



Hub Explore User Guide & Data Access Instructions

Content

Introduction.....	2
Data Access.....	2
CCDI Hub Explore Dashboard and Cart	3
Finding Participants, Studies, Samples, and Files.....	4
Creating and managing cohorts	7
Downloading Metadata from the Studies tab.....	8
Creating an Exportable File Manifest from the Cart	9
Accessing Help Documentation.....	11
Additional search features	11
Contact Information	12
Appendix A: database of Genotypes and Phenotypes (dbGaP)	14
Appendix B: NCI Data Commons Framework Services (DCFS): Controlled Data Access Instructions	15
File Download Procedure via User Interface.....	15
File Download Procedure via Command Line Interface (CLI) client	17
Appendix C: The Cancer Genomics Cloud (CGC)	18
Creating a CGC Account.....	18
Creating a DRS Manifest from downloaded Study metadata	22
Additional Resources on Working with Data at the CGC	30

Introduction

This document provides information about how to use the [Childhood Cancer Data Initiative \(CCDI\) Hub Explore Dashboard](#), including how to find, download, and export manifests for open CCDI data, and submit requests for controlled access data.

CCDI studies are indexed in the CCDI Hub Explore Dashboard at the file level, where row-level metadata for CCDI participants, diagnoses, studies, samples, and files can be exported. The shopping cart feature on the dashboard allows users to select and manage files of interest and generate a comma-separated values (CSV) file manifest. This manifest file can be downloaded locally or uploaded to the [Cancer Genomics Cloud](#) (CGC) for downstream data analysis. CGC is a cloud platform that enables the analysis, storage, and computation of large cancer data sets.

Additional CCDI applications and other resources are described in informational pages and available through the various Hub menus. Related resources—internal and external to CCDI—are linked where relevant. Application and data update summaries and other announcements are available to browse from the [News](#) page, with full release details available from the Release Notes page. Additional user documentation details general site usage and data use terms and best practices.

A list of all CCDI studies released is available through the [CCDI Childhood Cancer Data Catalog](#).

Data Access

The guiding principle of the [National Institutes of Health \(NIH\) Data Sharing Policy](#) is to make data available in a timely manner to the largest possible number of investigators. For human data, data are made available under terms and conditions consistent with the participants' informed consent, and the confidentiality of the data and the privacy of participants are maintained.

For CCDI, some resources contain open-access data, while others contain registered-access or controlled-access data sets.

Open Access: For public access; requires no special credentials

Examples: CCDI Childhood Cancer Data Catalog, CCDI Molecular Targets Platform

Registered Access: For anyone registered with the repository; usage may be monitored

Example: CCDI National Childhood Cancer Registry Explorer

Controlled Access: For credentialed users who have applied to access the data

Example: CCDI genomic data stored at the Cancer Data Service

To gain access to controlled data, researchers must:

1. Have an [NIH eRA Commons account](#) for authentication.
2. Request authorization by the NIH database of Genotypes and Phenotypes (dbGaP). Authorization is enforced by the Data Commons Framework Services (DCFS), whether accessing directly from DCFS or through the CGC. A step-by-step breakdown of the data access request process is located in this [guide](#).

Below is a guide to help you understand how these platforms are used to connect different components of a CCDI study:

Platform	Data Types
CCDI Hub Explore Dashboard	Basic de-identified information on participant, samples, files, etc., to build cohorts
database of Genotypes and Phenotypes (dbGaP)	CCDI study information, list of CCDI study subject IDs, sample IDs, and consents to register the controlled-access studies in dbGaP system
Data Commons Framework Services (DCFS)	Globally Unique Identifiers (GUIDs) for digital objects and authentication and authorization services
Cancer Genomics Cloud (CGC)	Tools, computing resources, etc., to analyze the data

Please note that additional information on using these resources for accessing or analyzing CCDI data is provided in the Appendices below; however, the most up-to-date documentation at any time will live with these external resources.

Reach out to the [CCDI mailbox](#) with any questions.

CCDI Hub Explore Dashboard and Cart

The [CCDI Hub Explore Dashboard](#) is a tool that allows for the exploration of participant-level, diagnoses, studies, samples, and files information for CCDI-managed data sets. The Explore Dashboard enables researchers to find CCDI data within a single study or across multiple studies and create synthetic cohorts based on filtered search (i.e., demographics, diagnosis, samples, etc.). Upon interaction with these filters (Figure 1A), users can review the open-access information through visual summaries (Figure 1B) and browse the row level data in tabs organized by participants, diagnosis, studies, samples, and files (Figure 1C) to determine which data sets are applicable to their research questions. Users can then add desired files to the cart (Figure 1D), from which they can download a manifest for the selected data or take the manifest directly into the CGC. To access the controlled data, users must request them at the [controlled-access login page on dbGaP](#). Note that bulk downloads are only possible through Command Line Interface (CLI) client as described in Appendices B and C.

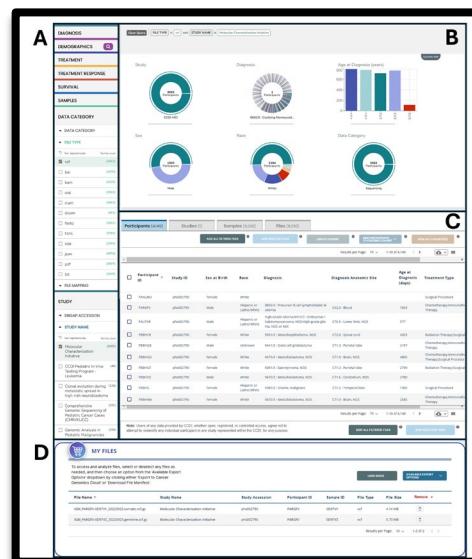


Figure 1: CCDI Hub Explore Dashboard and Cart features

Step-by-step instructions for finding and exporting data are included below.

Finding Participants, Studies, Samples, and Files

The CCDI Hub Explore Dashboard provides row-level metadata for CCDI study participants and their data objects for review with a filtered search, select visualizations, and an exportable table of results. Here's how to find and filter information on the Explore Dashboard:

1. The CCDI Hub is located at ccdi.cancer.gov. From the CCDI Hub Home page, navigate to the Explore Dashboard by clicking "Explore" (Figure 2).

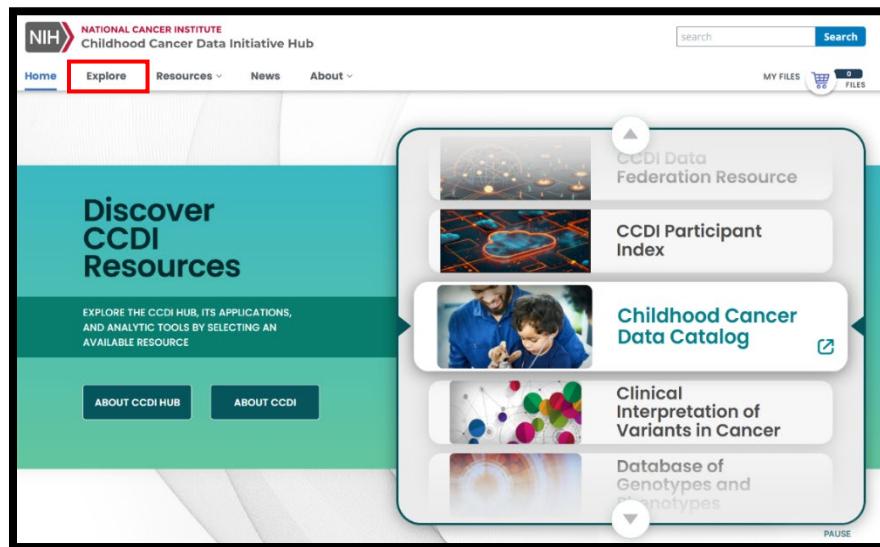


Figure 2: CCDI homepage with red box highlighting the “Explore” menu bar link

2. On the Explore Dashboard, you can filter row-level data and view them as visualizations (Figure 3). The Explore Dashboard is participant-centric, meaning that filtering criteria and results return de-identified information about a participant and their related studies, collected samples, or created files.

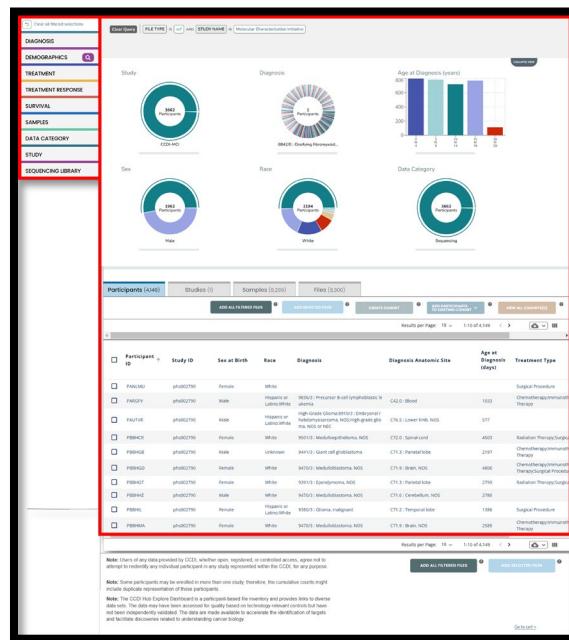


Figure 3: Explore Dashboard page with red boxes highlighting the search filters and results

3. Search criteria are displayed in the right panel (Figure 4A). Faceted filtering may be done by uploading a list of participant IDs (in “DEMOGRAPHICS” Figure 4B), text searches (“DIAGNOSIS,” “DIAGNOSIS ANATOMIC SITE,” and “SAMPLE ANatomic SITE” Figure 4C), numerical sliders (“AGE AT DIAGNOSIS” and “AGE AT COLLECTION” Figure 4D), or checkbox selections for the remaining properties. You can apply multiple filtering criteria at the same time in a search. You can view and clear your current selection(s) in the query summary at the top of the widgets (Figure 4E).

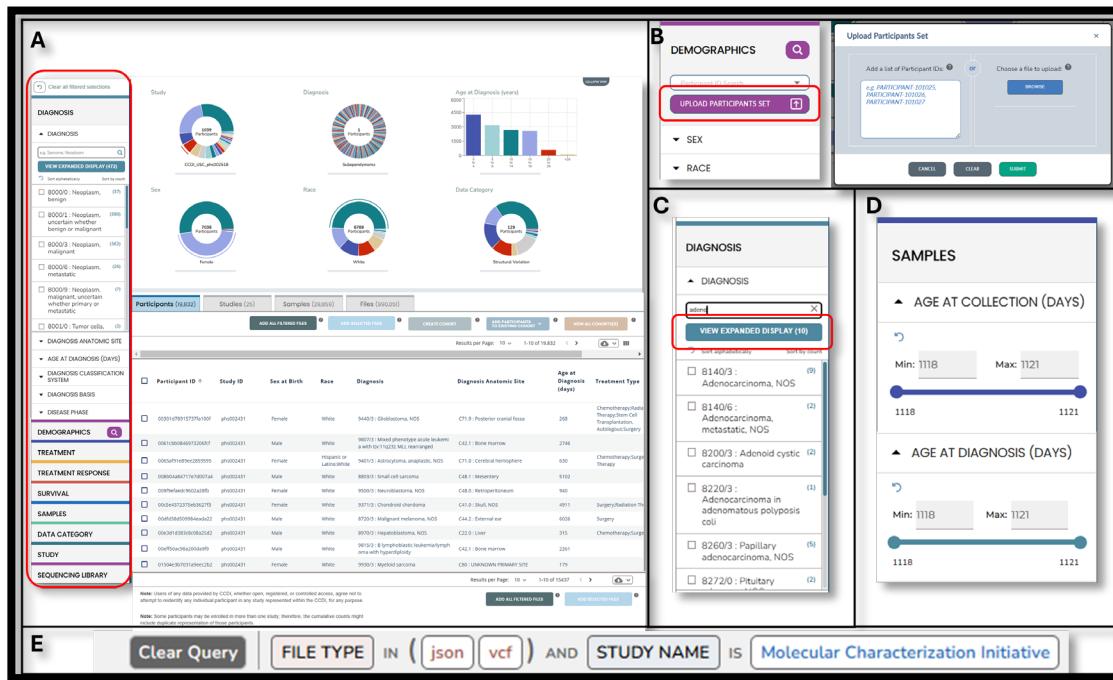


Figure 4: Full facet list in Explore Dashboard with highlights of various facet types and query display/clear function

4. Filtering your search will update the Explore Dashboard’s visualizations and the results tables (Figure 5). Each results table will be updated with information on the participants, samples, studies, or files that meet the filtered criteria. Information displayed by default on each table is described below:

- “Participants”**: Characteristics of a participant in the Explore Dashboard. Participants belong to a study, and they may have one or more samples, diagnoses, or files associated with them. Participants with mappings through the [CCDI Participant Index \(CPI\)](#) have a summary of these mappings available from this view.
- “Studies”**: Studies that are a part of the Explore Dashboard. Participants, diagnosis, samples, and files all belong to a CCDI study.
- “Samples”**: Samples available from participants within the Explore Dashboard. Samples belong to a participant and can be associated with one or more files.
- “Files”**: Files available from studies, participants, and samples within the Explore Dashboard. Files may belong to a study and may be associated with one or more participants or samples. Files may also be of many types, including sequencing, proteomics, imaging files, etc. DICOM imaging files are currently available for the Genomic Sequencing of Pediatric Rhabdomyosarcoma (phs000720) and Molecular Characterization Initiative (phs002790) studies and can be accessed directly from the [Imaging Data Commons \(IDC\) Data Portal](#) and

file paths to images are provided in the downloadable study manifest within Hub, described in the following section.

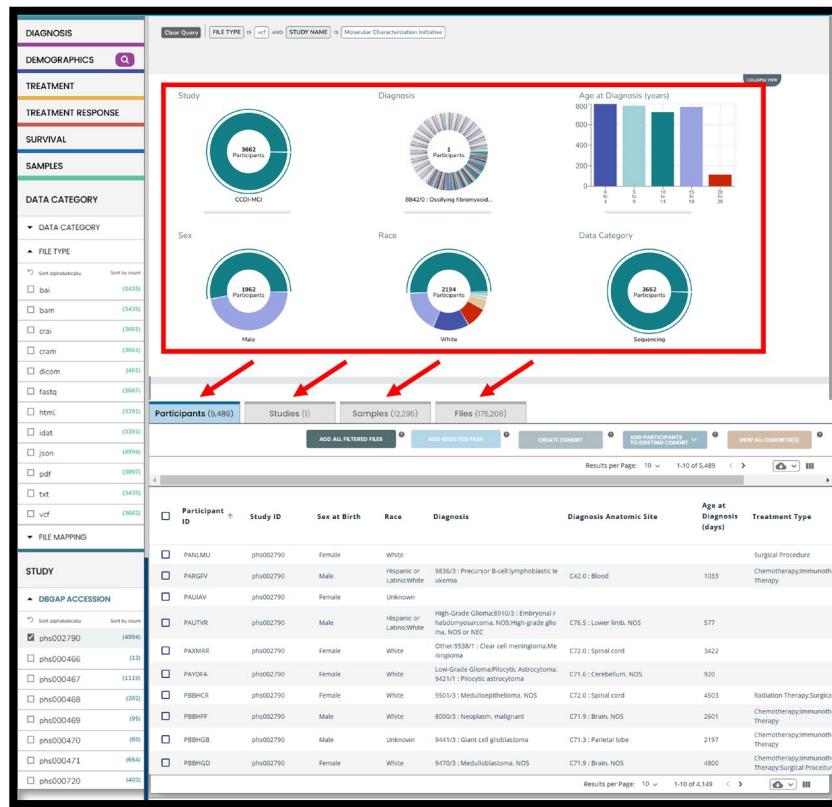


Figure 5: Explore Dashboard visualizations and results tables with arrows pointing to the available informational tables

- Visible columns in each table can be customized by clicking the “View columns” button in the upper righthand corner of the table and selecting or deselecting available columns (Figure 6). Note that Participant ID, Sample ID, and Study ID cannot be removed from any table, and File Name cannot be removed from the Files table.

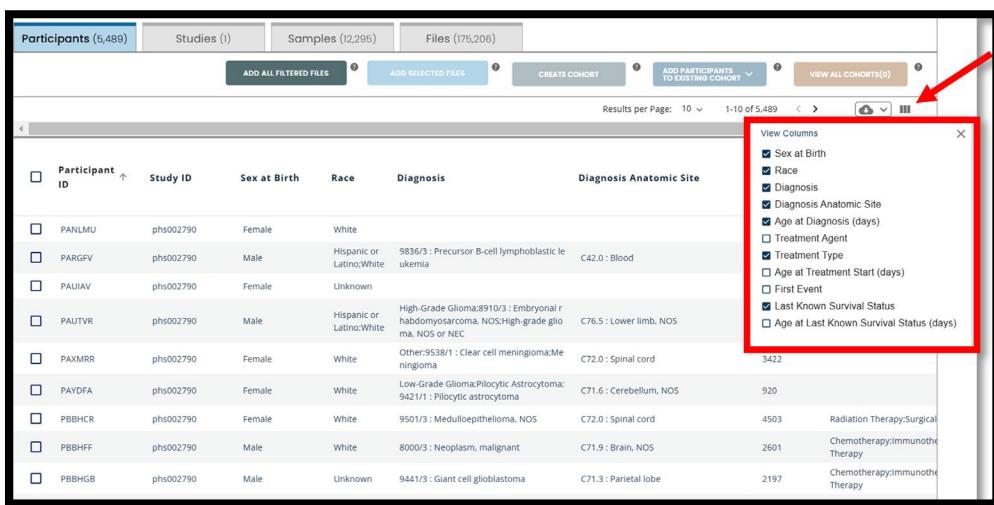


Figure 6: Interface for selecting and deselecting columns in table and downloads

- You can show, hide, and copy the URL used to construct your filtered view using the toggle and copy buttons at the top of the widgets (Figure 7).



Figure 7: Query URL

Creating and managing cohorts

From the CCDI Hub Explore Dashboard Participant table, or you can group participants into cohorts to find files of interest, or you can add files directly to the cart (read more in next section). To create a cohort:

- Using the process described above, apply any filters of interest from the lefthand facet menu.
- Navigate to the Participants table. On the results tables of the Explore Dashboard, you can select a row of metadata using the checkbox at the start of the row. Multiple rows can be selected within a table, even across pages of the table. Use the checkbox at the top of the checkbox column to select or deselect all rows.
- After selecting desired rows, select the “CREATE COHORT” button to add the selected participants to a cohort (Figure 8). A "View of All Cohorts" pop-up window will open in at least one participant row is selected.

	Participant ID	Study ID	Sex at Birth	Race	Diagnosis	Diagnosis Anatomic Site	Age at Diagnosis (days)	Treatment Type
<input checked="" type="checkbox"/>	PANLMU	phs002790	Female	White	9836/3 : Precursor B-cell lymphoblastic leukaemia	C42.0 : Blood	1033	Surgical Procedure
<input checked="" type="checkbox"/>	PARGFV	phs002790	Male	Hispanic or Latino/White	9836/3 : Embryonal rhabdomyosarcoma, NOS; high-grade glioma, NOS or NEC	C76.5 : Lower limb, NOS	577	Chemotherapy/Immunotherapy
<input checked="" type="checkbox"/>	PAUIAV	phs002790	Female	Unknown	Other 9538/1 : Clear cell meningioma; meningioma, NOS or NEC	C72.0 : Spinal cord	3422	
<input checked="" type="checkbox"/>	PAUTVR	phs002790	Male	Hispanic or Latino/White	Low-Grade Glioma; Pilocytic Astrocytoma; 9421/1 : Pilocytic astrocytoma	C71.6 : Cerebellum, NOS	920	
<input checked="" type="checkbox"/>	PAXMRR	phs002790	Female	White	9501/3 : Medulloepithelioma, NOS	C72.0 : Spinal cord	4503	Radiation Therapy/Surgery
<input checked="" type="checkbox"/>	PAYDFA	phs002790	Female	White	9800/3 : Neoplasm, malignant	C71.0 : Brain, NOS	2691	Chemotherapy/Immunotherapy
<input checked="" type="checkbox"/>	PBBHCR	phs002790	Female	White	9441/3 : Giant cell glioblastoma	C71.3 : Parietal lobe	2197	Chemotherapy/Immunotherapy
<input checked="" type="checkbox"/>	PBHHFF	phs002790	Male	Unknown				
<input checked="" type="checkbox"/>	PBBHGB	phs002790	Male	Unknown	9441/3 : Embryonal rhabdomyosarcoma, NOS; high-grade glioma, NOS or NEC			

Figure 8: Cohort creation and management

- From the "View of All Cohorts" window, you can view and delete all cohorts and see details about a selected cohort, which will be your newly created cohort, by default (Figure 9).
- In the selected cohort view, you can view current cohort attributes as well as change the name, add a description, search by participant ID, delete participants from the list (Figure 9).
- Click the “DOWNLOAD SELECTED COHORT” button to download a manifest json or csv file for the selected cohort (Figure 9).

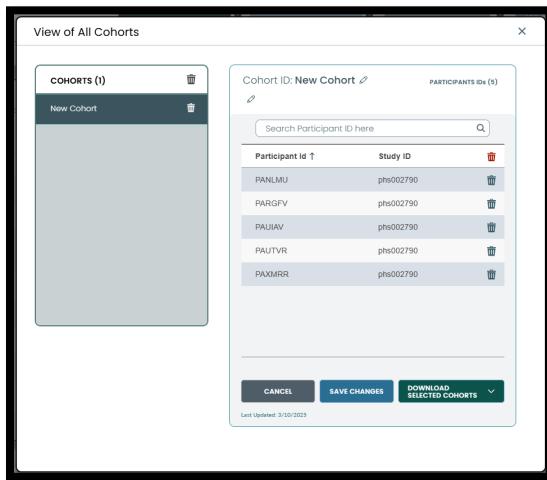


Figure 9: View of All Cohorts

7. Click the X button in the top right to return to the Participant table.

Once a cohort exists, you can easily add more Participants to a cohort by selecting at least one new participant, clicking the “ADD PARTICIPANTS TO EXISTING COHORT” button, and selecting the preferred cohort from the dropdown menu (Figure 10). Clicking the “VIEW ALL COHORTS” button from the Participant table will re-open the “View of All Cohorts” pop-up window described above.



Figure 10: Add Participants to Existing Cohort button

A user can create up to 20 cohorts to exist at any time – cohorts will be stored until a user deletes their browser history. An individual cohort can contain a maximum of 5,000 participants.

Downloading Metadata from the Studies tab

From the CCDI Hub Explore Dashboard, you can download all open metadata for each study from the “Studies” tab to further filter data and build cohorts. For instance, additional filtering by diagnosis of interest can generate a set of participants and the resulting manifest can be uploaded into the CGC. As an example, the following steps guide you on how to download the metadata for the CCDI Molecular Characterization Initiative:

1. Using the process described above, open the “STUDY” set of filters from the lefthand menu, expand the “STUDY NAME” category, and scroll down to find “Molecular Characterization Initiative.”
2. Select the checkbox corresponding to “Molecular Characterization Initiative” and see the Dashboard reload, filtered for this study’s details.
3. Navigate to “Studies” in the results tables and locate the “Manifest” column.

4. Click the “Download study manifest” icon in the “Manifest” column to download the metadata for this study (Figure 11).

Study Name	Study ID	Study Status	Manifest	Diagnosis (Top 5)	Diagnosis Anatomic Site (Top 5)	Number of Participants	Number of Samples
Molecular Characterization Initiative	phs002790	Active		see diagnosis_comment (10837) 9380/3 : Glioma, malignant (517) 9421/1 : Pilocytic astrocytoma (461) 9470/3 : Medulloblastoma, NOS (427) Low-Grade Glioma (337) Read More	C72.9 : Central nervous system (4339) C71.9 : Brain, NOS (3206) Not Reported (2336) C42.0 : Blood (2282) Invalid value (1885) Read More	5489	12295

Figure 11: Download metadata manifest for a given study

5. Open the resulting file on your local machine to browse the resulting metadata tables (Figure 12).

1	type	study.study_id	participant_id	race	sex_at_birth
2	participant	phs002790	PANLMU	White	Female
3	participant	phs002790	PARGFV	Hispanic or Latino;White	Male
4	participant	phs002790	PAUIAV	Unknown	Female
5	participant	phs002790	PAUTVR	Hispanic or Latino;White	Male
6	participant	phs002790	PAXMRR	White	Female
7	participant	phs002790	PAYDFA	White	Female
8	participant	phs002790	PBBHCR	White	Female
9	participant	phs002790	PBBHFF	White	Male
10	participant	phs002790	PBBHGB	Unknown	Male

Figure 12: Study metadata export file browsable on local machine

Appendix A details the process for generating a DRS manifest from the downloaded study metadata tables to be compatible with the CGC.

Creating an Exportable File Manifest from the Cart

In addition to the study-specific downloads, you can also export each row-level metadata element for CCDI participants, diagnoses, samples, or files based on your selections within the Explore Dashboard. Here's how to create a manifest file of filtered information on the Explore Dashboard:

1. On the results tables of the Explore Dashboard, you can select a row of metadata using the checkbox at the start of the row. Multiple rows can be selected within a table, even across pages of the table. Use the checkbox at the top of the checkbox column to select or deselect all rows.

2. After selecting desired rows, add files for that element to the My Files shopping cart (Figure 13) by clicking either the “ADD ALL FILTERED FILES” or “ADD SELECTED FILES” button. Note that selection of items in each tab depends on the specific content of that tab. For example, selecting an item in the “Participants” tab means every file associated with a participant will be added to the My Files shopping cart, whereas selecting an item in the “Files” tab will add that single selected file to the cart.

Participant ID	Study ID	Sex at Birth	Race	Diagnosis	Diagnosis Anatomic Site	Age at Diagnosis (days)	Treatment Type
<input checked="" type="checkbox"/> PANLMU	phs002790	Female	White				Surgical Procedure
<input type="checkbox"/> PARGFV	phs002790	Male	Hispanic or Latino;White	9836/3 : Precursor B-cell lymphoblastic leukaemia	C42.0 : Blood	1033	Chemotherapy;Immunotherapy
<input type="checkbox"/> PALIJAV	phs002790	Female	Unknown				
<input type="checkbox"/> PAUTVR	phs002790	Male	Hispanic or Latino;White	High-Grade Glioma;8910/3 : Embryonal rhabdomyosarcoma;NOS;High-grade glioma, NOS or NEC	C76.5 : Lower limb, NOS	577	
<input type="checkbox"/> PAXMRR	phs002790	Female	White	Other;S538/1 : Clear cell meningioma;Meningioma	C72.0 : Spinal cord	3422	
<input type="checkbox"/> PAYDFA	phs002790	Female	White	Low-Grade Glioma;Pilocytic Astrocytoma;9421/1 : Pilocytic astrocytoma	C71.6 : Cerebellum, NOS	920	
<input type="checkbox"/> PBBHCR	phs002790	Female	White	9501/3 : Medulloepithelioma, NOS	C72.0 : Spinal cord	4503	Radiation Therapy;Surgery
<input type="checkbox"/> PBBHFF	phs002790	Male	White	8000/3 : Neoplasm, malignant	C71.9 : Brain, NOS	2601	Chemotherapy;Immunotherapy
<input type="checkbox"/> PBBHGB	phs002790	Male	Unknown	9441/3 : Giant cell glioblastoma	C71.3 : Parietal lobe	2197	Chemotherapy;Immunotherapy
<input type="checkbox"/> PBBHGD	phs002790	Female	White	9470/3 : Medulloblastoma, NOS	C71.9 : Brain, NOS	4800	Chemotherapy;Immunotherapy;Surgical Procedure

Figure 13: Selection checkboxes and buttons to add files to the cart for the “Participants” table

3. To navigate to the shopping cart, select “MY FILES” or the shopping cart icon on the menu bar (Figure 14).

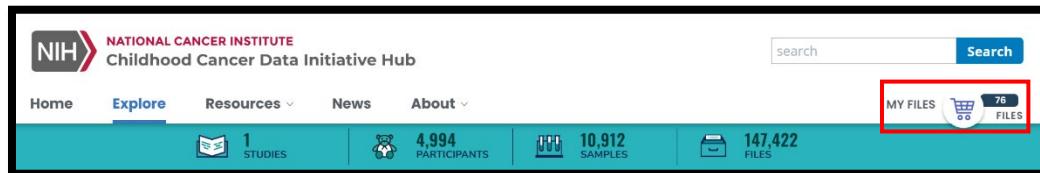


Figure 14: CCDI Hub menu bar with red box highlighting the My Files shopping cart

4. The shopping cart feature enables you to select and manage files. It's a simple way to keep track of data and files during your session. Selecting the “DOWNLOAD MANIFEST” button from the “AVAILABLE EXPORT OPTIONS” dropdown menu (Figure 15) will produce a comma-separated values (CSV) file manifest of the items within the cart.

MY FILES							
<small>To access and analyze files, select or deselect any files as needed, and then choose an option from the 'Available Export Options' dropdown by clicking either 'Export to Cancer Genomics Cloud' or 'Download File Manifest'.</small>							
<small>USER GUIDE</small>							
File Name	Study Name	Study Accession	Participant ID	Sample ID	File Type	File Size	Remove
IGM_PARGFV-0D97VY_20220923.somatic.vcf.gz	Molecular Characterization Initiative	phs002790	PARGFV	0D97VY	vcf	4.14 MB	<input type="button"/>
IGM_PARGFV-0D97VZ_20220923.germline.vcf.gz	Molecular Characterization Initiative	phs002790	PARGFV	0D97VZ	vcf	5.73 MB	<input type="button"/>

Figure 15: The Explore Dashboard Cart page with red box highlighting the “Available Export Options” button

- You can then download this manifest file locally or upload it in the CGC (Appendix C). Similarly, you can instead select the “EXPORT TO CANCER GENOMICS CLOUD” button from the “AVAILABLE EXPORT OPTIONS” dropdown menu to load the resulting manifest directly into your CGC account.

Note that the Cart has a maximum capacity of 200,000 files, which may limit the ability to create very large manifests for use in the CGC. Should you need to create a manifest containing more than 100,000 files, you can either create manifests from the cart in batches (containing up to 100,000 files in each batch) or use the comprehensive metadata downloads from the Explore page “Studies” tab to create a manifest that can take all data for a given study into the CGC. Longer term solutions are being researched.

Accessing Help Documentation

The contents of this document are also available from a Help Browser, accessible from the top of the Explore Dashboard (Figure 16). From this browser, you can scroll or click through the available topics and readily navigate back to the main UI as needed.

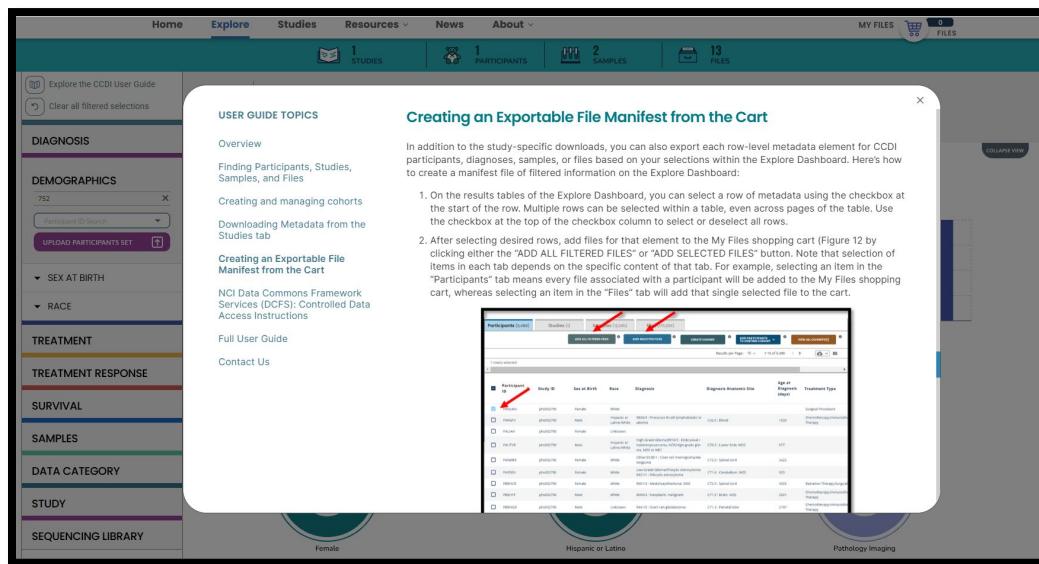


Figure 16: In-page help text browser

Additional search features

For users who prefer text searching, the global search bar at the top of each page allows you enter a query and report all findings in metadata and page text throughout the CCDI Hub. Search results are categorized as Participant, Studies, Samples, Files, Data Model, or About (Figure 17). Where applicable, you can link from these findings back to the Explore Dashboard or other pages with information relevant to your search.

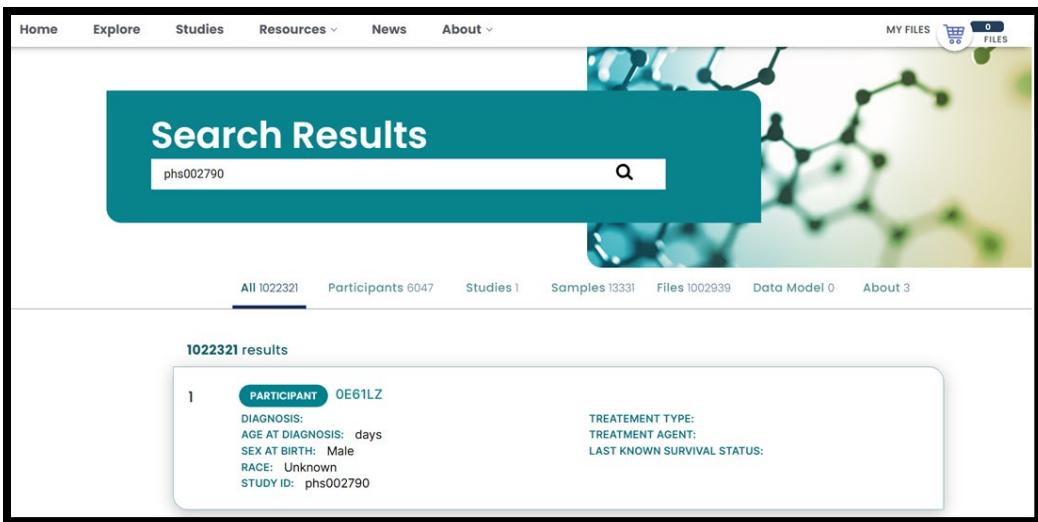


Figure 17: Global Search results

For users interested in the data model, you can browse the model nodes and properties by selecting “CCDI Data Model” from the About dropdown menu.

Contact Information

Please direct any questions or requests for further information to the [CCDI mailbox](#).

THIS PAGE IS INTENTIONALLY LEFT BLANK

Appendix A: database of Genotypes and Phenotypes (dbGaP)

CCDI studies with controlled-access data are registered with the National Center for Biotechnology Information's database of Genotypes and Phenotypes (dbGaP), which maintains a list of the studies' subject IDs, sample IDs, and consents.

Eligible investigators interested in obtaining a controlled data set should watch the [instructional video](#) on applying for controlled access data and consult the [Tips on Preparing a Successful Data Access Request](#). A step-by-step breakdown of the data access request process is located in this [guide](#).

Appendix B: NCI Data Commons Framework Services (DCFS): Controlled Data Access Instructions

NCI Data Commons Framework Services (DCFS), powered by [Gen3](#), facilitates data authorization in a secure and scalable manner. DCFS's Indexd service provides permanent digital IDs for data objects. These IDs can be used to retrieve the data or query the metadata associated with the object.

CCDI data is available for download using the DCFS and bulk downloads are only possible using this method. To gain access to controlled data, researchers must first have an [NIH eRA Commons account](#) for authentication, after which they will need to obtain authorization (via an active DCFS [login account](#)) to access the data in [dbGaP](#).

Below are instructions for using the Data Commons Framework (DCF) user interface or the DCF Gen3-client to access CCDI data.

File Download Procedure via User Interface

To download a study-specific research data distribution file with the DCF Services Portal interface, a researcher must execute the following steps:

1. Please navigate to nci-crcdc.datacommons.io and click the “RAS Login” button. Login to the [NCI DCF Services](#) portal at nci-crcdc.datacommons.io/login (Figure B1).

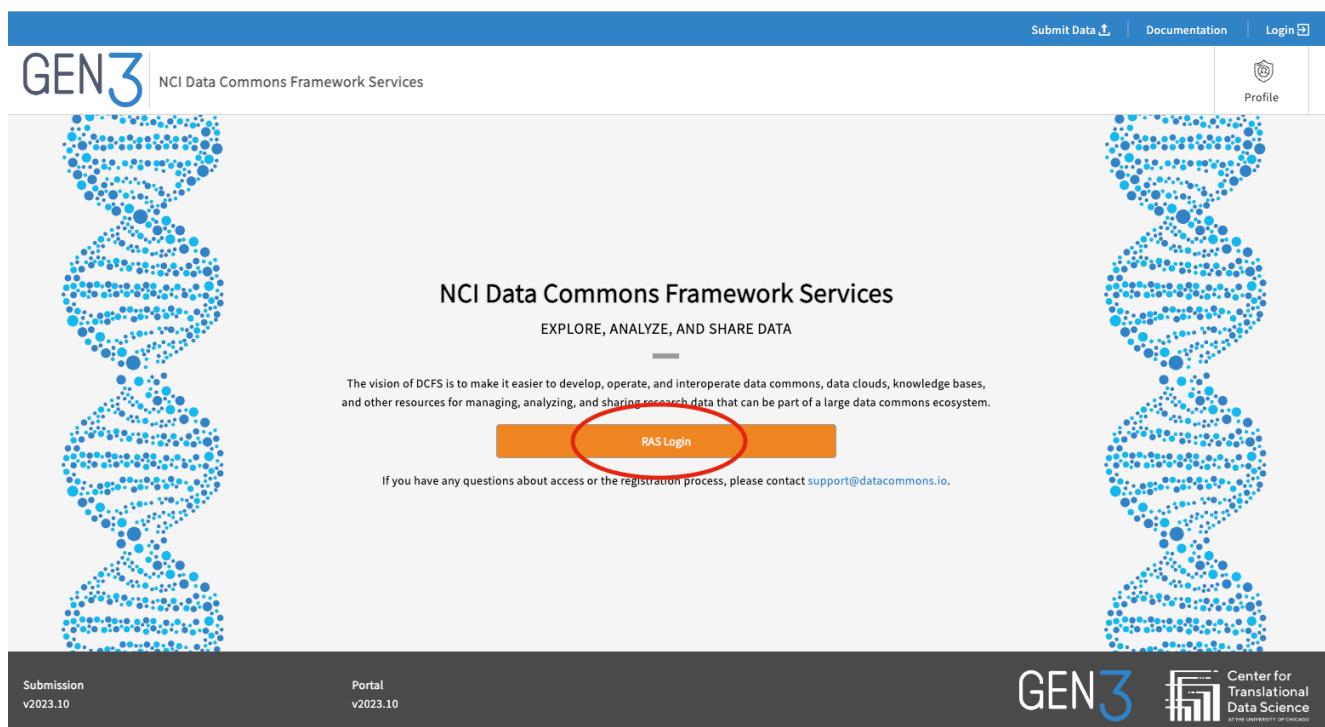


Figure B1: NIH DCF homepage with the NIH Researcher Auth Service (RAS) login highlighted

2. Once logged in, click the “Profile” section in the top right corner and review your project access to confirm study access (Figure B2).

Submit Data Documentation ryuss Logout

GEN3 NCI Data Commons Framework Services

1. click profile

Create API key

You don't have any API key. Please create one!

You have the following API key(s)

API key(s)	Expires

You have access to the following project(s)

Project(s)	Method(s)
phs003291	read-storage, read

Submission v2023.10 Portal v2023.10

GEN3 Center for Translational Data Science AT THE UNIVERSITY OF CHICAGO

Figure B2: DCF Profile page highlighting “Profile” and the accessible projects.

3. Add the file GUID after the final backslash in this URL: <https://nci-crdc.datacommons.io/user/data/download/>. Paste the URL you created in a browser address field and press Enter or Return.
4. The NCI DCF Services Portal will respond by providing a JSON document with a new (signed) URL for the requested data file. Copy the signed URL.
5. Paste this new signed URL into the browser address field and press Enter or Return (Figure B3).
 - a. Note: Once issued, the signed URL provided is valid for a relatively short period of time.
6. The NCI DCF Services Portal will respond by displaying a URL. Click the URL to download the file (Figure B3).

```

{
  "url": "https://cds-305-phs002790-7353b646a8979b9e3b9554d35c014cc924ed6d7601bb7dfb"
}

```

Figure B3: DCF Service Portal displaying the signed access URL and the file download URL

Note: If errors or problems are experienced during the file downloading process above, please contact the [CCDI mailbox](#) for assistance.

File Download Procedure via Command Line Interface (CLI) client

To download a study-specific research data distribution file with a CLI client, a researcher must execute the following steps:

1. Obtain the [Gen3-client command-line tool](#) from GitHub.
2. Install and configure the client based on the [Gen3 instructions](#).
 - a. These instructions include signing into the DCF web client and obtaining a downloaded JSON API key from the Profile page, and then configuring the client.
 - b. The API endpoint that will be used for DCF configuration is ‘<https://nci-crdc.datacommons.io>’.
3. Obtain either a GUID or manifest of GUIDs for the data files of interest from the [CCDI Explore page](#) or the [Explore Dashboard exportable manifest](#).
4. Create a Gen3 structured manifest:

```
[  
  {  
    "object_id": "dg.4DFC/{guid_1}"  
  },  
  {  
    "object_id": "dg.4DFC/{guid_2}"  
  },  
  ...  
  {  
    "object_id": "dg.4DFC/{guid_n}"  
  }  
]
```

5. Download the file(s) using the Gen3 client (either the [single](#) or [multiple](#) download option).

For more information on this process, please visit the [Gen3 documentation page](#).

Appendix C: The Cancer Genomics Cloud (CGC)

The Seven Bridges Cancer Genomics Cloud (CGC), powered by Velsera and funded by NCI, is a flexible cloud platform that enables analysis, storage, and computation of large cancer data sets. The CGC provides a user-friendly portal to access and analyze cancer data without having to learn how to program.

Creating a CGC Account

You need an account to access and analyze CCDI data on the CGC platform. Note that a [list of all CCDI studies](#) released is also available. The following instructions describe the process to create a CGC account.

1. From the CGC home page, cancergenomicscloud.org, click “REGISTER” or “LAUNCH” in the center of the page (Figure C1).

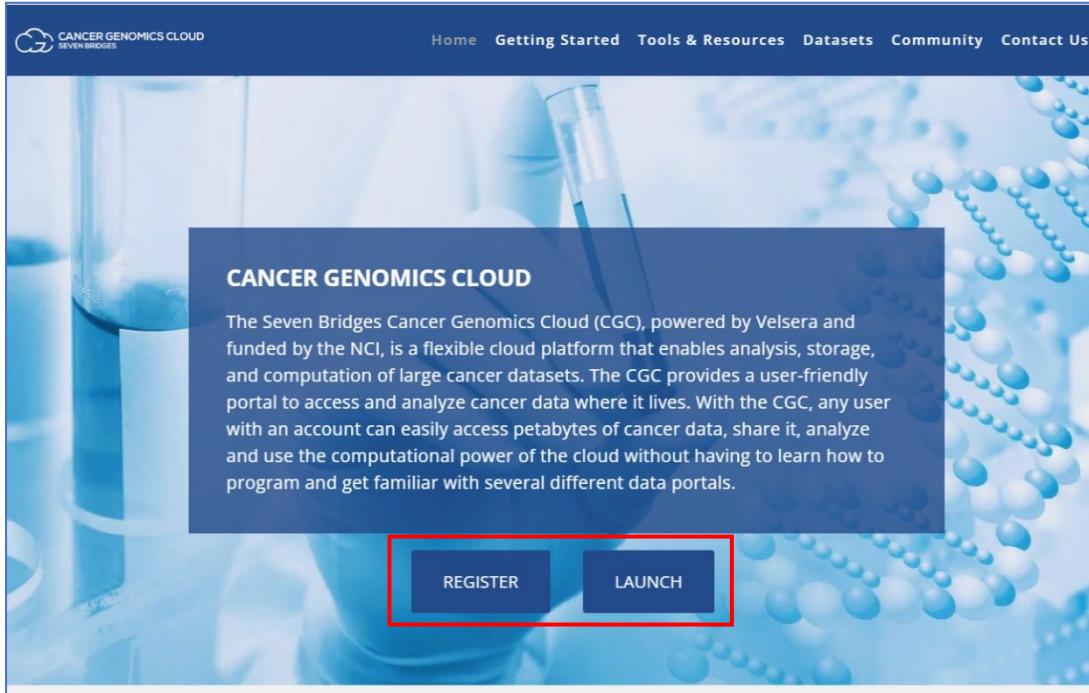


Figure C1: CGC home page with “REGISTER” and “LAUNCH” buttons highlighted in a red box.

2. On the login screen, click on “Log in with eRA Commons” (Figure C2).

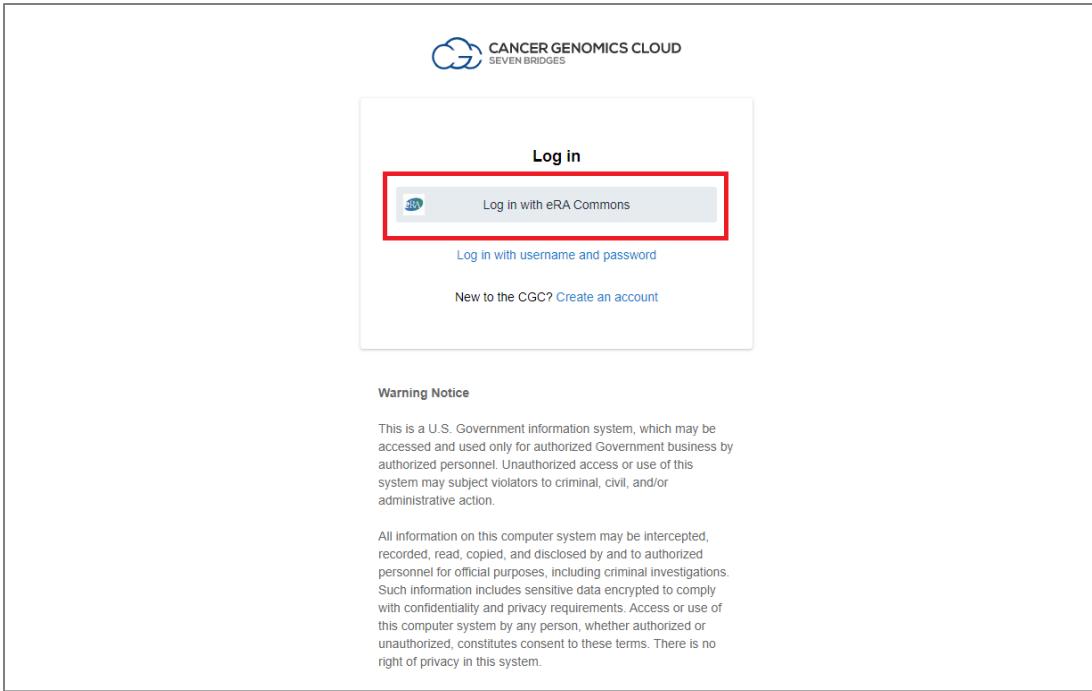


Figure C2: eRA Commons login screen for the CGC, with login button highlighted in a red box

3. On the login screen, enter the eRA Commons account credentials associated with your approved dbGaP study and click “Sign In” (Figure C3).

Please note that if you receive an error message when logging in here, you can confirm that your eRA Commons username and password are correct by logging in to the [eRA Commons site](#).

For additional details about ensuring your Velsara eRA Commons account is correct and correctly configuring the download restriction and network access toggle at the project level, please refer to the following documentation:

- docs.cancergenomicscloud.org/docs/modify-project-settings
- docs.cancergenomicscloud.org/docs/define-download-restriction

If you've previously logged into the eRA Commons site, you may need to clear your web browser's cache or use "incognito" mode to ignore cached data and cookies so you can enter and test your credentials. If you receive an error message on that site as well, you may need to reset your eRA Commons password.

Figure C3: Login page for CGC, with username and password credentials sections highlighted in a red box.

4. If you agree to the “Consent to Share Information” on the following page, click the “Grant” button to continue (Figure C4).

Figure C4: Consent to share information page with red box highlighting the “Grant” button to confirm consent to share information with the CGC

5. If you agree to authorize Gen3 DCFS to share your account and authorization information to access the data sets for which you have been approved, click the “Yes, I authorize” button (Figure C5). Note

that Gen3 is an authorization system that uses eRA Commons as an authentication tool and allows access to the CGC system.

The screenshot shows the Gen3 Data Commons Framework Services interface. At the top is the Gen3 logo with the text "Data Commons". Below it is a blue header bar with the text "Data Commons Framework Services". The main content area has a title "Authorize SBG-20201217 to:" followed by a list of permissions:

- Know your Google and NIH and RAS basic account information and what you are authorized to access.
- Retrieve controlled-access datasets to which you have access on your behalf.
- Retrieve GA4GH Passports and Visas

At the bottom right of the main area are two buttons: "Cancel" and "Yes, I authorize.", with "Yes, I authorize." highlighted by a red box. Below the main area is a link "Gen3 Privacy Policy".

Figure C5: Gen3 DCFS authorization page with “Yes, I authorize” button highlighted in a red box

6. On the next page, confirm that the information listed for you is correct (if this page appears). If you agree to the Terms of Service, Data Use, and Privacy policies, click the two related checkboxes, and then click on “Proceed to the CGC” (Figure C6).

The screenshot shows the Cancer Genomics Cloud (CGC) registration confirmation page. At the top is the CGC logo with the text "CANCER GENOMICS CLOUD" and "SEVEN BRIDGES". The main content area has a heading "We just need a couple more pieces of information from you, then you're all set." followed by several input fields:

- First name / Last name
- Email (example: username@yourdomain.com)
- CGC Username (example: CGCUsername)
- Your Organization Name
- Location (example: United States)

Below these fields is a red box containing two checkboxes:

- I have read and agreed to the [Terms of Service](#) and all relevant [Data Use](#) policies.
- I have read and agree to the [Privacy Policy](#) and the processing of my personal data as described therein.

At the bottom is a blue "Proceed to the CGC" button.

Figure C6: Confirmation of terms and policy for CGC registration highlighted in a red box

- If the CGC questionnaire appears, complete it to continue (Figure C7).

Figure C7: Screenshot of the CGC questionnaire

Creating a DRS Manifest from downloaded Study metadata

The Study Manifest file contains all metadata about a study available in the CCDI Hub. This file can be used to create a Data Repository Service (DRS) file that includes identifiers and metadata that allows users to locate and access specific files or data sets within a data repository for use on the CGC. To convert the Study Manifest file into a DRS manifest, please follow these steps:

- Go to ccdi.cancer.gov/explore
- Select "Molecular Characterization Initiative" under Study Name, or choose "phs002790" under Study dbGaP Accession.
- Download the Study Manifest (contains metadata for the study).
- Find the tabs ending with "_file" (look for FASTQ, JSON, etc.).
- Create a DRS Manifest for Cloud Analysis
 - Extract and rename the column headers: file_name (name); dcf_indexd_guid (drs_uri); study_id (study_code); sample.sample_id/participant_id (case_id).
- Add drs://nci-crdc.datacommons.io/ to all the beginning of all rows in the drs_uri column. Go to <https://www.cancergenomicscloud.org/>:
- In CGC, go to your Project. Under Files, find Add Files and GA4GH Data Repository Service (DRS).
- Upload your manifest.
- On the next page, select the "From a manifest file" tab and drag and drop your manifest onto the page. Go through the next steps agreeing to the use of the system and ingest the manifest.
- After being redirected back to the file page for the project, it will begin to load the files into the Files section of the CGC project.

Downloading files from the CGC

Data exported to CGC from CCDI Hub Explore or produced from either CGC Public Apps or from data analysis in the CGC Data Studio will be output to Files tab. *Please note, if the user intent is to download large and/or many files indexed in CCDI Hub Explore, it is highly recommended that users utilize the Gen3 client with file DCF GUIDs as outlined in [Appendix B: NCI Data Commons Framework Services \(DCFS\): Controlled Data Access Instructions](#).*

There are several methods to download files from the CGC:

1. CGC Web Interface
2. Seven Bridges Command Line Interface (CLI)
 - a. Recommended method for downloading large and/or many files
3. Using Download Links

1. Download Files Using the CGC Web Interface

1. Navigate to the Project: Select the project (workspace) where your data is stored.

The screenshot shows the CGC Web Interface homepage. At the top, there is a navigation bar with tabs: Home, Projects (which is highlighted with a red box), Data, Public Apps, Public Projects, and Developer. Below the navigation bar is a search bar labeled "Search projects..." with a magnifying glass icon. To the right of the search bar are filters for Type, Category, Toolkit, and Language. The main content area displays a list of projects. One project, "test dl", is listed under "JSON2TSV COG IGM". Another project, "CCDI MCI JSON2TSV COMMIT", is listed under "CCDI MCI JSON2TSV". On the right side, a detailed view of a pipeline is shown. The pipeline is titled "CWL Single Cell Multi Sample Pairwise...". It has a brief description: "This pipeline performs differential expression analysis on single cell data between pairs of user defined conditions. ...". Below the description are two buttons: "Copy" and "Run". At the bottom left of the main content area, there are buttons for "View all projects" and "+ Create a project".

2. Open the Files Tab: Click on the “Files” tab within the project workspace.
3. Locate Your File(s): Browse or search for the specific files or folders you want to download.

Dashboard **Files** Apps Tasks Data Studio Interactive Apps **CONTROLED JSON2TSV COG IGM** Interactive Browsers Settings

Root **New Folder** **Add file**

Search Extension Tags Clear filters

<input type="checkbox"/>	Name	Size	Extension	Task ID	Created on
<input type="checkbox"/>	DRS PBLPL_archer_fusion.clinical_report.json	472 bytes	JSON	-	May. 16, 2025
<input type="checkbox"/>	DRS PBBKNA_archer_fusion.clinical_report.json	482 bytes	JSON	-	May. 16, 2025
<input type="checkbox"/>	DRS PBBHWP_archer_fusion.clinical_report.json	480 bytes	JSON	-	May. 16, 2025
<input type="checkbox"/>	DRS PBBVYL_archer_fusion.clinical_report.json	479 bytes	JSON	-	May. 16, 2025
<input type="checkbox"/>	DRS archer_fusion_IGM_PBCBSV-0DR0C2_20231016.clinical_report.json	484 bytes	JSON	-	May. 16, 2025
<input type="checkbox"/>	DRS PBCAXG_archer_fusion.clinical_report.json	484 bytes	JSON	-	May. 16, 2025
<input type="checkbox"/>	DRS PBBJJS_archer_fusion.clinical_report.json	474 bytes	JSON	-	May. 16, 2025
<input type="checkbox"/>	DRS archer_fusion_IGM_PBCEHX-0DPYDY_20230918.clinical_report.json	479 bytes	JSON	-	May. 16, 2025
<input type="checkbox"/>	DRS archer_fusion_IGM_PBCCZB-0DQ7WK_20230919.clinical_report.json	484 bytes	JSON	-	May. 16, 2025
<input type="checkbox"/>	DRS PBBISG_archer_fusion.clinical_report.json	462 bytes	JSON	-	May. 16, 2025

4. Download Files:

- o Select one or more files using the checkboxes.
- o Click on the Download icon (cloud with a downward arrow).
- o If prompted, confirm the download location or process.

Dashboard **Files** Apps Tasks Data Studio Interactive Apps **CONTROLED JSON2TSV COG IGM** Interactive Browsers

Root **New Folder**

5 items selected **Copy** **Rename** **Move** **Metadata** **Edit tags** **Download** **...**

<input type="checkbox"/>	Name	Size	Extension	Task ID
<input checked="" type="checkbox"/>	30463dee-4ecb-9907-86d646fb8f55.csv	1.63 kB	CSV	-
<input checked="" type="checkbox"/>	DRS archer_fusion_IGM_PBCBSV-0DR0C2_20231016.clinical_report.json	484 bytes	JSON	-
<input checked="" type="checkbox"/>	DRS archer_fusion_IGM_PBCCZB-0DQ7WK_20230919.clinical_report.json	484 bytes	JSON	-
<input checked="" type="checkbox"/>	DRS archer_fusion_IGM_PBCEHX-0DPYDY_20230918.clinical_report.json	479 bytes	JSON	-
<input checked="" type="checkbox"/>	DRS PBBHWP_archer_fusion.clinical_report.json	480 bytes	JSON	-
<input type="checkbox"/>	DRS PBBISG_archer_fusion.clinical_report.json	462 bytes	JSON	-
<input type="checkbox"/>	DRS PBBJJS_archer_fusion.clinical_report.json	474 bytes	JSON	-
<input type="checkbox"/>	DRS PBBKNA_archer_fusion.clinical_report.json	482 bytes	JSON	-
<input type="checkbox"/>	DRS PBLPL_archer_fusion.clinical_report.json	472 bytes	JSON	-
<input type="checkbox"/>	DRS PBBVYL_archer_fusion.clinical_report.json	479 bytes	JSON	-

Note: The CGC has limitations on the number of files that can be downloaded at once.

2. Download Files Using Seven Bridges CLI

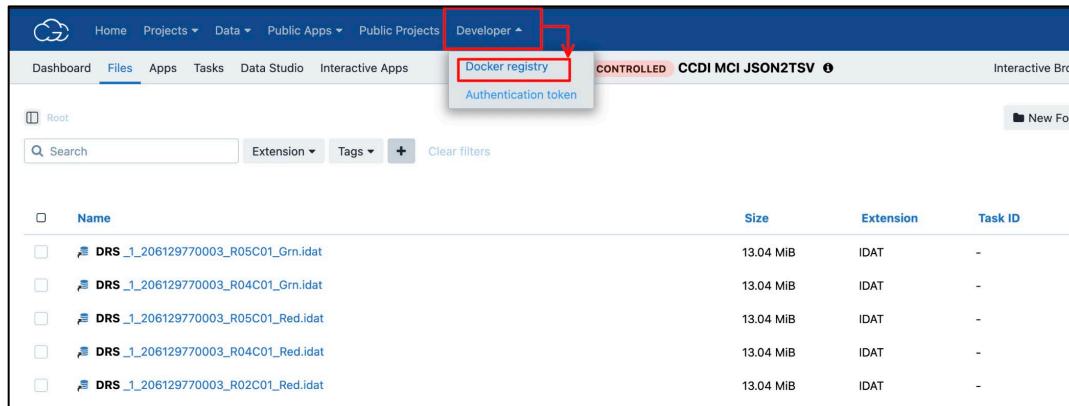
This method is suitable for downloading large datasets or automating workflows.

Prerequisites:

- Install the Seven Bridges CLI
 - Installation Guide: <https://docs.cancergenomicscloud.org/docs/command-line-interface>
- Authenticate the CLI with your CGC account
 - Enter following command in terminal: `sb configure`
 - User will then be prompted to enter in the following information:

```
USER$ sb configure
Seven Bridges API endpoint [None]:
Seven Bridges authorization token [None]:
Seven Bridges advance access (true/false) [None]: false
```

- API Endpoint
- Authorization token
- Advance access True/False: Enter `false` unless advance access applies to your endpoint
- The API Endpoint and Authorization token can be found from the Developer > Authentication token page in the CGC Dashboard:



The screenshot shows the Seven Bridges Cloud Platform interface. At the top, there's a navigation bar with links for Home, Projects, Data, Public Apps, Public Projects, and a 'Developer' dropdown. The 'Developer' dropdown is expanded, showing 'Docker registry' and 'Authentication token'. Below the navigation is a search bar and filter options for 'Search', 'Extension', 'Tags', and 'Clear filters'. The main area displays a table of files with columns for Name, Size, Extension, and Task ID. The table contains five entries, all of which are 13.04 MiB in size, have an IDAT extension, and no Task ID assigned. The file names follow a pattern: DRS_1_206129770003_R05C01_Grn.idat, DRS_1_206129770003_R04C01_Grn.idat, DRS_1_206129770003_R05C01_Red.idat, DRS_1_206129770003_R04C01_Red.idat, and DRS_1_206129770003_R02C01_Red.idat.

Name	Size	Extension	Task ID
DRS_1_206129770003_R05C01_Grn.idat	13.04 MiB	IDAT	-
DRS_1_206129770003_R04C01_Grn.idat	13.04 MiB	IDAT	-
DRS_1_206129770003_R05C01_Red.idat	13.04 MiB	IDAT	-
DRS_1_206129770003_R04C01_Red.idat	13.04 MiB	IDAT	-
DRS_1_206129770003_R02C01_Red.idat	13.04 MiB	IDAT	-

- And then copied from Authentication token page:

The screenshot shows the 'Developer' section of the CGC platform. In the 'Authentication Token' box, a token is displayed with a red box around it. Below it, the expiration date is Aug. 13, 2025, and the last used date is Jun. 9, 2025. In the 'API' box, the 'API base endpoint' is listed as <https://cgc-api.sbggenomics.com>, which is also highlighted with a red box.

Example Commands

For following commands, replace project-name with actual name of your project, can be found from project URL in browser as “username/project_name”. For example:

The screenshot shows a browser window with the URL <https://cgc.sbggenomics.com/u/user12345/json2tsv-cog-igm>. A red box highlights the URL in the address bar. The page title is "JSON2TSV COG IGM".

In example above, project name is “user12345/json2tsv-cog-igm”

List files in a project

```
sb files list --project project-name
```

Lists file_id, file_name and project location, can pipe to a file to see all files and ids in one location:

```
sb files list --project project-name > project_files.txt
```

Use returned file_ids to download specific files with CLI.

Download a file

```
sb download --file <file_id_value> ... [--destination <destination_value>] [--overwrite] [flags]
```

Download multiple files

```
sb download --file <file_id_1> --file <file_id_2> [--overwrite]
```

Display help information

`sb help` or `sb help [command]`, e.g. `sb help files` for seeing help menu for sb files command

For more information, please see the following relevant links at the CGC Documentation site:

<https://docs.cancergenomicscloud.org/docs/command-line-interface>

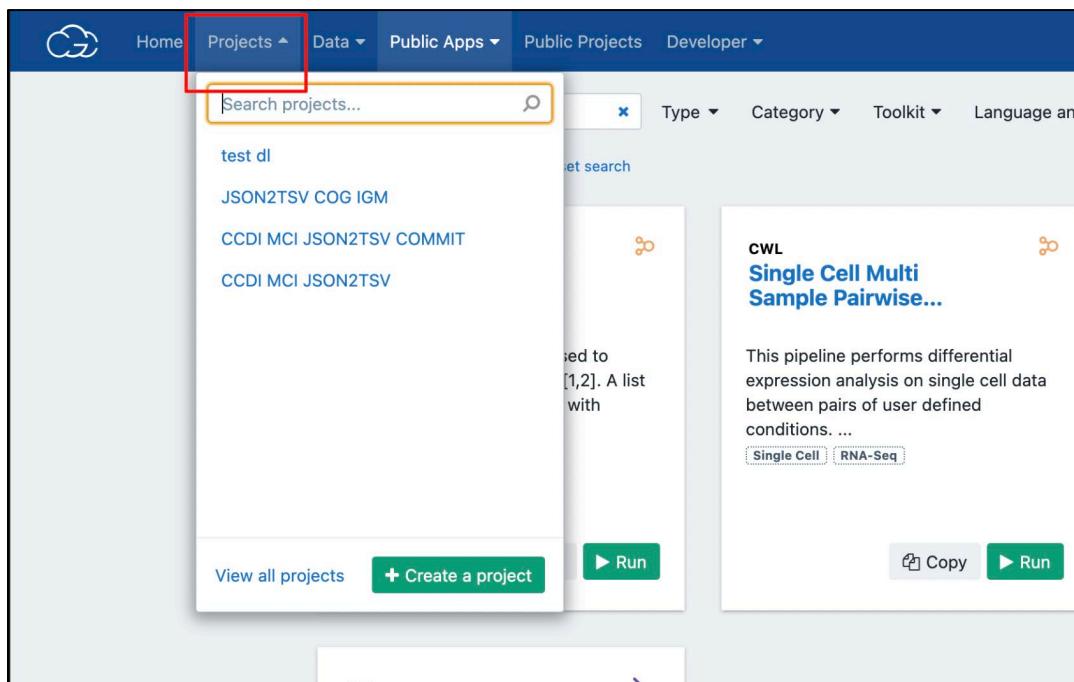
<https://docs.cancergenomicscloud.org/docs/upload-and-download-files>

3. Download Files Using CGC Generated Download Links

Generated Download Links enable users to download files using curl, wget, [aria2](#) or other preferred third-party download tools. **Please note:** there is a limit to the number of download links that can be generated at given time for file downloads, this method is **not** recommended for bulk download of files from the CGC. For more information, please see CGC Documentation site regarding Download Links:

<https://docs.cancergenomicscloud.org/docs/download-results#download-files-using-direct-download-links>

1. Navigate to the Project: Select the project (workspace) where your data is stored.



2. Open the Files Tab: Click on the “Files” tab within the project workspace.
3. Locate Your File(s): Browse or search for the specific files or folders you want to download.

The screenshot shows the Qeazy Cloud interface with the 'Files' tab selected. At the top, there's a banner indicating the folder is 'CONTROLLED' and contains 'JSON2TSV COG IGM'. Below the banner, there's a search bar, filter buttons for 'Extension' and 'Tags', and a 'Clear filters' button. A 'New Folder' button and an '+ Add file' button are also visible. The main area displays a table of files under the 'Root' folder. The columns are 'Name', 'Size', 'Extension', 'Task ID', and 'Created on'. The files listed are all clinical reports in JSON format, mostly from May 16, 2023.

Name	Size	Extension	Task ID	Created on
DRS PBLPL_archer_fusion.clinical_report.json	472 bytes	JSON	-	May. 16, 2023
DRS PBBKNA_archer_fusion.clinical_report.json	482 bytes	JSON	-	May. 16, 2023
DRS PBBHWP_archer_fusion.clinical_report.json	480 bytes	JSON	-	May. 16, 2023
DRS PBBVYL_archer_fusion.clinical_report.json	479 bytes	JSON	-	May. 16, 2023
DRS archer_fusion_IGM_PBCBSV-0DROC2_20231016.clinical_report.json	484 bytes	JSON	-	May. 16, 2023
DRS PBCAXG_archer_fusion.clinical_report.json	484 bytes	JSON	-	May. 16, 2023
DRS PBBJJS_archer_fusion.clinical_report.json	474 bytes	JSON	-	May. 16, 2023
DRS archer_fusion_IGM_PBCEHX-0DPYDY_20230918.clinical_report.json	479 bytes	JSON	-	May. 16, 2023
DRS archer_fusion_IGM_PBCCZB-0DQ7WK_20230919.clinical_report.json	484 bytes	JSON	-	May. 16, 2023
DRS PBBISG_archer_fusion.clinical_report.json	462 bytes	JSON	-	May. 16, 2023

- Select the files you would like to download. **Note:** There is a limit to the number of files that download links can be generated for at a given time, ~50 files.

The screenshot shows the Qeazy Cloud interface with the 'Files' tab selected. At the top, there's a banner indicating the folder is 'CONTROLLED' and contains 'JSON2TSV COG IGM'. Below the banner, there's a search bar, filter buttons for 'Copy', 'Rename', 'Move', 'Metadata', 'Edit tags', 'Download', and an ellipsis button. A 'New Folder' button and an '+ Add file' button are also visible. The main area displays a table of files under the 'Root' folder. The files listed are all clinical reports in JSON format, mostly from May 16, 2023. The first ten files have been selected, as indicated by checked checkboxes in the 'Selected' column. A new file, '30463dee-6ecb-4bed-9907-86d646fb8f55.csv', has been added at the bottom, which is not selected.

Name	Size	Extension	Task ID	Created on
DRS PBCAXG_archer_fusion.clinical_report.json	484 bytes	JSON	-	May. 16
DRS PBBVYL_archer_fusion.clinical_report.json	479 bytes	JSON	-	May. 16
DRS PBLPL_archer_fusion.clinical_report.json	472 bytes	JSON	-	May. 16
DRS PBBKNA_archer_fusion.clinical_report.json	482 bytes	JSON	-	May. 16
DRS PBBJJS_archer_fusion.clinical_report.json	474 bytes	JSON	-	May. 16
DRS PBBISG_archer_fusion.clinical_report.json	462 bytes	JSON	-	May. 16
DRS PBBHWP_archer_fusion.clinical_report.json	480 bytes	JSON	-	May. 16
DRS archer_fusion_IGM_PBCEHX-0DPYDY_20230918.clinical_report.json	479 bytes	JSON	-	May. 16
DRS archer_fusion_IGM_PBCCZB-0DQ7WK_20230919.clinical_report.json	484 bytes	JSON	-	May. 16
DRS archer_fusion_IGM_PBCBSV-0DROC2_20231016.clinical_report.json	484 bytes	JSON	-	May. 16
30463dee-6ecb-4bed-9907-86d646fb8f55.csv	1.63 KiB	CSV	-	May. 16

- After files have been selected, select the ellipses button next to the "Download" button, and then from drop down menu select "Get Download Links"

6. A popup with download links for selection of files is displayed; download links can be then copied from popup by clicking copy button or downloaded to a TXT file by clicking green “Download link” button

Please note: download links provide access to controlled access files and should not be shared with outside parties or unauthorized individuals. Download links expire 48 hours after generation.

7. Users can then download files with download links using preferred third party download app.

Download individual files with download link, example with curl:

```
curl -O "https://example-download-link-url"
```

Since output is streamed, user should specify **-O** flag ahead of link so that file download is saved to file with same name as source file. Depending on terminal settings, you may need to place parentheses around file download link to perform special character escaping.

Additional Resources on Working with Data at the CGC

- CGC Documentation: docs.cancergenomicscloud.org/docs
- Importing General Commons Data: <https://docs.cancergenomicscloud.org/docs/import-data-from-the-gc>
- Common Workflow Language Workflows and Apps: cgc.sbggenomics.com/public/apps
- Volumes: docs.cancergenomicscloud.org/docs/volumes-1
- Tool Editor Tutorial: docs.cancergenomicscloud.org/docs/tool-editor-tutorial
- About the Editor: docs.cancergenomicscloud.org/docs/about-the-editor