**NSC Data Pulls - Major Quarterly Queries, All Non-Graduated Students**

The National Student Clearinghouse provides data on an individual's college engagement. It can answer questions about a student's college journey once that student is no longer working with Bottom Line. This data is provided by the NSC and stored in Salesforce as multiple records per student. We can request data in bulk for large numbers of students through an upload process.

***\*\*Note:*** we have a limited number of rows we can request in a year. It is **very important** to verify your steps and make sure everything is in order before submitting to NSC.  If time allows, we recommend that you get a second set of eyes on the final file before submitting.

**To request bulk data from the NSC**

1. Use the Salesforce report NSC Pull Template[Quarterly](https://bottom-line.lightning.force.com/lightning/r/Report/00O1Y000007HgBrUAK/view?queryScope=userFolders) - do not adjust the filters
   1. ​​​​​​​in March of 2022, we made the following modifications to the filters to accommodate for limitations in the number of Data Pull credits we receive from NSC:​​​​​​​
      1. A filter was added for Student Type NOT EQUAL TO "ERROR NO PROGRAM DATA"- those student records were added to Salesforce for a special project, and were neither Bottom Line students nor were they ever part of the program.
      2. A filter was added for current Access students- they are currently in High School and therefore have no enrollment records in NSC. **THIS FILTER WILL HAVE TO BE CHANGED EVERY YEAR BEFORE THE JULY DATA PULL**
      3. A filter was added for any students without a Date of Birth -those inquiries come back empty every time we make a request and therefore are wasteful.
2. Export it as a CSV
3. Edit it following instructions found starting p19 [here](https://www.google.com/url?q=https://studentclearinghouse.info/onestop/wp-content/uploads/STCU_User_Manual.pdf&sa=D&sntz=1&usg=AFQjCNHuYYF3GlovWn7NZRJoXzbjFf-ZtQ) or use the [example spreadsheet available here](https://drive.google.com/file/d/11q_nQXnhOPHawbS_ZxvX6nZ6ZNSbmf9u/view?usp=sharing) as a guide.  
   **\*\***If you use the sample spreadsheet, make sure to open the full document- the browser preview may not show all the columns.
   1. Trim off the "Internal use only" rows at the end of the report
   2. Add 2 columns to the left of Column A
   3. Col A - fill all the way down with D1
   4. Col B - leave blank
   5. Col C - First Name
   6. Col D - Middle Initial
   7. Col E - Last Name​​​​​​​​​​​​​​
   8. Col F - Suffix
      1. ​​​​​​​NSC files **DO NOT LIKE** special characters.
         * ​​​​​​​Search for (- ? . 1 \_ \* () and áéíóú, ñ, ü, àùè, ' `  )
           + "-"; Hyphens are the most common character found in this dataset
           + "?"; use "~" before the question mark, otherwise excel will use as wildcard.  
             The CSV interprets "ñ" as either a question mark or a missing character
           + "."; common in abbreviations and names that start with "St."
           + "1"; a few records have this as a marker for students with the same name (old)

This search will point towards "Anonymous student 1". Remove that student from the data request.

* + - * + "\_"; another interpretation of "ñ"
        + "\*"; use "~" before the question mark, otherwise excel will use as wildcard.
        + ","; this is especially problematic because it sends the return file with fields displaced.
      * Also make sure to check the last name field for hidden suffixes
        + Check for JR/Jr and III/Iii and move them to the Suffix field
  1. Col G - Date Of Birth
     1. This must be in YYYYMMDD format
     2. It comes from a formula in Salesforce and is already in the template report, so there's no need to transform it
  2. Col H - the date that starts your search request
     1. This must be in YYYYMMDD format
     2. It tells the NSC what date to begin its search.
     3. Set this to the date of the **last data pull** (for March it would be a date in November of the prior year, for August it would be March, and for November it would be August)
  3. Col I - leave empty
  4. Col J - leave empty
  5. Col K - make this a text field and fill down 00 (zero zero)
  6. Col L - this must contain a unique identifier from Salesforce.
     1. If you use the template report, it will be in the format of SPR-000001750
     2. DO NOT USE a Salesforce 15 or 18 character ID (e.g. a3F1Y000001Lwe9) - the returned data from the NSC will make it all caps and these IDs are case sensitive.

1. Add a header row above row 1 that contains:
   1. Col A: H1
   2. Col B: 600173
   3. Col C: 00
   4. Col D: Bottom Line
   5. Col E: Today's date in YYYYMMDD format
   6. Col F: SE
   7. Col G: S
2. Add a footer row below the last row of data that contains:​​​​​​​
   1. Col A: T1
   2. Col B: the row number of this row.
3. ​​​​​​​Save the file as a **tab-delimited text** with 600173 at the beginning of the name.
   1. ​​​​​​​MAKE SURE ALL OF THE 00 COLUMNS REMAIN CORRECTLY FORMATTED before you export - Excel loves to change double zeros to single zeros.
4. Log into the [NSC FTP](https://www.google.com/url?q=https://ftps.nslc.org&sa=D&sntz=1&usg=AFQjCNEhtWVreMZZhWv-GOy56MqCzSXlyg) server. Use the upload wizard on the landing page to upload the bulk request.​​​​​​​

**What to expect after submitting a request**

The NSC will automatically process the uploaded file. If there are errors in the file format, someone from the NSC will contact Sarah Place. Sarah also gets emails for all requests and is emailed when a file has been processed and the results are available.

The NSC takes up to 72 hours (sometimes a bit more) to process a request. Large requests take longer than small ones.

You will get 3 files back from the NSC - one with details about the students, one that shows aggregated data about the colleges, and one HTML file that we don't use. See instructions on this page for processing the results.

**Getting Data From NSC into Salesforce**

NSC data is meant to be stored only for students who have participated in BL programming in the past. We do not store NSC data for anyone that is not in the data warehouse ([Student Progressions](https://sites.google.com/deepwhydesign.com/bl-reporting-documentation/technical-documentation/student-progressions) object). The NSC Records object is a child of the Student Progressions object. We use this relationship to link NSC data back to the correct student without having to worry about over-writing Bottom Line data.

**Retrieving Data from the NSC**

When your bulk request file has been processed, Sarah Place will get an email notification.

1. Log into the NSC FTP server ([https://ftps.nslc.org](https://www.google.com/url?q=https://ftps.nslc.org/&sa=D&sntz=1&usg=AFQjCNG8NixFr-Yloiwm0UNQE-w_gOzDxw))
2. Click on "Go to your Home Folder" from the homepage
3. Click on "receive" from the list of folders that appears
4. Download the files you need to work with. Description of each file is below the image.

Each request you upload will result in THREE files. All 3 will be prefixed with the BL NSC ID (600173) and may look very similar to other files in the folder. There will be 6 numbers after the "T" in the file names (represented below by ######) and the date of the request will be shown in the file names (represented below by YYYYMMMDD)

1. 600173\_T######cntlrpt.YYYYMMDD\_SE.htm - it is safe to ignore this file, it is a control file that is delivered as a web page
2. 600173\_T######aggrrpt.YYYYMMDD\_SE.csv - this shows data aggregated by college (e.g. how many degrees awarded by each college to the students in the request set, how many students / what percent of the request set have records at a college, etc). It is not useful for upload to Salesforce but may occasionally be of interest for comparison or research reporting.
3. 600173\_T######.YYYYMMDD\_SE.csv - this file is the one to work with throughout the remainder of this how-to. It contains one record per student per college reporting period and is used to provide "gold standard" reporting on graduation, persistence, and attendance.

**Processing the 600173\_T######.YYYYMMDD\_SE.csv file for upload to Salesforce**

The NSC Records object contains **one row per student per reporting period**. This means that students will have multiple records that represent their postsecondary academic history.

For relatively easy data management in Salesforce, we have constructed a **unique key pattern** that concatenates the Student Progression Name (in the format SPR-XXXXXXXX) with the Enrollment Begin field from the NSC results. Both of these data elements are provided with the NSC results, and can be concatenated into a globally unique record identifier that is able to be easily linked back to a unique Salesforce record.

The NSC file does require some preparation. The steps are:

1. Open the CSV that was returned from the NSC and is named as 600173\_T######.YYYYMMDD\_SE.csv (this will be the largest file returned by the NSC).
2. The fields returned are as follows (do not import any data for fields in italics)
   * ***Your Unique Identifier****- this will be empty*
   * ***First Name****- student's first name. NOT imported into Salesforce*
   * ***Middle Initial****- student's middle initial. NOT imported into Salesforce*
   * ***Last Name****- student's last name. NOT imported into Salesforce*
   * ***Name Suffix****- student's name suffix (e.g. Jr, III, etc). NOT imported into Salesforce*
   * **Requester Return Field** - the name of the Student Progression record that this NSC record is a child of, with an extra underscore appended (e.g. SPR-000000001\_). This is the critical key for both unique record ID construction and the link to the student's BL record.
   * ***Record Found Y/N****- was this student located in the NSC? Y if yes, N if no. Note that there are many reasons that a student might not be located, but typically a mismatched name or birthdate are to blame.*
   * ***Search Date****- the date you requested that the NSC begin the search for records. This is set when you*[*submit the bulk*](https://sites.google.com/deepwhydesign.com/bl-reporting-documentation/technical-documentation/requesting-data-from-nsc)*request.*
   * **College Code/Branch** - The NSC code for the college, in the format of ######-##. This does not currently link to anything else in Salesforce. ([\*\*\* this could be improved \*\*\*](https://sites.google.com/deepwhydesign.com/bl-reporting-documentation/improvements-suggestions))
   * **College Name**- The name of the college that this record relates to. Note that a single student may have multiple colleges in their history.
   * **College State** - The US state that the college is in (e.g. MA, IL, NY, etc.)
   * **2-year / 4-year**- Whether this school is coded as a 2 or 4 year institution. Note that it is possible, but rare, for NSC data to differ from ours.
   * **Public / Private** - Whether this school is listed as a public or private college. Note that it is possible, but very rare, for NSC data to differ from ours.
   * **Enrollment Begin**- the date (in YYYYMMDD format) that begins this enrollment reporting period. This field is the second part of the unique ID key for most records and ensures we are able to upsert NSC data cleanly over time.
   * **Enrollment End** - the date (in YYYYMMDD format) that ends this enrollment reporting period. Informational only.
   * **Enrollment Status** - A single-letter code to indicate the student's enrollment status for this reporting period. Note this field is sometimes blank if the college did not provide it to the NSC.
     + F = Full time
     + Q = Three-quarters time
     + H = Half time
     + L = Less than half time
     + A = Leave of absence
     + W = Withdrawn
     + D = Deceased
   * **Class Level**- A single-letter code to indicate whether, for this reporting period, a student is classified as the following. Note this field is often blank if the college did not provide it to the NSC.
     + F - freshman
     + S - sophomore
     + J - junior
     + S - senior
   * **Enrollment Major 1** - The student's primary major for this reporting as reported by the college to the NSC. Note this field is often blank.
   * **Enrollment CIP 1** - The student's primary CIP for this reporting as reported by the college to the NSC. Note this field is often blank.
   * **Enrollment Major 2**- Additional major for this reporting as reported by the college to the NSC. Note this field is often blank.
   * **Enrollment CIP 2**- Additional CIP for this reporting as reported by the college to the NSC. Note this field is often blank.
   * **Graduated?**- If a student has graduated, there will be one record per degree and this field will contain "Y" for the graduation record. For standard enrollment records, this will be "N"
   * **Graduation Date** - If the student has graduated, this will have a date in YYYYMMDD format. Records will either have an Enrollment Begin or a Graduation Date but not both. For Graduation records, this is the date that is used to create the unique key (instead of Enrollment Begin)
   * **Degree Title -**If the student has graduated, the degree title will be in this field if it was reported by the college. This is occasionally blank.
   * **Degree Major 1**- if the student has graduated and the college reported a graduating major, it will be in this field.
   * **Degree CIP 1** - if the student has graduated and the college reported a graduating CIP, it will be in this field.
   * **Degree Major 2**- if the student has graduated and the college reported a graduating major, it will be in this field.
   * **Degree CIP 2**- if the student has graduated and the college reported a graduating CIP, it will be in this field.
   * **Degree Major 3** - if the student has graduated and the college reported a graduating major, it will be in this field.
   * **Degree CIP 3** - if the student has graduated and the college reported a graduating CIP, it will be in this field.
   * **Degree Major 4** - if the student has graduated and the college reported a graduating major, it will be in this field.
   * **Degree CIP 4** - if the student has graduated and the college reported a graduating CIP, it will be in this field.
   * **College Sequence** - if the student has transferred, this field will contain information about the sequence of colleges
3. If you have to use the dataloader app (rather than dataloader.io), run a report from Salesforce to get the Student Progression record ID (the 15 or 18 character Salesforce ID) and the record name (SPR-000000001). This forms a crosswalk between the data returned from the NSC and the record ID so that dataloader can use the record ID. If you are using dataloader.io you can simply lookup as using the name.

**Data Prep - Part 1: Clean up and Format**

The file requires a small amount of processing before being uploaded to Salesforce.

1. In column F (Requester Return Field) - remove the \_ character from all rows in the column
2. Sort by column G (Record Found Y/N) and remove all rows with "N" in this column.
3. Remove columns A-E (Your Unique Identifier through Name Suffix) and G and H (Record Found Y/N and Search Date)
4. Add two empty columns to the beginning of the file (new cols A and B). The headers should be "External ID from NSC" and "Student Progression" (only relevant if NOT using dataloader.io)
5. Sort by Enrollment Begin
   1. Find the empty enrollments at the bottom first​​​​​
   2. If Enrollment Begin is empty
      1. "External ID from NSC" = Requester Return Field & Graduation Date & "G"
      2. Leave no spaces or punctuation
      3. The result should look like SPR-00000000120180801G and each row will be unique \*
   3. If enrollment Begin is NOT empty
      1. ​​​​​​​"External ID from NSC" = Requester Return Field & Enrollment Begin
      2. Leave no spaces or punctuation
      3. The result should look like SPR-00000000120180801 and each row will be unique \*
6. If you are using dataloader.io you can skip this step. If not, use VLOOKUP to populate the "Student Progression" field with the student progression record ID, linking by name from "Requester Return Field"
7. Add a new column "Enrollment Begin Date" and copy the "Enrollment Begin" into it. Format it as a regular date (MM/DD/YYYY) - you will upload *both* dates.
   1. Formula is:    =RIGHT(LEFT(I2,6),2)&"/"&RIGHT(I2,2)&"/"&LEFT(I2,4)   
      **NOTE**: use this formula as guide; ALWAYS double-check that the reference cell is using  the value for "Enrollment Begin"
8. Add a new column "Enrollment End Date" and copy the "Enrollment End" into it. Format it as a regular date (MM/DD/YYYY) - you will upload *both* dates.
   1. Formula is:    =RIGHT(LEFT(K2,6),2)&"/"&RIGHT(K2,2)&"/"&LEFT(K2,4)  
      **NOTE**: use this formula as guide; ALWAYS double-check that the reference cell is using  the value for "Enrollment End"
   2. IF there is ever a need to reverse the date to the original value, you can use this formula =RIGHT(Cell#,4)&LEFT(Cell#,2)&RIGHT(LEFT(Cell#,5),2) OR you can use the special formatting for cells, under "Custom" options.
9. Format Graduation Date as a regular date (MM/DD/YYYY). You will upload only the date-formatted field. (use formula above- reference column will be column T; keep in mind same considerations)
10. Add a new column "Type of Diploma" right after the "Degree Title" column. Translate the field "Degree Title" into one of the following four options:
    1. Certificate or Other
    2. Associates
       1. Watch out for [GECC](https://www.ibhe.org/board/2019/03March2019/Item_G-3-GECC_update.pdf) degrees- they should be marked as "Certificate or Other", not associates
       2. An AAS from Berkeley College in NY is an Associates
    3. Bachelors
       1. Watch out for some of the schools that name their degrees something different.  
          i.e. Harvard and University of Miami give "AB" instead of BA, and Dominican University abbreviates Bachelors as "BCHL"
       2. Worcester State University sometimes returns BAs in teaching as something along the lines of "Teaching (5-12)" or "FOREIGN LANG:SPANISH(5-12)". But these credentials can also be post-bacs
       3. Juniata College in PA calls their undergraduate majors "POE"s (Programs of Emphasis). These are bachelors degrees
    4. Masters
       1. Watch out for some of the schools that name their degrees something different.  
          i.e. Harvard gives "AM" instead of MA
    5. Doctorate
    6. No Data [**only if Graduated = Y and Graduation Date is not blank; otherwise this field would be blank, meaning the student had not graduated this would be blank**]
11. Save the file as a csv
12. Upsert the records into the NSC Records object, using the External ID from NSC field as the key
    1. Select the NSC Record Object and select UPsert
    2. Map External ID from NSC to External ID from NSC (will auto-map in Dataloder.io)
    3. Unmap Student Progression
    4. Map Requester Return Field to Student Progression
       1. Lookup via Student Progression: Student Progression ID
    5. Map 2-year/4-year to 2 or 4 Year College
    6. Map Public/Private to Type of College
    7. The only unmapped column from the spreadsheet will be Student Progression (see above)

\* Often the NSC returns multiple of the exact same enrollment record. so you may have two records with the same unique ID. This is fine, the upsert process will reject the duplicates.

13.  You may see some errors from your NSC Record Upsert. If this happens, you will run an NSC Record                 UPDATE with the error records.

* Save the error file from dataloader.io as a CSV
* Set up an NSC Record Update
* Map External ID from NSC to \*Record ID\* and lookup via External ID from NSC
  + this will be the key "Use ID as key for update
* As before, map Requester Return Field to Student Progression and lookup via Student Progression: Student Progression ID
* Map 2-year / 4-year to 2 or 4 year College
* Map Public/Private to Type of College
* The only unmapped fields will be ERRORS and Student Progression
* You can ignore "record ID not found" errors at this point and move on to the next step.

**Data prep- Part 2: Persistence Records**

Because the NSC can provide data that is not quite aligned with persistence reporting requirements, once the data is uploaded and stored in SF, additional information should be populated on the linked Student Progression record. \*Can use this report: <https://na88.lightning.force.com/lightning/r/Report/00O1Y000007HhpnUAC/view?queryScope=userFolders> [NSC Persistence from NSC to SP](https://na88.lightning.force.com/lightning/r/Report/00O1Y000007HhpnUAC/view?queryScope=userFolders)

1. Create a Student Progression with NSC report: all student progressions where high school grad year is [year in question, eg, 2019] AND (edit for 3 exports, identified below)
   1. Fall of Freshman Year
      * nsc enrollment date (Enrollment Begin Date) >= 8/1/2019
      * nsc enrollment date (Enrollment Begin Date) <= 12/31/2019
   2. Spring of Freshman Year
      * nsc enrollment date (Enrollment Begin Date) >= 1/1/2020
      * nsc enrollment date (Enrollment Begin Date) <= 4/30/2020
   3. Fall of Sophomore Year
      * nsc enrollment date (Enrollment Begin Date) >= 8/1/2020
      * nsc enrollment date (Enrollment Begin Date) <= 12/31/2020

For batch 1 (matching filter criteria set 1 above) - update Student Progressions as follows:

* + Set NSC Semester 1 - Term to "Fall"
  + Set NSC Semester 1 - Year to "2019"

For batch 2 (matching filter criteria set 2 above)

* + Set NSC Semester 2 - Term to "Spring"
  + Set NSC Semester 2 - Year to "2019" (\*\*\* important that you upload the STARTING YEAR of the acad year notation, NOT the actual calendar year)

For batch 3 (matching filter criteria set 3 above)

* + Set NSC Semester 3 - Term to "Fall"
  + Set NSC Semester 3 - Year to "2020"

DO NOT UPLOAD ANY OTHER DATES - everything is to be based off of FTFT. Either they enrolled that semester or they didn't, we don't start the clock randomly.

**Data prep - Part 3: Graduation Date Updates**

There are currently several fields where the graduation dates are stored. Whenever we get new info from the NSC, we update these dates (current practice as of 3/20/2020).

Run a Student Progressions with Contacts report that shows all Student Progression records where Graduation Year (Numeric) is null and NSC Date First Bachelors is not null. Include Student Progression ID, Contact ID, NSC Date First Bachelors. Use this report: <https://na88.lightning.force.com/lightning/r/Report/00O1Y000006q5sjUAA/view> [NSC Grad Populate StuPro and Contact](https://na88.lightning.force.com/lightning/r/Report/00O1Y000006q5sjUAA/view)

Export to Excel.

1. Create a new column labeled Graduation Year (Numeric). Set this to the end of the academic year, with graduations dates running from Nov 1 - Oct 31. For example, if an NSC Date First Bachelors is 10/10/2019, the Graduation Year (Numeric) is 2019 (because it falls in the 2018-19 academic year). If the NSC Date First Bachelors is 11/1/2019, the Graduation Year (Numeric) is 2020 (because it falls in the 2019-20 academic year).
2. Create a new field labeled NSC Confirmed Date Earned BA. Copy the NSC Date First Bachelors into this field.
3. Update Student Progressions and populate the Graduation Year (Numeric).
4. Update Contact and populate the NSC Confirmed Date Earned BA.

Create another report - Student Progressions with Success Lookup - where Graduation Year (Numeric) is not null and Stage is not Graduated or Timed Out. Include Program Record ID and Stage. Use this report: <https://bottom-line.lightning.force.com/lightning/r/Report/00O5d000007BbeyEAC/view>

Update the Stage to "Graduated" and update the Program Record.