Dear SER Workshop Participant,

Thank you for your interest and participation in our workshop “An introduction to transporting treatment effects from randomized clinical trials to clinical practice” scheduled for **Friday, January 8, 2021 from 12-2pm ET**.

To make this workshop as interactive as possible, we wanted to provide you with access to “real data” to use for the workshop exercises. To that end, you will be required to take a few steps to apply for and obtain access to the two data sources that we will be using during the workshop:

1. Clinical trial data accessed through Project Data Sphere: <https://data.projectdatasphere.org/projectdatasphere/html/home>
2. Surveillance, Epidemiology, and End Results Program (SEER) cancer registry data: <https://seer.cancer.gov/seertrack/data/request/>

Below you will find a step-by-step guide to applying for an obtaining access to the two data sources used for the workshop. Please note: you will be asked to download the relevant trial dataset from Project Data Sphere; however, we will be providing you with a SEER cancer registry extract, so no download on your side is required (i.e., you only need to receive approval for data access).

Please follow the instructions listed below to obtain access to the two data resources. Once you receive email confirmation of access to both data sources, please save and copy those emails into a document and upload them to the Google form that can be found here: [Insert link]

**The deadline for uploading the approval documents is Monday, January 4, 2021.**

We will be sending another set of workshop material to all participants who have received approvals on Wednesday, January 6, 2021, so please ensure that all approvals are submitted prior to this date.

If you have any questions or encounter problems with the approval process, please do not hesitate to contact the lead course instructor, Jennifer Lund ([Jennifer.Lund@unc.edu](mailto:Jennifer.Lund@unc.edu)).

Once again, thank you for your interest and participation in this workshop. We are excited to “meet” you in January!

Sincerely,

Jennifer Lund (course lead)

Michael Webster-Clark, Alexander Keil, Daniel Westreich (co-instructors)

Shahar Shmuel, Allison Musty (workshop support)

**Data Source #1: Project Data Sphere Clinical Trial Data**

# **Applying for Access to Project Data Sphere (Trial Data)**

Please note: it may take up to 7 business days for access to these data to be granted, so plan accordingly!

In this workshop, we will provide an overview of methods for generalizing and transporting treatment effects from randomized clinical trials (RCTs) to defined target populations of patients in clinical practice. Participants will receive SAS and R code to combine publicly available phase III RCT and patient registry data.

However, for these data to be accessible, each participant must apply for access. This guide will take participants through the process of applying for, obtaining, and accessing the necessary data for this workshop.

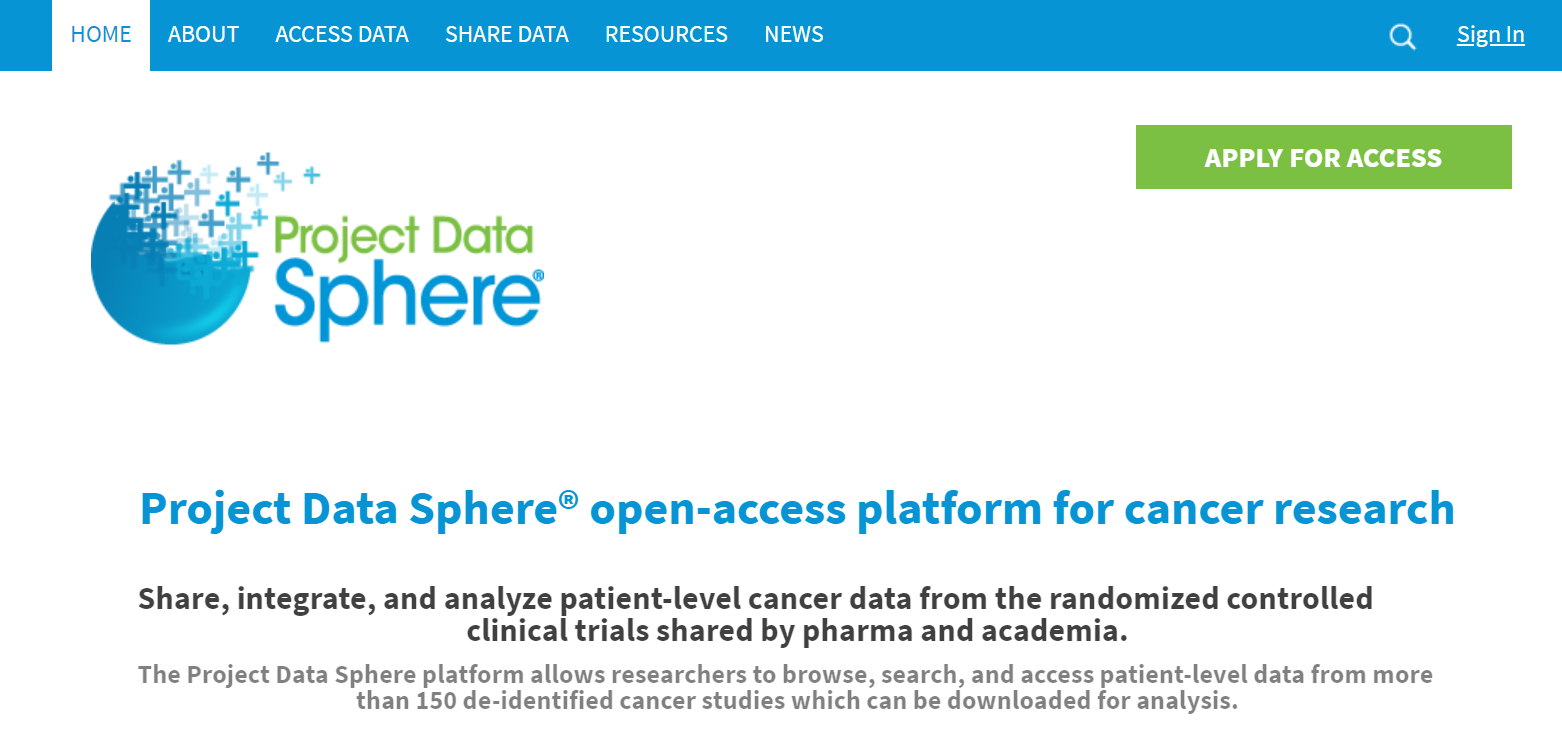
It is recommended that one creates a folder for this workshop which can easily be accessed, in order to keep downloaded data organized.

## STEP 1: Applying for Data Sphere Access.

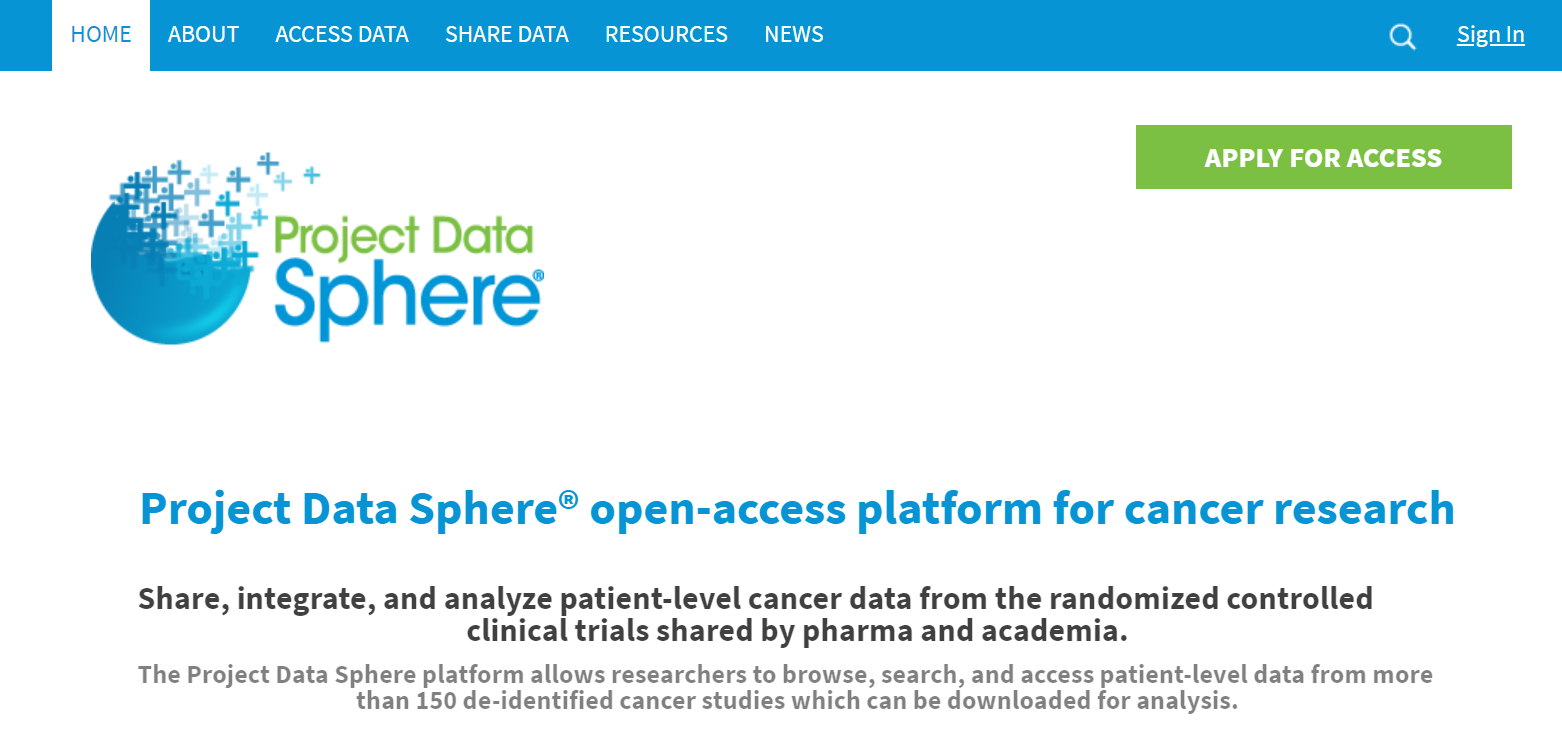
If one already has Project Data Sphere access, feel free to skip to the next section of this guide.

Before beginning, ensure that information on place of work/education is on hand, as this will accelerate the process. Application information for this workshop will be provided below.

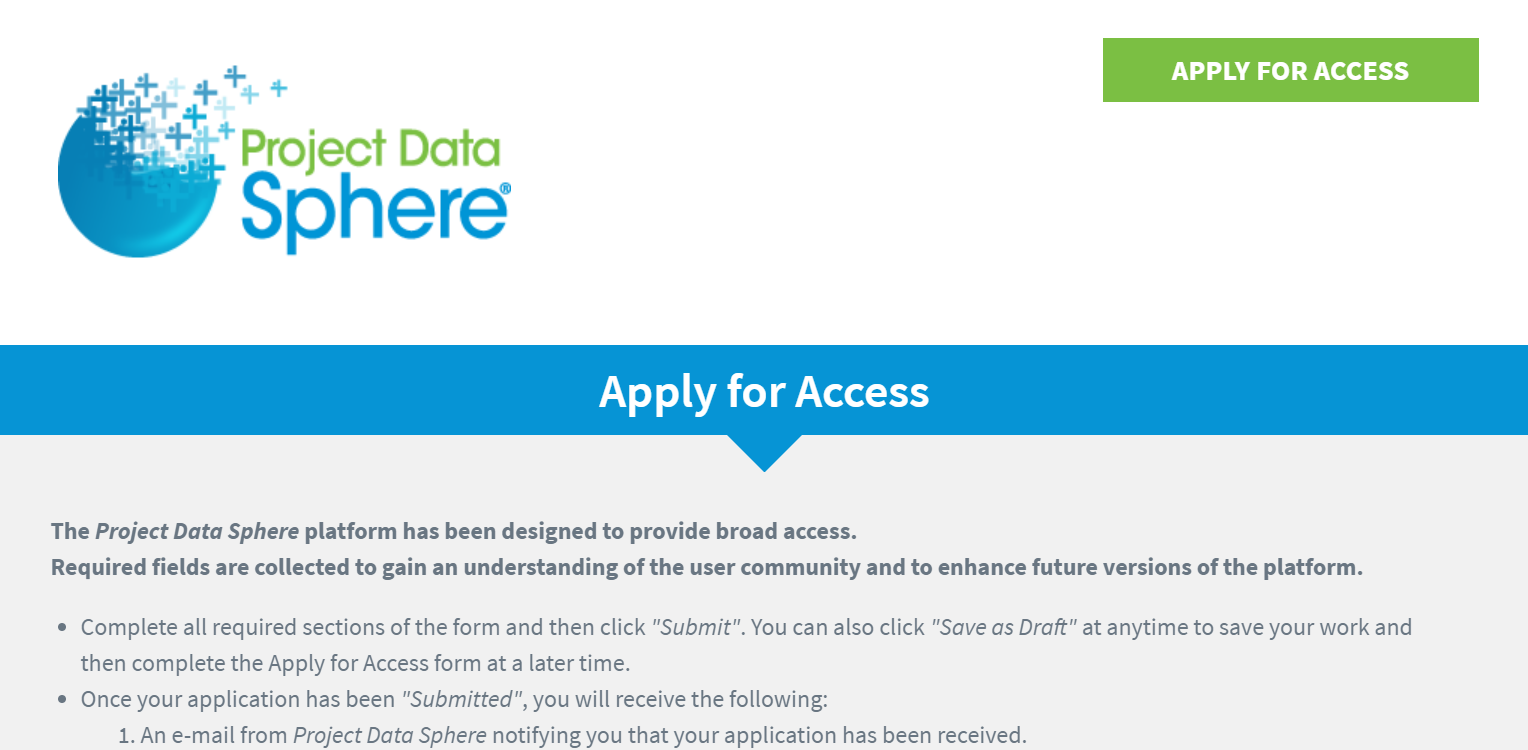
First, visit <https://data.projectdatasphere.org>. This will bring up the following webpage:



Next, select “apply for access” in the upper right-hand corner.

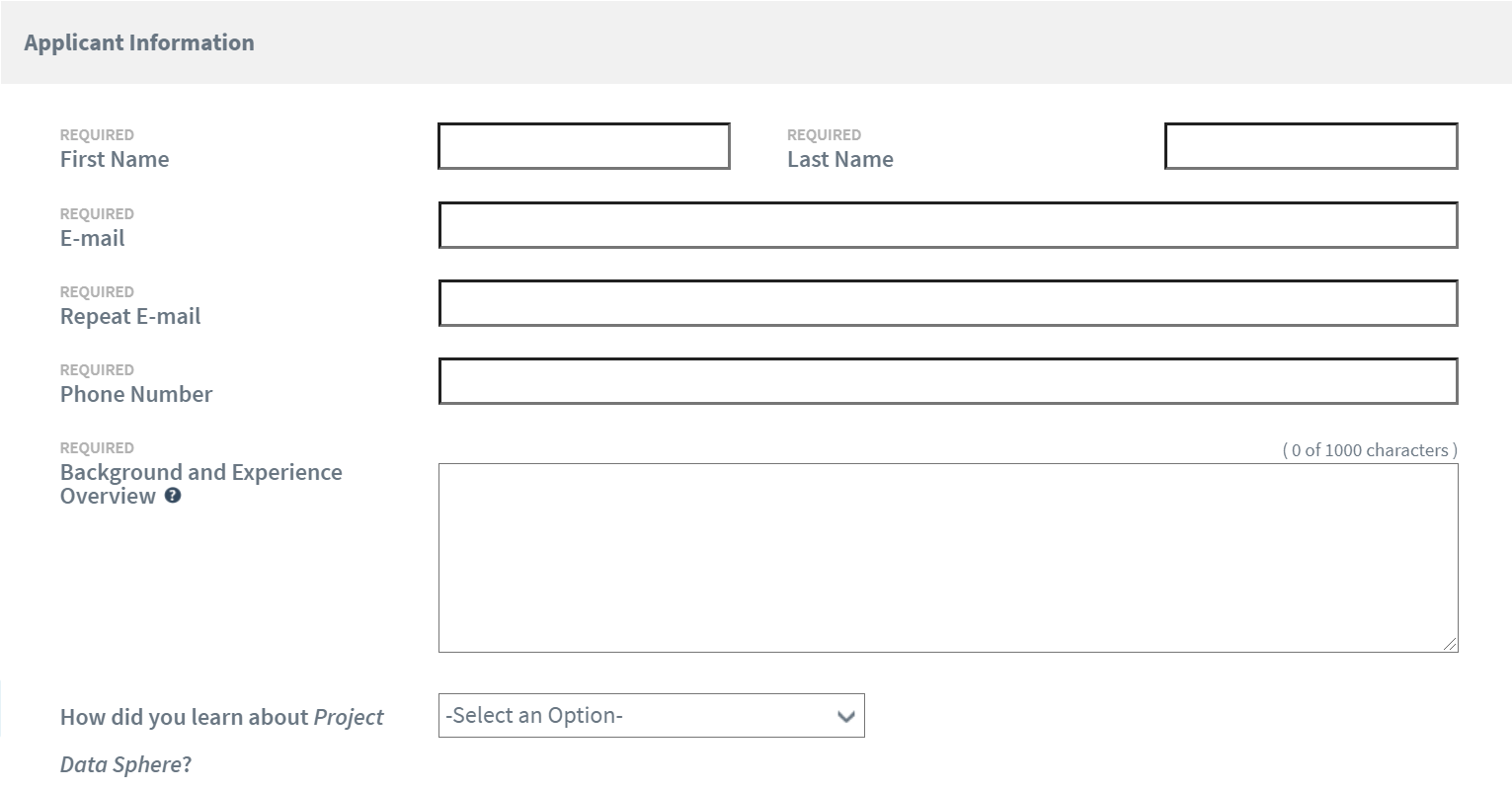


This will bring up the following page.



Scroll until the “applicant information” fields are visible as seen below. Complete all of the required fields.

To streamline the process, the “research description” and “research goals” information is provided below. Please copy/paste it into the appropriately labeled fields on the application.



Research description: “When response to therapy varies across subgroups, differences between trial and clinical population composition can contribute to the “efficacy-effectiveness gap” – where a treatment’s efficacy in a trial differs from its effectiveness in clinical practice. Methods for generalizability and transportability can help bridge this gap. I will use data from Project Data Sphere to combine RCT and publicly available cancer registry target population data to visualize differences in patient characteristics between the trial and target populations and to evaluate therapy effectiveness for a specific target population.”

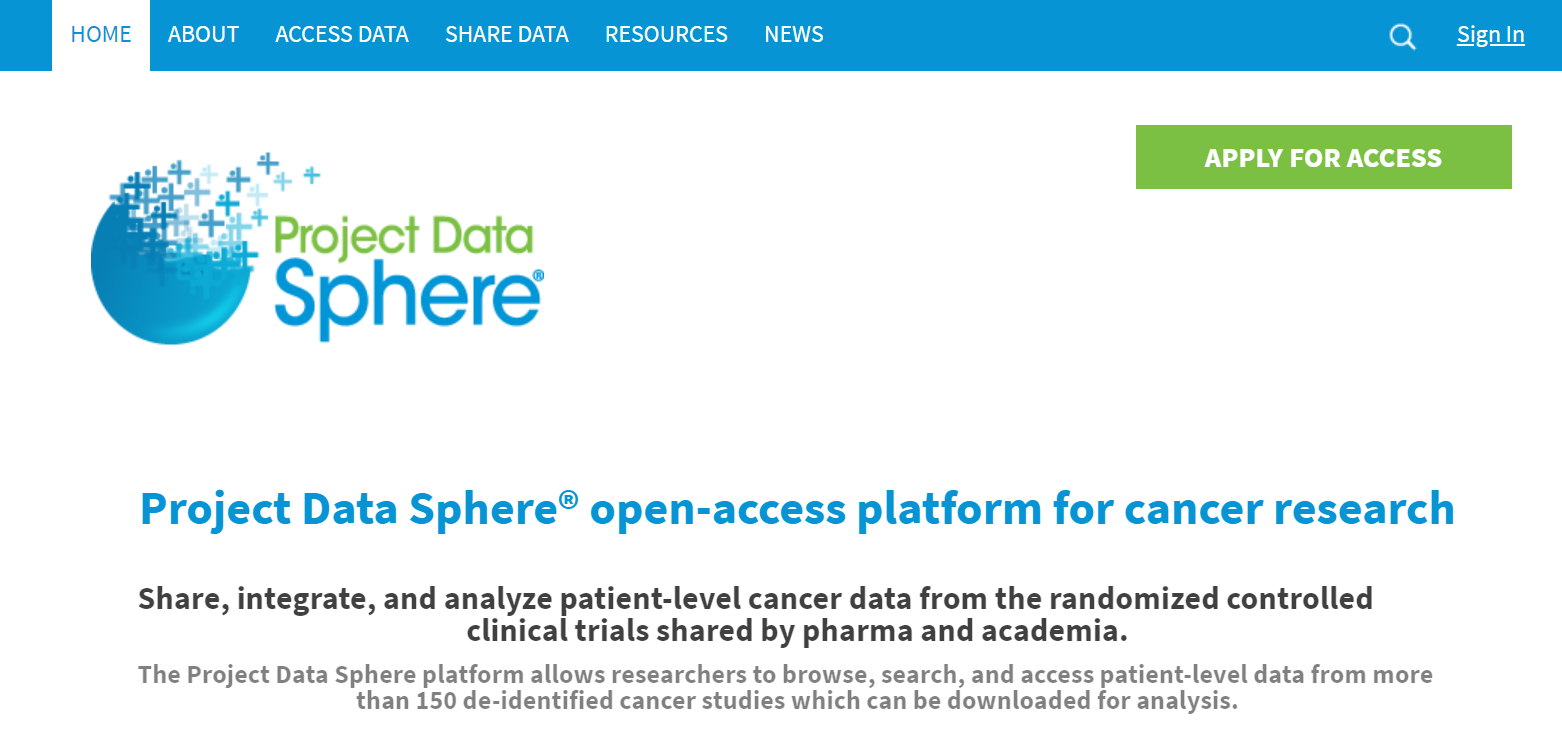
Research goals: “I will be participating in a workshop on January 8, 2021 as part of the Society for Epidemiologic Research (SER) Annual Meeting. During this workshop, I will (1) learn about the foundations of epidemiologic methods for generalizability and transportability, (2) use visual tools to characterize trial and target populations and inform variable selection for transport models, and (3) apply this knowledge to an example of a trial for chemotherapy in metastatic colorectal cancer.”

Once finished, agree to the Data User Agreement and submit.

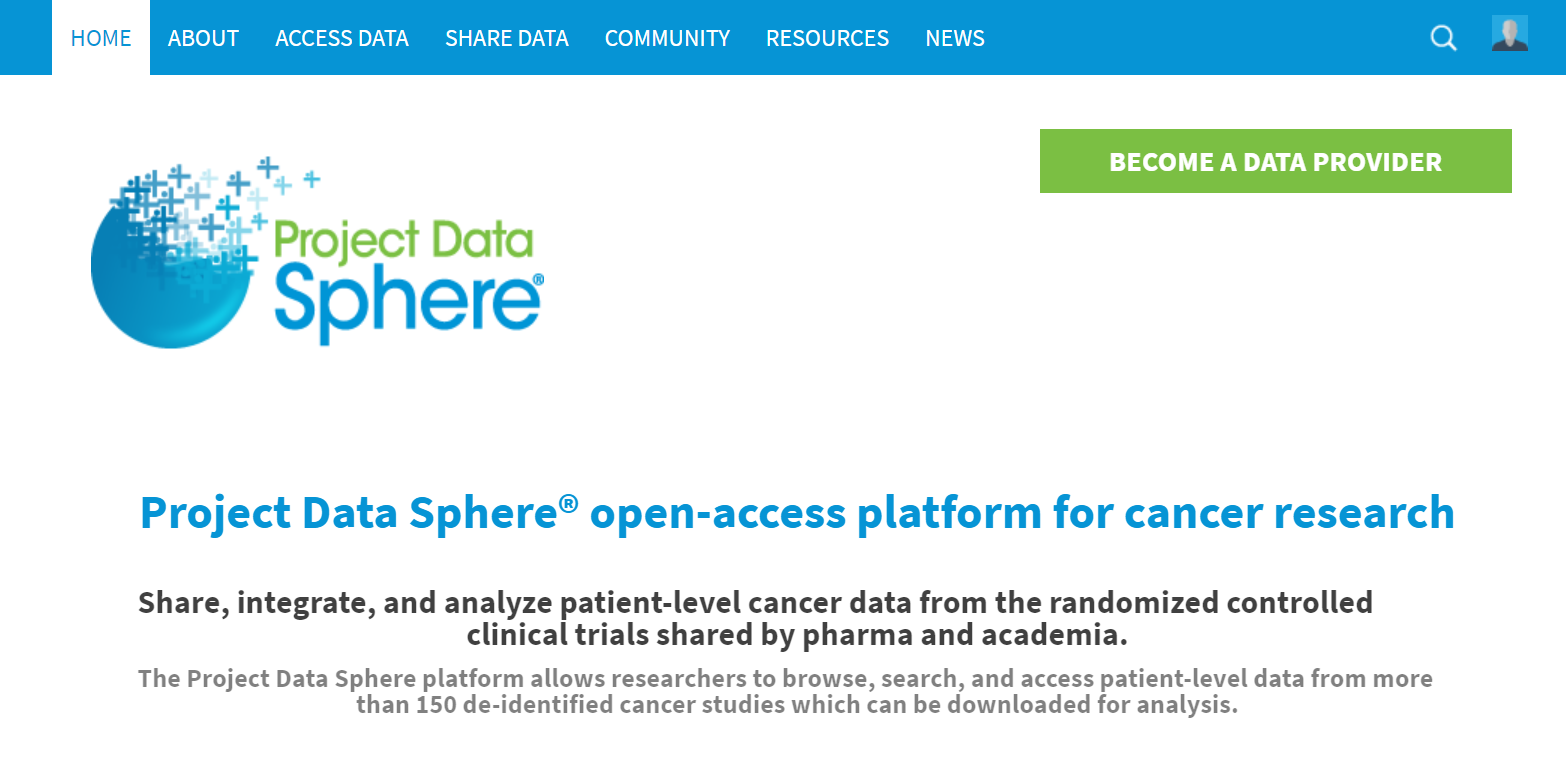
Following submission, please allow up to 7 days for the application to be accepted. Once the application is accepted, you will receive an email at the address provided in the application.

# STEP 2: Downloading the Data.

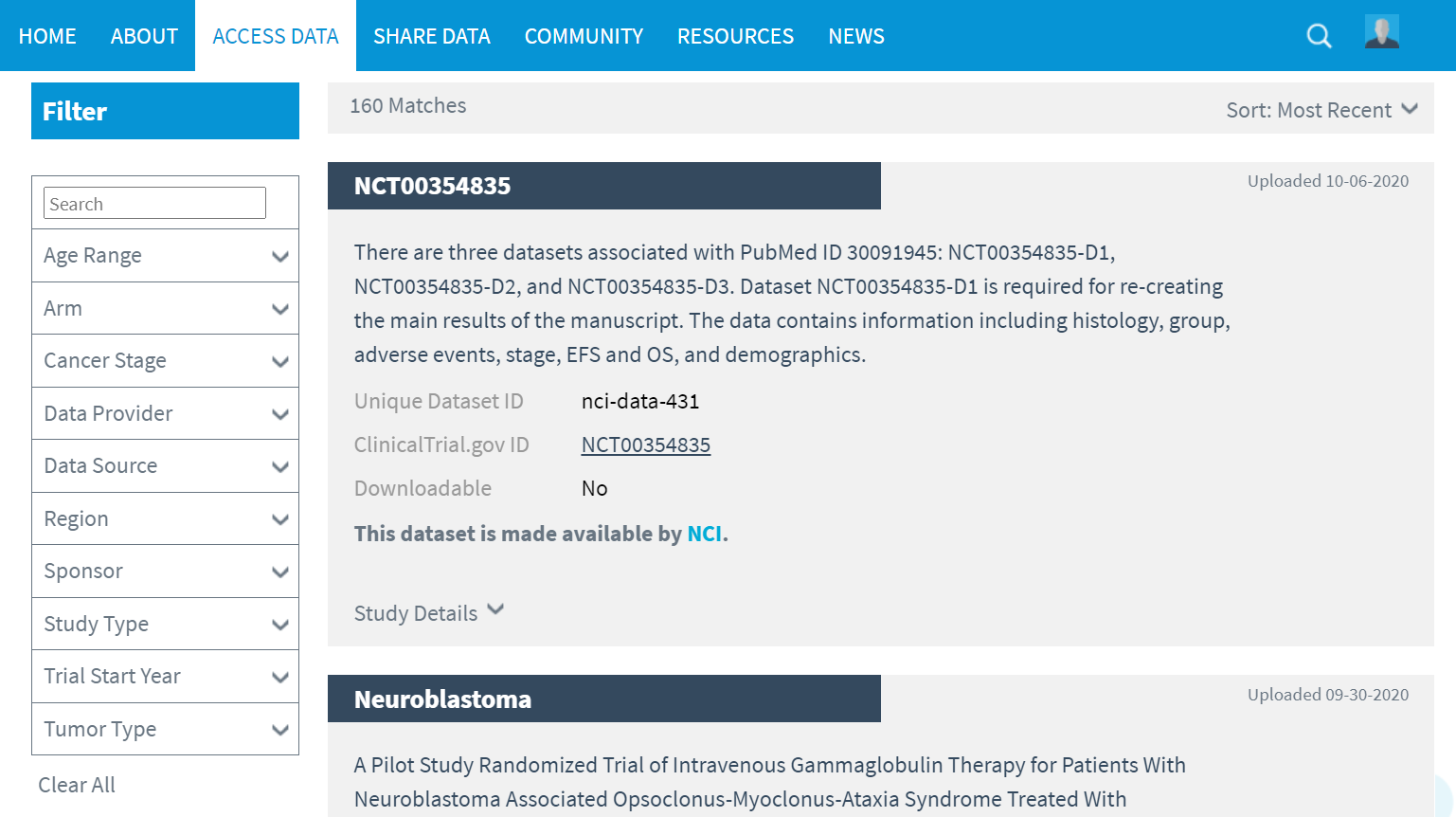
Return to <https://data.projectdatasphere.org> and sign in using the link in the upper right-hand corner.



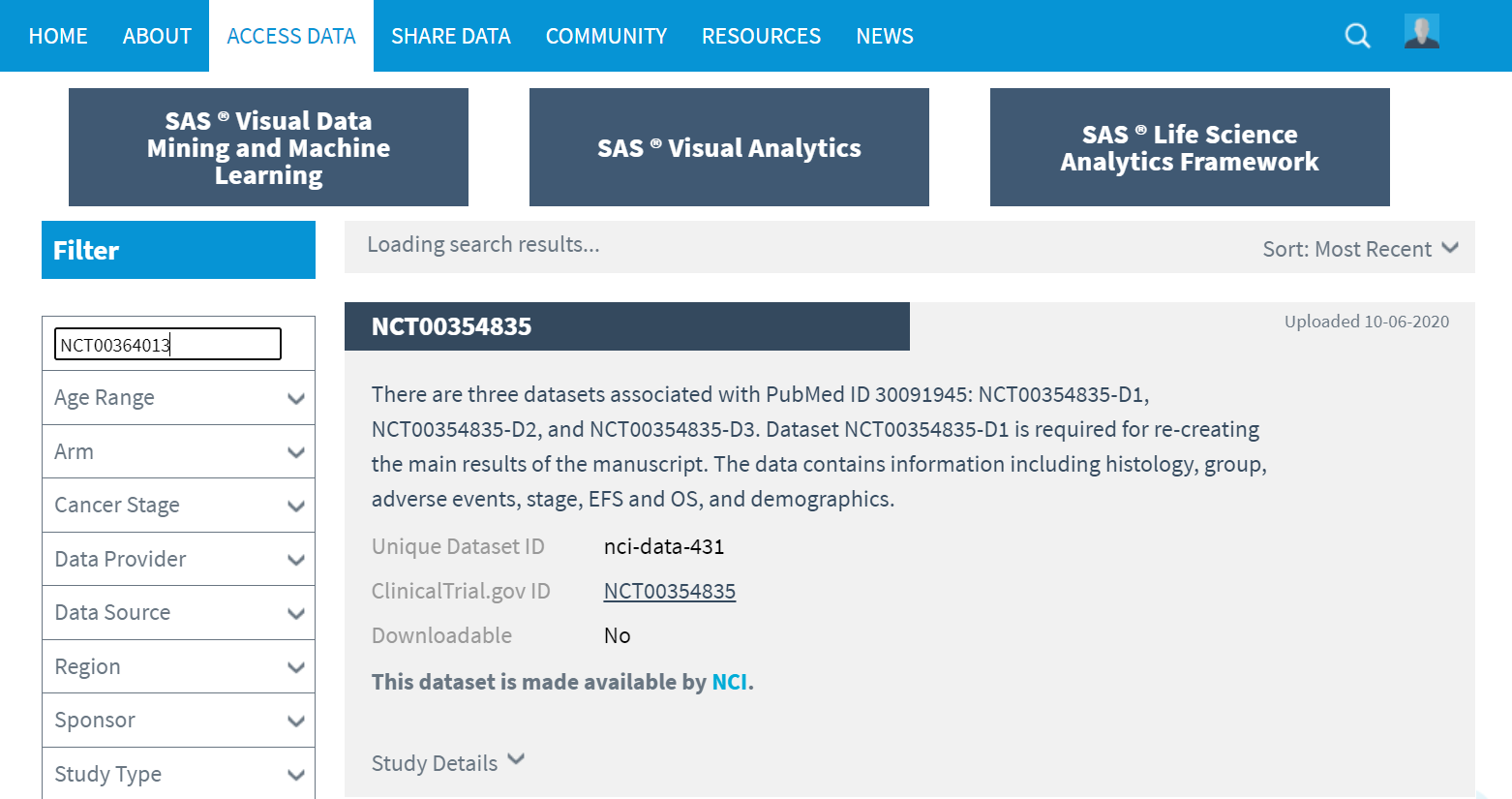
Then, select “access data” from the blue banner along the top.



Scroll until the data filters are visible on the left.



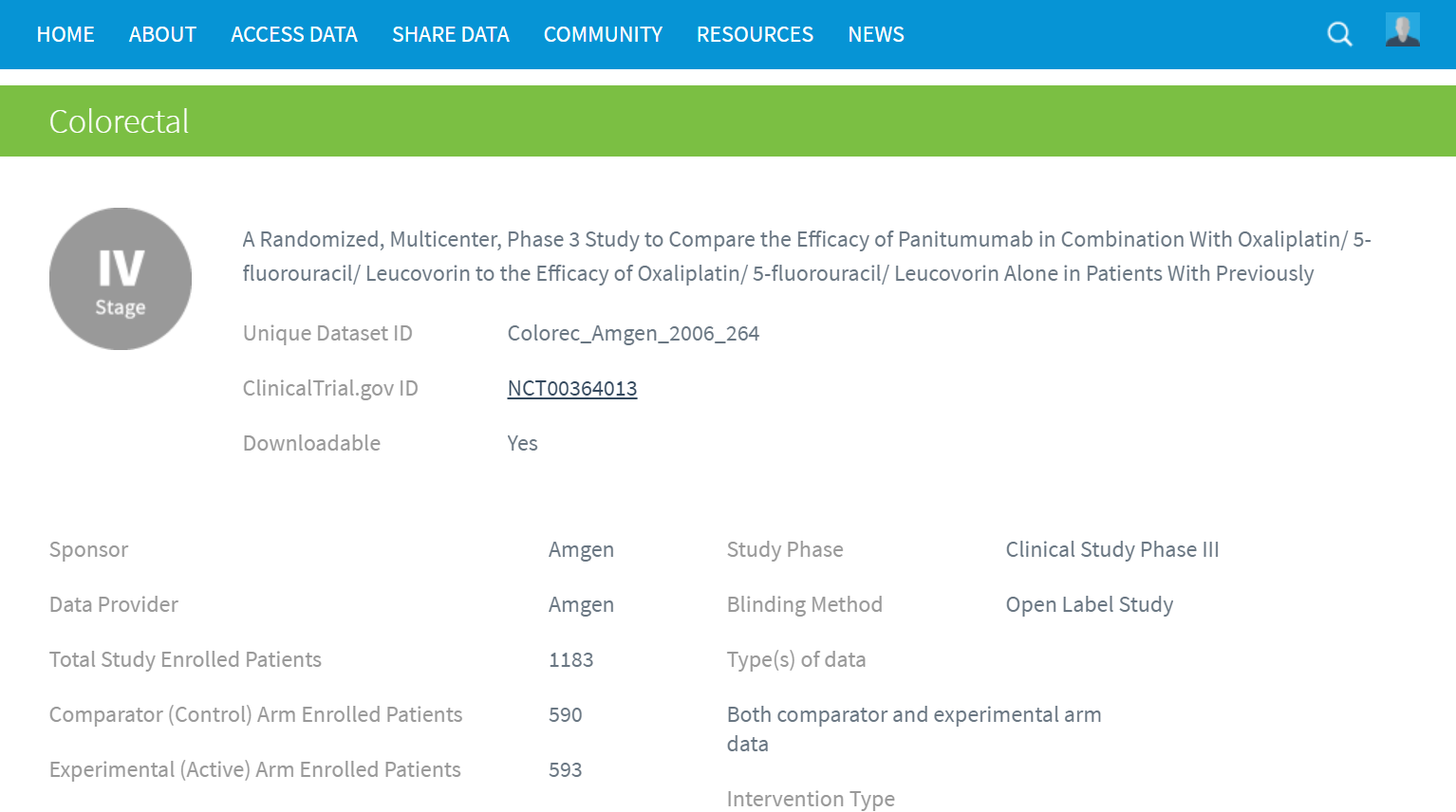
In the search bar, type the study ID: “NCT00364013”



This will bring up two results. Scroll down and select the second result. This study will possess the title “A Randomized, Multicenter, Phase 3 Study to Compare the Efficacy of Panitumumab in Combination…”



This should bring up the following page.



Only certain sections of the dataset must be downloaded. Scroll down to the bottom of the page, under “SAS dataset - 20050203.zip Contents”



Select “download” for the first data file in the list: “a\_eendpt.sas7bdat” Please note that this data is in SAS format. If using R, conversion would be required.



**Data Source #2: Surveillance, Epidemiology, and End Results Program (SEER) Cancer Registry Data**

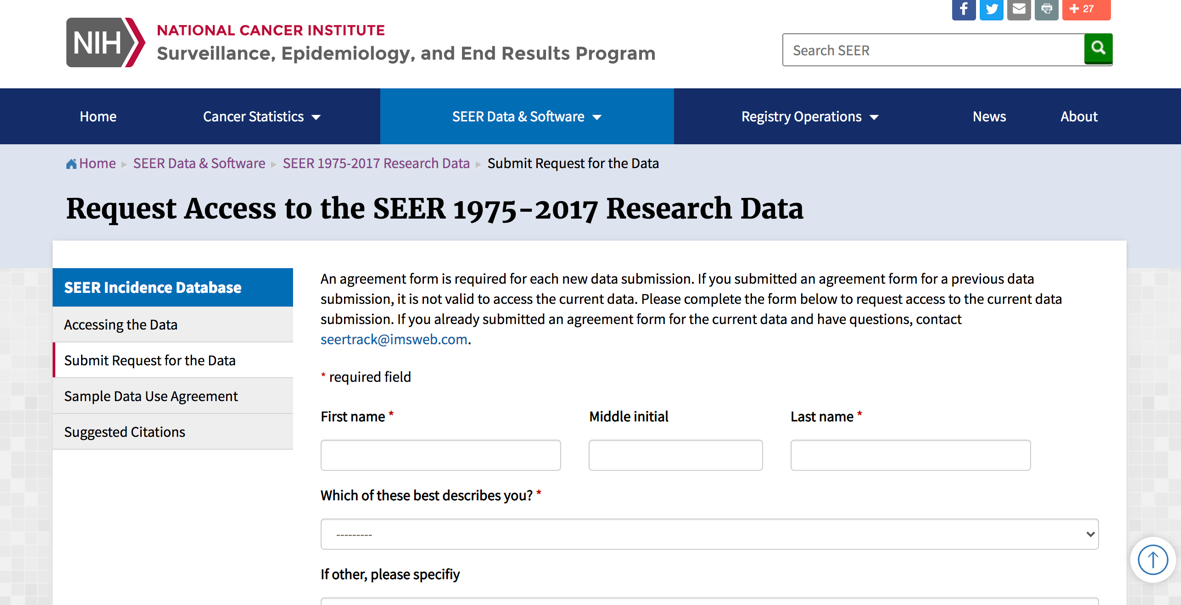
**Applying for Access to the Surveillance, Epidemiology and End Results Program (SEER) Cancer Registry Research Data (Target Population Data)**

Accessing the SEER data is 2-step process.

|  |  |
| --- | --- |
|  | **Before you begin, please note:**   1. It can take **2 business days** to be granted access, so please plan accordingly. |

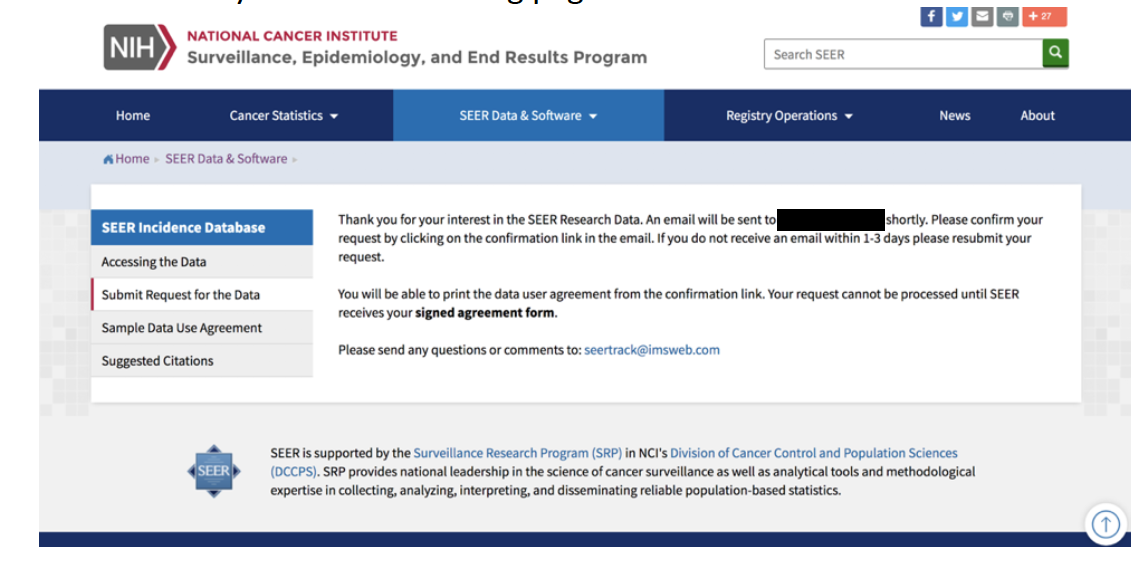
**STEP 1: Submit a Request.**

1. Complete the agreement form (<https://seer.cancer.gov/seertrack/data/request/>) by filling in all required fields:

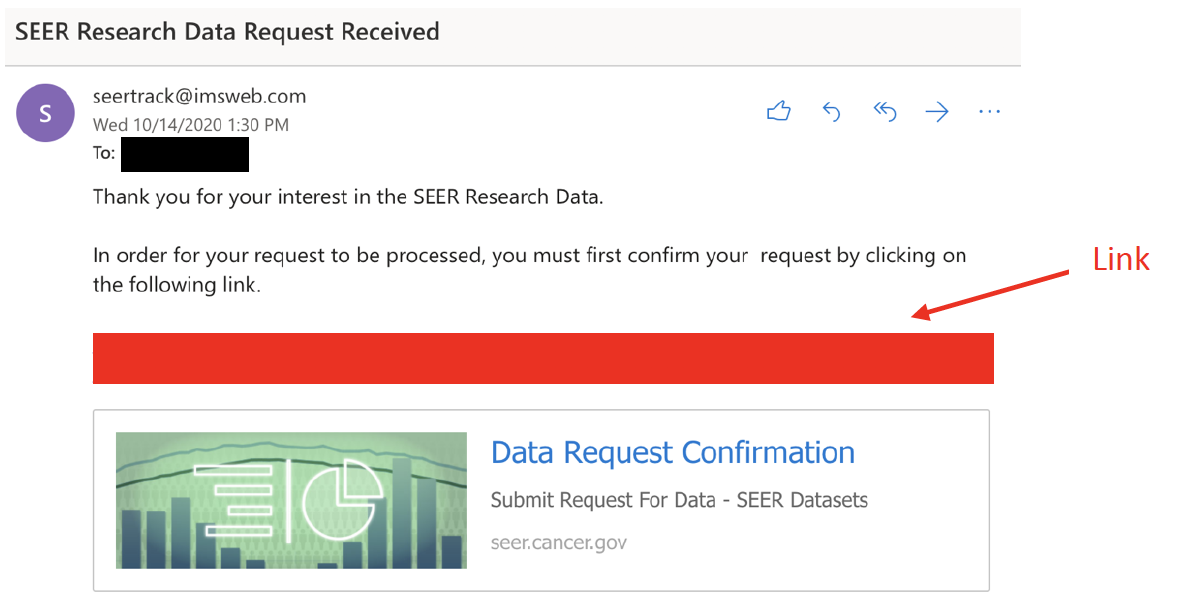


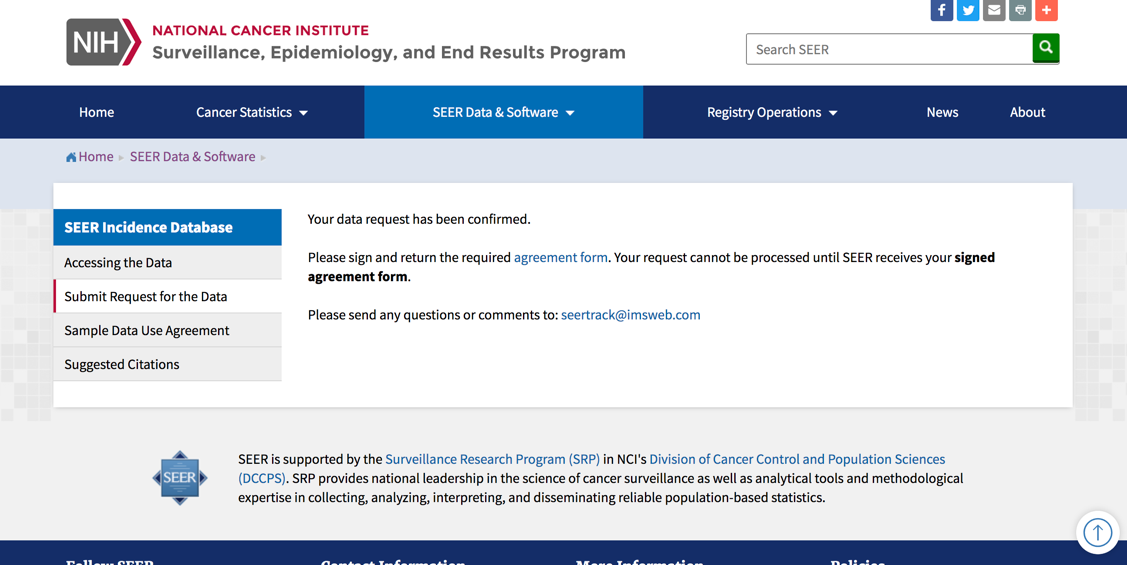
and press .

1. This will lead you to the following page:



1. You should then receive an email that looks like this:

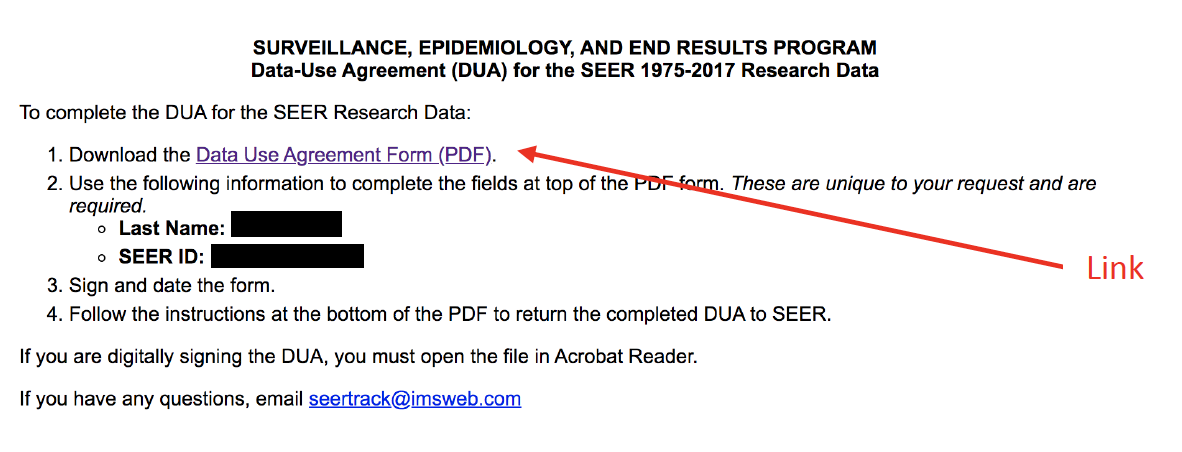


1. Click the link provided in the email above.
2. This should bring you to the following webpage:
3. 

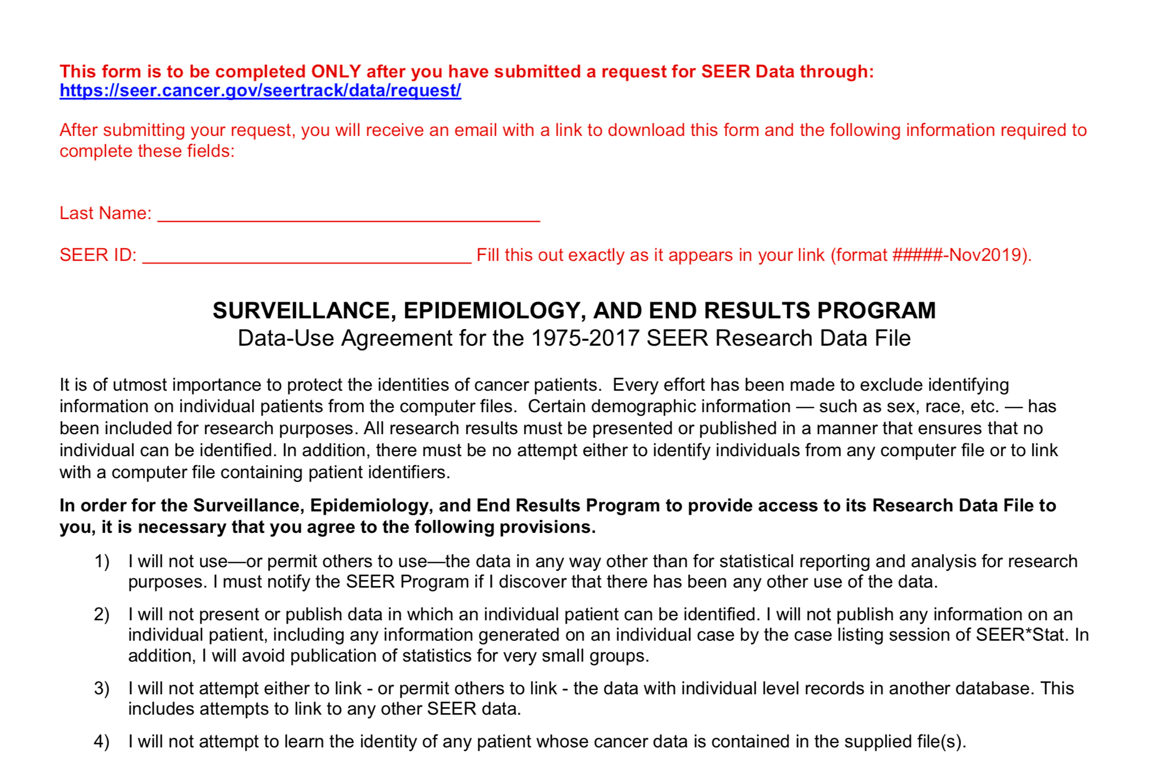
Link

**STEP 2: Complete and Return the SEER Research DUA**

1. Click on the “agreement form” link. This should bring you to the following webpage:



1. Click on the “Data Use Agreement Form (PDF)” link, fill it out (SEER ID was provided on the previous screen), and sign it:



1. Send the PDF back to the SEER Program via email ([seerfax@imsweb.com](mailto:seerfax@imsweb.com)). **Note:** you will also receive your SEER ID and a link to the PDF by email, in case you accidentally close out of the previous window.
2. Within **2 business days**, you should receive the following email containing your username and password for accessing the data:

