written notes.

Notes from all interviewees were reviewed, and responses were grouped where similarities existed. These groupings represent the specific issues institutions encountered with the PDP. Because institutions did not answer a similar set of questions, it was not appropriate to report the number of institutions providing specific responses. Instead, this report will make use of broad categories, such as *some institutions* or *many institutions when* appropriate.

Not all issues identified are specifically a hindrance to the data load but may represent a hindrance to an institution's adoption of the PDP. For the sake of reporting, issues have been grouped together into related-issue topics or groupings. Some issues could be categorized under multiple topics; however, they have only been reported under one topic area. Based on responses and observations, a selection of Key Findings, page 8, were identified and reported.

General Findings

General issues identified by interviewees are listed below, grouped by topic.

Data Definitions

Some institutions expressed concerns about the *definitions* of select data elements. Some participants were uncertain of some definitions used in preparing the submission file. Additionally, concern was expressed about the usability of the PDP submission data because key definitions used in PDP and dashboard reports are inconsistent with those used by the institution.

PDP Data Concerns

Two types of concerns were brought up regarding the PDP data. These concerns can be grouped into concerns about the quality of data and concerns regarding data security/ethical use of data.

Regarding the *quality of data,* several institutions expressed concerns about "how good" their data files were, given the complexity of the extracts. Even if the file adhered to the format and passed the data-quality checks, it is not an indicator the submission file represents what

actually occurred at the institution. Further concerns were expressed that submissions by less technically sophisticated institutions are more likely to contain errors.

The centralized collection of student-level data was also identified as a concern, from the perspective of *data security*. If an institution can be subject to a data breach, so might NSC and the PDP. Others expressed concern about personal privacy and the ethical questions about the collection and use of personal information in a centralized database without the ability for individuals to "opt out."

Formatting

Formatting required for data files has been identified by institutions as an area of concern. There is a feeling among institutional representatives that subtle changes in data formats can result in significant issues during the load process.

The rules on special characters are dependent on the field in the data load file. This can lead to confusion including the belief the PDP does not accept any special characters.

Field lengths were also identified as an issue, particularly with regard to addresses. Length restrictions create challenges with international student records, students with long street addresses or long city names. These restrictions make it difficult for institutions to validate and fix addresses.

One positive comment that came out of institutional discussions on data formatting is that many institutions used the development of the data files as an opportunity to readdress their data standards and clean up their records.

Data Elements

Three concerns were identified by institutions regarding required data files. These concerns were associated with the use of course Classification of Instructional Program (CIP) codes, coding of dual majors and identification and coding of gateway courses.

Many institutions do not currently have *CIP codes* associated with courses. These institutions find it challenging to take time and staff to identify and code CIP codes for individual courses.

Some institutions indicated the PDP does not accommodate *dual majors* and are asking NSC to include this data element in the PDP.

The *identification and coding of gateway courses* continue to be identified as an issue. These courses are frequently perceived to be courses at the institutional level for all students, or the rubric for determining gateway courses is perceived to be confusing.

Load Process

The actual PDP data-file-load process has been identified as being problematic by many institutions. The *load process* has been characterized as being more complex than other clearinghouse submissions.

Error codes generated during a failed submission are difficult to interpret/understand. These codes could be clearer or more helpful. Some types of errors block further analysis of a file. Frequently in these cases, once one problem is solved, another problem occurs.

Overall, most institutions indicate there is a steep learning curve associated with setting up the initial data file and troubleshooting errors in the submission file. Despite regular changes in the data elements requested by NSC, institutions report that after the initial load, subsequent loads are significantly easier and require less time.

Tools

Based on an institution's information systems and the technical capabilities of staff, a wide range of *tools* are used to pull information for a data load.

Data is not always pulled from the student-information system. Several institutions use their existing reporting infrastructure and/or data warehouses for part or all of the data used to create the report files that limit the reusability of scripts.

Many institutions make use of MS Excel to manipulate or review data prior to submission. Unfortunately, it is not uncommon for MSExcel (depending on user settings and expertise) to change data-field types and/or eliminate key data, such as leading zeros, which can result in a data file that will be rejected upon load. Sophisticated users, however, report MS Excel is a valuable tool for reviewing and finalizing institutional load files.

There are information-system-specific issues and benefits mentioned by institutions. Key items identified included:

- institutions using Jenzabar generally have a difficult time pulling information to create their load files.
- Empower has developed a process/report to automate generation of a submission file.
- Ellucian has the ability to generate the PDP submission file; however, implementation of Ethos is required. At this point, few institutions have implemented Ethos.

Client Internal Issues

Not all issues institutions report are technical in nature. Submitting data in the PDP can require a high degree of cooperation across many functional areas of an institution.

The registrar's office is frequently the office within the institution that initially controls access to NSC products. Registrars and their staff can have a negative influence on the load process, if they do not understand the security roles associated with each NSC product.

Given the cross-functional nature of the PDP data file, some institutions have reported data-ownership issues when a department is reluctant to release information to be included in the data file. A subset of these institutions has speculated problems may have arisen out of a lack of communication between departments, the reluctant department not understanding the security inherent in the process.

It was not uncommon for institutions to report starting the PDP process resulted in reviews of internal processes. Pulling the PDP data file can highlight data-quality issues, which in turn can result in review of internal processes for fixing data. Some institutions reported the data-load process was delayed until internal processes were updated and data was corrected in their information systems.

The story told by data can be nuanced, based on the assumptions inherent in a data set and the use of similar data labels that may have different meanings. Some institutions expressed concerns these differences may result in confusion among institutional faculty and staff or even the general public. Concerns over misinterpretation of data and dashboard reports were reported as extremely concerning to some institutions.

Many medium- and large-sized institutions report having their own data warehouses and/or dashboards. These institutions have dedicated significant efforts to development of these resources to meet specific institutional needs. In a number of cases, institutional representatives expressed doubt in the ability of the PDP's dashboards to compete with institutional-specific resources. Additionally, because significant institutional-client input was

used to develop these resources, institutional representatives expressed concerns that institutional clients would not make use of the PDP's dashboards.

As anticipated, staff burnout, staff turnover and other priorities were mentioned as issues for institutions adopting the PDP. These concerns were more prominent among smaller institutions, as opposed to larger ones.

Support

Several institutions mention that the National Student Clearinghouse could use a more comprehensive *support* structure.

Although a help desk exists, several institutions were not aware of it. Of those institutions reporting contact with the help desk, about half mentioned positive interactions. The other half did not find the help desk helpful. A few institutional contacts indicated having a personal contact at NSC would be beneficial, particularly when these same institutions reported it can take weeks to receive support from NSC.

Many institutions characterized the PDP load as being more difficult than other clearinghouse-load processes. Several mentioned having an NSC contact who could review submission files for new institutions would be quite helpful.

Contracting Process

Only one institution discussed challenges with the NSC-PDP *contracting process.* In this instance, both the institution's processes and NSC's processes were perceived as a cumbersome delay to the enrollment process.

Additional Observations

At most institutions, Institutional Research/Institutional Effectiveness is generally the office responsible for the PDP. The Registrar's office is generally a peripheral participant.

Most institutions participating in the PDP have joined as part of a grant, an initiative, a statewide activity or to meet accreditation requirements. Very few institutions adopt the PDP as an enrollment-management tool.

Institutions generally fall into three categories-=-small, medium and large. Small institutions struggle due to a lack of specialized staff and other reporting requirements. Medium-sized institutions are well positioned both in terms of having specialized staffing (not necessarily a large number of them) and no competing data-reporting systems. Large institutions have specialized staff with the technical capabilities, but institutional staff are generally already invested in an existing reporting infrastructure.

There are currently a wide range of effective training and technical resources available to institutions. Most institutions access some of these resources but are not aware of the full, and growing, collection of resources available.

One strategy is the development of libraries of programs or scripts for collecting data. Based on discussions, collecting a library of scripts is less viable than envisioned, due to differences in institutional definitions of data fields and the use of intermediate reporting tools to develop report files.

Key Findings

- Registrars and enrollment managers have little role in the PDP-submission process. The function is dominated by Institutional Research/ Institutional Effectiveness.

- Many institutions perceive the PDP to be a compliance tool, not an enrollment-management tool.

- Due to differences in institutional-data standards, institutional practices, student-information systems and reporting tools, there appear to be few opportunities for script repositories to be useful. There are two key exceptions;state systems, in which institutions share similar infrastructure, and when a reporting tool is developed by the student- information-system vendor.

- Many institutions struggle with the PDP-data-load-feedback mechanism. Most indicate it has limited use for troubleshooting load issues. They perceive there is little to no support available from NSC-PDP to help troubleshoot/resolve load issues.

- Institutions fall into three categories–small, medium and large. Small institutions seem to struggle due to a lack of specialized staff and competing reporting requirements. Medium-sized institutions are well positioned, both in terms of having specialized staffing (not necessarily a large number of them) and no competing data-reporting

systems. Large institutions have specialized staff with the technical capabilities, but institutional staff are generally already invested in an existing reporting infrastructure.

● There are concerns among institutions that PDP dashboards provide information that may be perceived as conflicting with existing institutional reports/data warehouses, due to the differences between PDP data definitions and institutional data definitions.