



# CHILDHOOD CANCER CLINICAL DATA COMMONS (C3DC) User Guide

11/07/2024

Version	Date	Description	Author
1.0.0	03/05/2024	Initial Release	C3DC Team
2.0.0	05/01/2024	Release 2	C3DC Team
3.0.0	07/31/2024	Release 3	C3DC Team
4.0.0	11/07/2024	Release 4	C3DC Team

## Table of Contents

<b>INTRODUCTION AND OVERVIEW</b>	<b>4</b>
<b>C3DC HOME PAGE</b>	<b>5</b>
FACET SEARCH	7
CHART VISUALIZATION	7
TABLE VISUALIZATION	8
DOWNLOAD HARMONIZED DATA	8
COHORT SELECTOR	9
<b>C3DC STUDIES PAGE</b>	<b>11</b>
C3DC STUDIES DETAILS PAGE	11
<b>C3DC RESOURCE PAGE</b>	<b>13</b>
<b>C3DC ABOUT PAGE</b>	<b>14</b>
ANNOUNCEMENTS	15
RELEASE NOTE	16
USER GUIDE	17

## Introduction and Overview

**The Childhood Cancer Clinical Data Commons (C3DC)** enables searching demographic and phenotypic clinical data of childhood cancers. These data have been harmonized to a standard set of common data elements (CDEs). C3DC empowers researchers to search for participant-level data to create synthetic cohorts and export data for analysis.

This document describes a high-level overview of the features of C3DC. Investigators are encouraged to explore C3DC themselves, using this guide as a primer.

## C3DC Home Page

The C3DC Home page allows users to navigate pages such as Explore and About.



Figure 1: C3DC Home Page - top

At the bottom of the Home page, there are links to brief descriptions of key sections of the C3DC application.

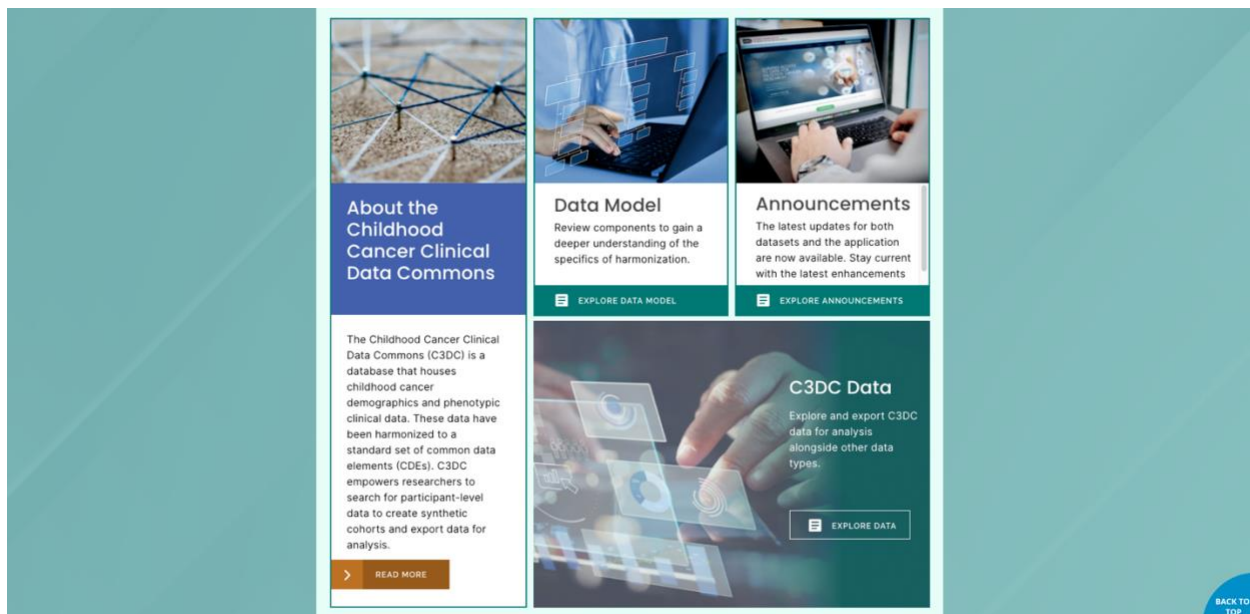


Figure 2: C3DC Home Page - bottom

## C3DC Explore Page

Users can navigate to the Explore Page by clicking the link on the Home Page menu bar, which shows various sections for exploring the data.

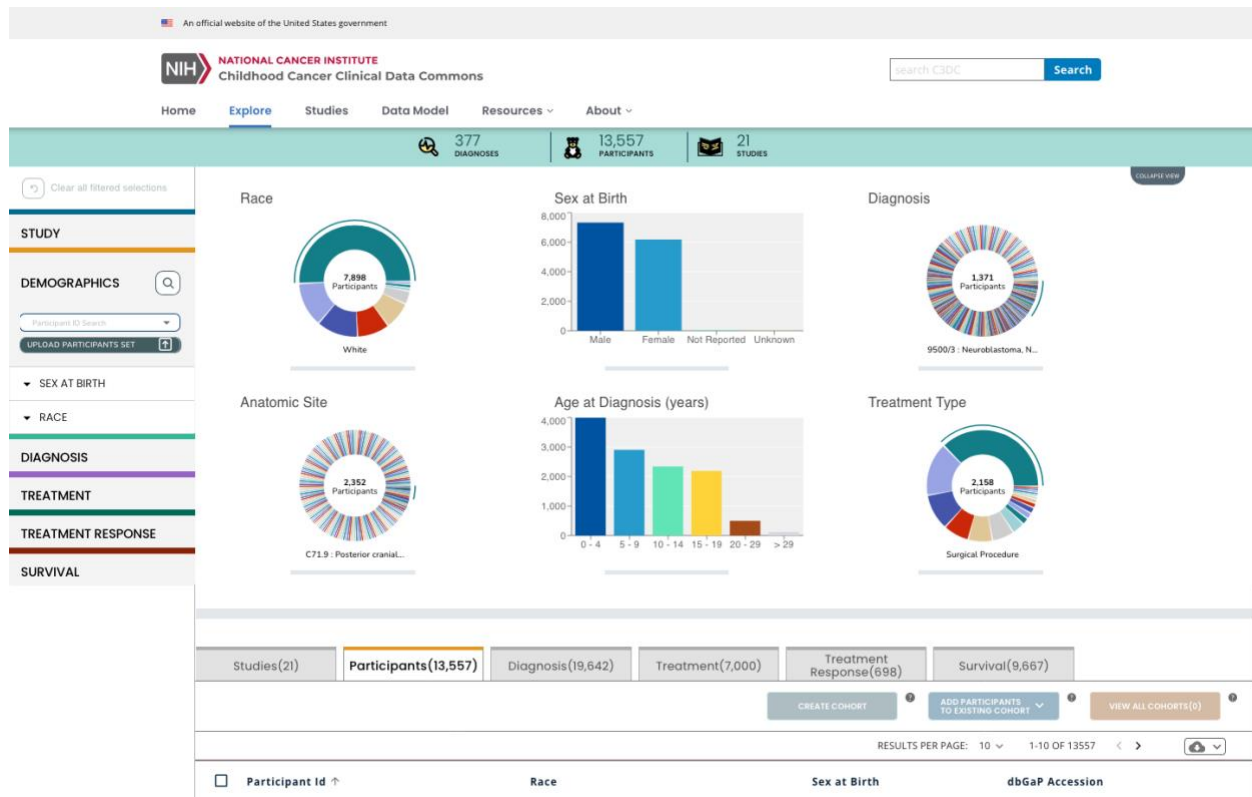


Figure 3: The Explore Page – Faceted Search (left), Visualization (top), and Data Tables (bottom)

**Facet Search:** Users can select a subset of the childhood cancer participants by choosing filtering options from drop-down lists within four main categories: Study, Demographics, Diagnoses, Treatment, Treatment Response, and Survival.

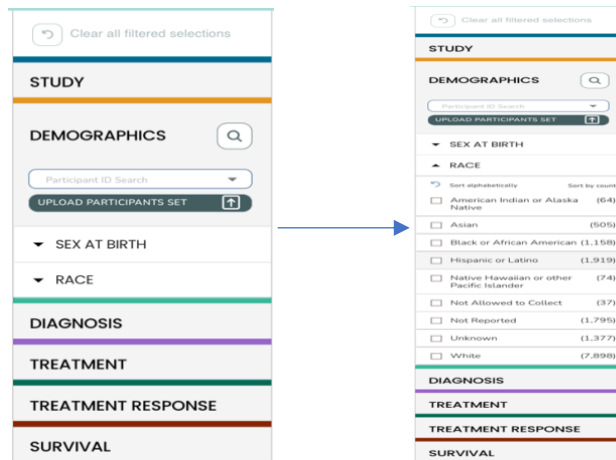


Figure 4: Performing a faceted search on the C3DC Explore Page

**Chart Visualization:** The faceted search results are automatically updated in the Stats Bar, the Visualization, and the Table sections.

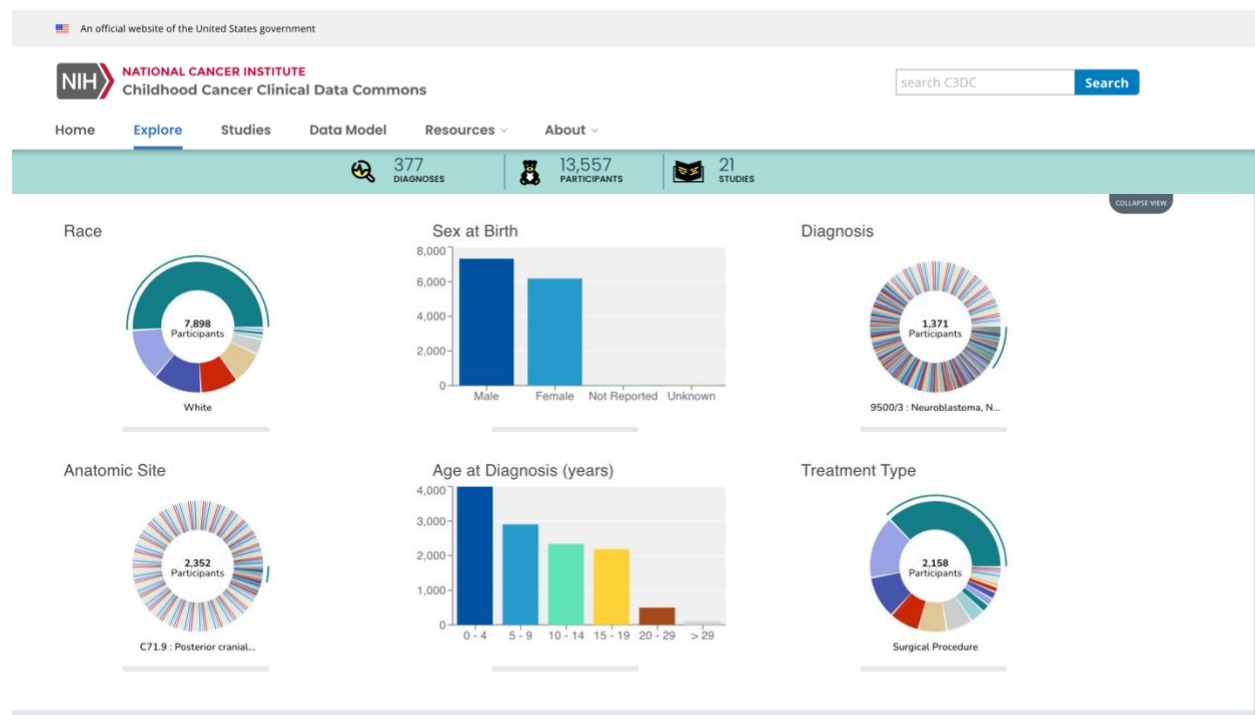


Figure 5: Results returned from a faceted search in six graphs in the Visualization section

Count of Participant Record

Studies(21)

**Participants(13,557)**

Diagnosis(19,642)

Treatment(7,000)

Treatment Response(698)

Survival(9,667)

CREATE COHORT

ADD PARTICIPANTS TO EXISTING COHORT

VIEW ALL COHORTS(3)

RESULTS PER PAGE: 10 1-10 OF 13557

<input type="checkbox"/>	Participant Id ↑	Race	Sex at Birth	dbGaP Accession
<input type="checkbox"/>	00301d78915737fa100f	White	Female	phs002431
<input type="checkbox"/>	0061cbb0846973206cfc	White	Male	phs002431
<input type="checkbox"/>	0065af91e89ee2859595	Hispanic or Latino;White	Female	phs002431
<input type="checkbox"/>	008b04a84717e7d007a4	White	Male	phs002431
<input type="checkbox"/>	009f9efaedc9602a28fb	White	Female	phs002431
<input type="checkbox"/>	00c5e4372375eb3627f3	White	Female	phs002431
<input type="checkbox"/>	00dfd38d509984eada22	White	Male	phs002431
<input type="checkbox"/>	00e3d1d383c8c08a25d2	White	Male	phs002431
<input type="checkbox"/>	00eff50ac98a200da9f9	White	Male	phs002431
<input type="checkbox"/>	01504e3b7031a9eec2b2	White	Female	phs002431

RESULTS PER PAGE: 10 1-10 OF 13557

**Download Harmonized Data:** Users can download the contents of the Participants, Diagnosis, Survival, and Studies tabs by selecting the "Download Data" button under the Table tab headers. The data will be downloaded in CSV file format.

A1	Participant ID																						
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R					
1	Participant ID	Diagnosis ID	Diagnosis	Diagnosis C1	Diagnosis B	Diagnosis C	Disease Path	Tumor	Class	Anatomic SI	Age at Dx	Jiang	Toronto	Chikl	Tumor Grade	Tumor Stage	Tumor Stage	Study ID	Dogba Accession				
2	PCHD0L	02296835-611-4787-49fa-1f105b3977bc	9421/1: Piloc: ICD-O-3.2	Clinical	Not Reported	Initial Diagnostic	Primary			C71.6: Cereb	2005	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs022790	phs022790				
3	P17A22ASA	0743006cc6a11-6b22-4c365f58be43221.1	High-grade GNS CNDiagnosis	Pathological	High-grade g1 Initial Diagnostic	Primary				C71.6: Cereb	2712	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs020517	phs020517				
4	P1CT1H0GJ	0be0c38e16335f5853b05a36288c65b3.1	9400/0: Astro ICD-O-3.2	Pathological	Low-grade g1 Initial Diagnostic	Primary				C71.6: Cereb	1546	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs020517	phs020517				
5	P_VVV047R	0e95c55474d5e5508a344efc825e7.1	Medulloblast: CNS Diagnosis	Pathological	Medulloblast Initial Diagnostic	Primary				C71.6: Cereb	2212	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs020517	phs020517				
6	P1_DZ8EYNG	0e90db0f3c7b0c003dac0b69ca87513.1	9400/0: Astro ICD-O-3.2	Pathological	Low-grade g1 Initial Diagnostic	Primary				C71.6: Cereb	1805	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs020517	phs020517				
7	P1_GBN25651	00f4d0e3013b287ca7669887879b026.5	9500/3: Atypyl ICD-O-3.2	Pathological	Atypical Tera Post-Mortem	Primary				C71.6: Cereb	3562	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs020517	phs020517				
8	P1_G62TCBVR	0f6c4040a3c653f86446582c5c7c_1	Low-Grade GNS CNDiagnosis	Pathological	Low-grade g1 Progression	Primary				C71.6: Cereb	2402	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs020517	phs020517				
9	P1_GACHC0Y2	146dd653ca7b05e93717270b0c3b9.1	Low-Grade GNS CNDiagnosis	Pathological	Low-grade g1 Initial Diagnostic	Primary				C71.6: Cereb	1372	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs020517	phs020517				
10	P1_6VKN63NS	1629909022:14d375f6065a5d3ea87374.1	9400/0: Astro ICD-O-3.2	Pathological	Low-grade g1 Initial Diagnostic	Primary				C71.6: Cereb	3698	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs020517	phs020517				
11	P1_641N1C02	157a17b2b96186142cb3a19467f60.1	Low-Grade GNS CNDiagnosis	Pathological	Low-grade g1 Initial Diagnostic	Primary				C71.6: Cereb	1473	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs020517	phs020517				
12	P1_K07J21Z7	155550275c6ba0899f6857f60602b24.1	9400/0: Astro ICD-O-3.2	Pathological	Low-grade g1 Initial Diagnostic	Primary				C71.6: Cereb	2326	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs020517	phs020517				
13	P1_G62TCBVR	155b2b566dbd73c7de890fa1fac9a25.1	Low-Grade GNS CNDiagnosis	Pathological	Low-grade g1 Recurrent DI	Primary				C71.6: Cereb	1717	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs020517	phs020517				
14	P1B9JZ	195c131ab6c-4661-4060-76702740D05	9421/1: Piloc: ICD-O-3.2	Clinical	Not Reported	Initial Diagnostic	Primary			C71.6: Cereb	1040	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	phs022790	phs022790				

8



**Cohort Selector** (New Key Feature): The new Cohort Selector feature enables users to create and manage up to 10 cohorts. This feature offers flexibility to researchers, allowing them to create cohort groups according to their specific requirements. Users can do the following:

Studies(21)

Participants(13,557)

Diagnosis(19,642)

Treatment(7,000)

Treatment Response(698)

Survival(9,667)

CREATE COHORT

ADD PARTICIPANTS TO EXISTING COHORT

VIEW ALL COHORTS(3)

RESULTS PER PAGE: 10 1-10 OF 13557

10 row(s) selected

Participant Id	Race	Sex at Birth	dbGaP Accession
00301d78915737fa100f	White	Female	phs002431
0061cbb0846973206fcf	White	Male	phs002431
0065af91e89ee2859595	Hispanic or Latino;White	Female	phs002431
008b04a84717e7d007a4	White	Male	phs002431
009f9efaedc9602a28fb	White	Female	phs002431
00c5e4372375eb3627f3	White	Female	phs002431
00dfd38d509984eada22	White	Male	phs002431
00e3d1d383c8c08a25d2	White	Male	phs002431
00eff50ac98a200da9f9	White	Male	phs002431
01504e3b7031a9eec2b2	White	Female	phs002431

RESULTS PER PAGE: 10 1-10 OF 13557

Figure 8. Cohort Selection features visible on the Explore page Table tabs

- **Create New Cohort:** Select participant IDs from the facet search and create a new cohort. Users can name and describe the cohort for easy reference.
- **Add to Existing Cohort:** Add participants to existing cohorts or remove them. Entire cohorts can also be deleted as needed.
- **View All Cohorts:** View a list of all created cohorts, making it easier to manage and analyze groups.
  - Cohort ID: Create your own id to identify saved cohorts
  - Cohort description: Create a description for saved cohorts
  - Save Changes: Save the changes made to the selected cohort. This includes changes to cohort id, cohort description and any participants removed.
  - Download Selected Cohort:
    - Download the selected cohort in one of two formats.
      - Manifest CSV: a list of participant ids and high-level data.
      - Metadata JSON: a JSON file containing all metadata information for the participants in the selected cohort.

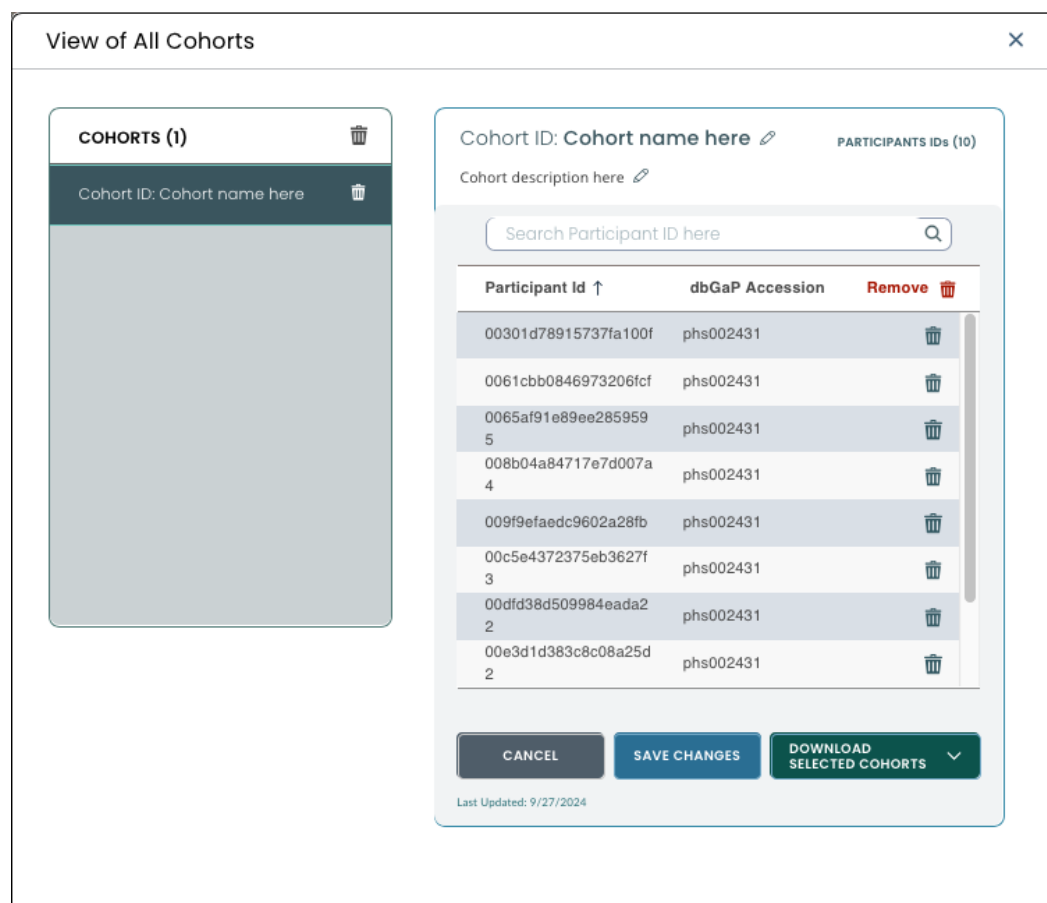
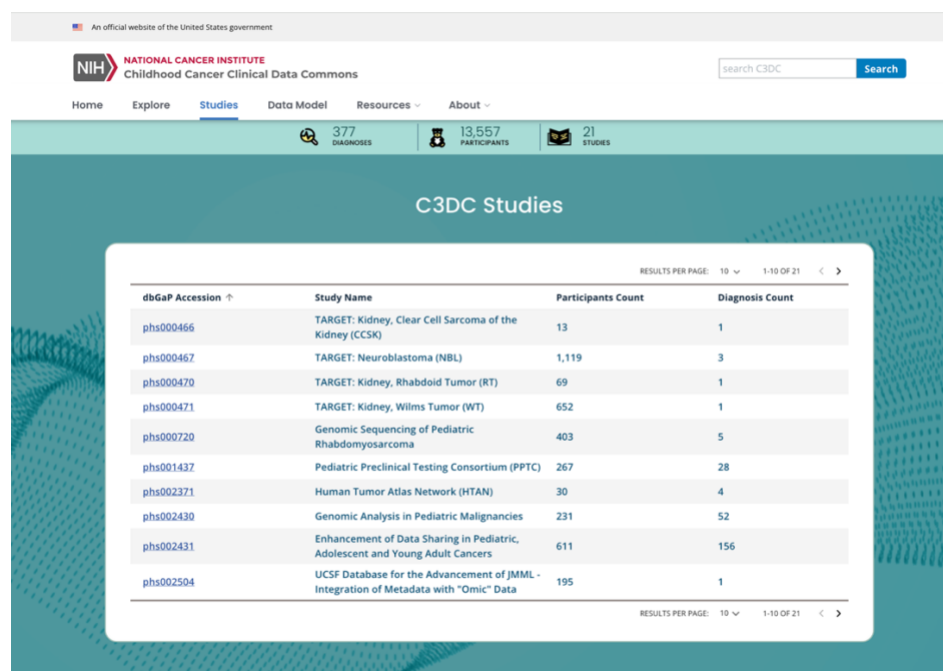


Figure 9. View All Cohorts and edit capabilities

## C3DC Studies Page

Users can navigate to the Studies Page to view the list of dbGaP accessions, study names, and counts for participants and diagnoses.



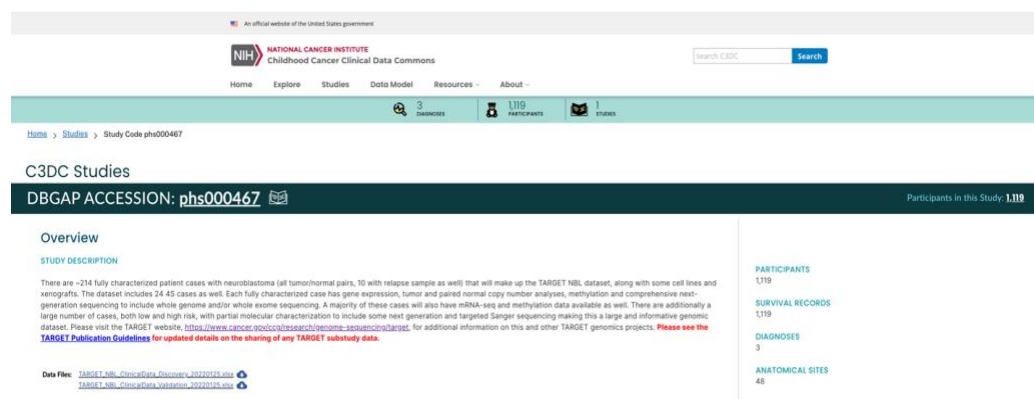
The screenshot shows the C3DC Studies page. At the top, there is a navigation bar with links: Home, Explore, Studies (selected), Data Model, Resources, and About. Below the navigation bar, there are three summary cards: 377 DIAGNOSES, 13,557 PARTICIPANTS, and 21 STUDIES. The main content area is titled "C3DC Studies" and displays a table of studies. The table has four columns: dbGaP Accession, Study Name, Participants Count, and Diagnosis Count. The table lists 10 studies, with the first 9 visible in the screenshot. The table is paginated, showing results 1-10 of 21.

dbGaP Accession	Study Name	Participants Count	Diagnosis Count
phs000466	TARGET: Kidney, Clear Cell Sarcoma of the Kidney (CCSK)	13	1
phs000467	TARGET: Neuroblastoma (NBL)	1,119	3
phs000470	TARGET: Kidney, Rhabdoid Tumor (RT)	69	1
phs000471	TARGET: Kidney, Wilms Tumor (WT)	652	1
phs000720	Genomic Sequencing of Pediatric Rhabdomyosarcoma	403	5
phs001437	Pediatric Preclinical Testing Consortium (PPTC)	267	28
phs002371	Human Tumor Atlas Network (HTAN)	30	4
phs002430	Genomic Analysis in Pediatric Malignancies	231	52
phs002431	Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	611	156
phs002504	UCSF Database for the Advancement of JMML - Integration of Metadata with "Omic" Data	195	1

Figure 10: The Studies Page

C3DC studies details page, by clicking on the dbGaP accession number, users can access detailed information about the studies. This action will redirect users to the dbGaP page to view a high-level overview of the information. Currently, source data files are only available for open access data (phs000466, phs000467, phs000470, and phs000471). If you are interested in accessing the controlled access data, please follow this [link](#) for instructions on how to access it

For other CCDI studies, source data can be found in the [CCDI Hub](#).



The screenshot shows the C3DC Studies details page for study phs000467. The page has a dark blue header with the text "C3DC Studies" and "DBGAP ACCESSION: phs000467". Below the header, there is a section titled "Overview" with a "STUDY DESCRIPTION" sub-section. The description states: "There are ~214 fully characterized patient cases with neuroblastoma (all tumor/normal pairs, 10 with relapse sample as well) that will make up the TARGET NBL dataset, along with some cell lines and xenografts. The dataset includes 24.45 cases as well. Each fully characterized case has gene expression, tumor and paired normal copy number analyses, methylation and comprehensive next-generation sequencing to include whole genome and/or whole exome sequencing. A majority of these cases will also have whole-exome and methylation data available as well. There are additionally a large number of cases, both low and high risk, with partial molecular characterization to include some next generation and targeted Sanger sequencing making this a large and informative genomic dataset. Please visit the TARGET website, <https://www.cancer.gov/ccr/research/ncmi-acquainting-target>, for additional information on this and other TARGET genomics projects. Please see the TARGET Publication Guidelines for updated details on the sharing of any TARGET study data." Below the description, there is a "Data Files" section with two links: "TARGET NBL Clinical Data Release v1.0.20210125.xlsx" and "TARGET NBL Clinical Data Release v1.0.20210125.xlsx". On the right side of the page, there is a sidebar with summary statistics: PARTICIPANTS (1,119), SURVIVAL RECORDS (1,119), DIAGNOSES (3), and ANATOMICAL SITES (48).

STUDY DESCRIPTION
There are ~214 fully characterized patient cases with neuroblastoma (all tumor/normal pairs, 10 with relapse sample as well) that will make up the TARGET NBL dataset, along with some cell lines and xenografts. The dataset includes 24.45 cases as well. Each fully characterized case has gene expression, tumor and paired normal copy number analyses, methylation and comprehensive next-generation sequencing to include whole genome and/or whole exome sequencing. A majority of these cases will also have whole-exome and methylation data available as well. There are additionally a large number of cases, both low and high risk, with partial molecular characterization to include some next generation and targeted Sanger sequencing making this a large and informative genomic dataset. Please visit the TARGET website, <a href="https://www.cancer.gov/ccr/research/ncmi-acquainting-target">https://www.cancer.gov/ccr/research/ncmi-acquainting-target</a> , for additional information on this and other TARGET genomics projects. Please see the TARGET Publication Guidelines for updated details on the sharing of any TARGET study data.

Summary Statistics
PARTICIPANTS
1,119
SURVIVAL RECORDS
1,119
DIAGNOSES
3
ANATOMICAL SITES
48

Figure 11: The Studies Details page

## C3DC Data Model Page

The [data model](#) is developed collaboratively with multiple organizations to establish standard terms for pediatric cancer. In this harmonization effort, we are using CDEs (Common Data Elements) to enhance data accuracy, consistency, and interoperability across health research studies. CDEs are defined in the caDSR (Cancer Data Standards Registry and Repository) and provide controlled terms, vocabularies, detailed information on data representation, and robust metadata. The C3DC data model schema consists of well-defined classes with attributes and permissible values. The model focuses on a set of data elements, with plans to include more elements in future releases.

The screenshot shows the NIH National Cancer Institute Childhood Cancer Clinical Data Commons website. The header includes the NIH logo, the text "NATIONAL CANCER INSTITUTE Childhood Cancer Clinical Data Commons", and a search bar labeled "search C3DC". The navigation menu includes "Home", "Explore", "Studies", "Data Model", "Resources", and "About". The main content area has a teal background with the title "C3DC Data Model" and a background image of code. Below the title, there is a paragraph of text describing the C3DC harmonized Data Model and its purpose. To the right of the text is a small image showing a complex network diagram with many nodes and connections.

The C3DC harmonized [Data Model](#) unites multiple childhood cancer studies and drives the search experience. Initially, the C3DC database model will focus on a limited set of data elements, such as participant diagnosis, to establish processes for data ingestion, harmonization, data modeling, and storage. The C3DC model is developed collaboratively with contributions from Data for the Common Good at University of Chicago, Frederick National Laboratory, NCI's Semantics Infrastructure, Cancer Research Data Commons, and National Cancer Childhood Registry team members. The resulting model will collectively establish the standard terms for pediatric cancer and will serve as guidance to researchers.

NCI encourages researchers to use CDEs to enhance data accuracy, consistency, and interoperability across diverse areas of health research. CDEs are pieces of data captured in multiple datasets across different studies. The caDSR, short for the [Cancer Data Standards Registry and Repository](#) is a structured repository for clinical and research data. Its content semantically defines data through controlled terms and vocabularies, offering detailed information about data representation, including permissible values, data types (numeric, date, text), and the ability to associate various local or source names with the same data value. Moreover, CDEs provide robust metadata, encompassing references to external standards, usage instructions, and examples of expected data formats. The C3DC target schema will consist of well-defined classes with attributes and permissible values.

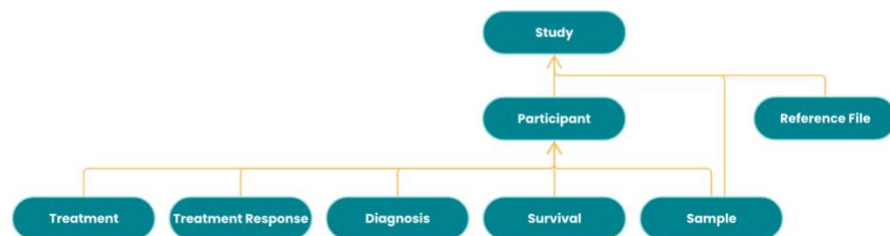
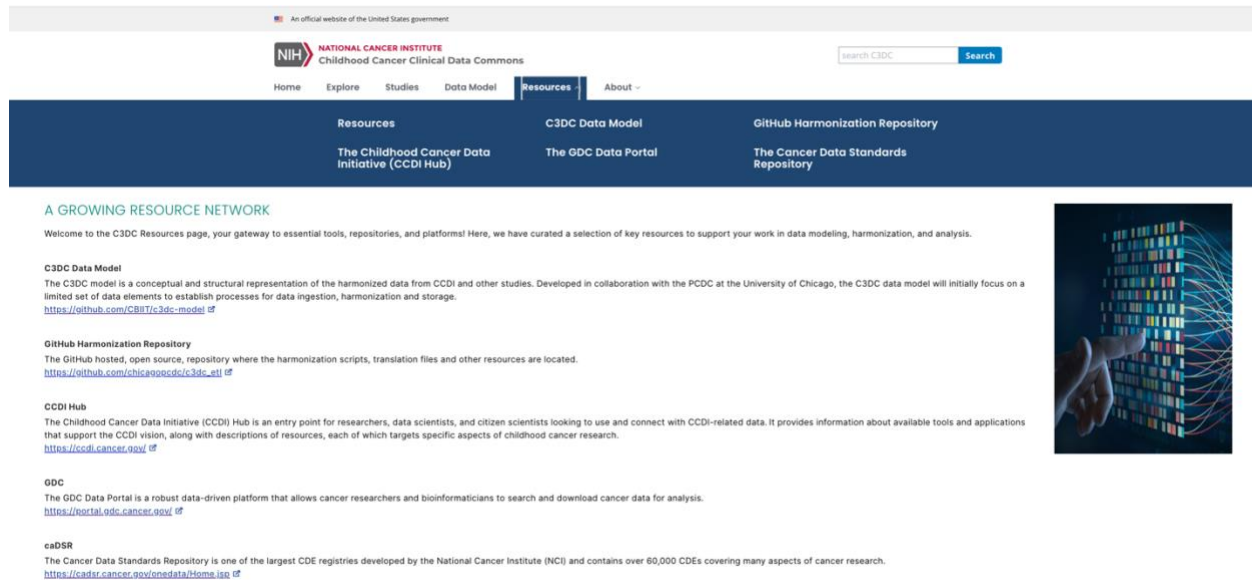


Figure 12: The Data Model page

## C3DC Resource Page

By clicking on each resource, users can access a range of useful tools and information available on the site.



The screenshot displays the C3DC Resource Page, an official website of the United States government. The page features a header with the NIH logo and the text "NATIONAL CANCER INSTITUTE Childhood Cancer Clinical Data Commons". A search bar is located in the top right corner. The main navigation menu includes links for Home, Explore, Studies, Data Model, Resources, and About. The Resources section is highlighted, showing three sub-sections: The Childhood Cancer Data Initiative (CCDI Hub), The GDC Data Portal, and The Cancer Data Standards Repository. Below this, a section titled "A GROWING RESOURCE NETWORK" welcomes users and provides links to various resources. The resources listed are: C3DC Data Model, GitHub Harmonization Repository, CCDI Hub, GDC, and caDSR. Each resource is accompanied by a brief description and a link to the resource.

An official website of the United States government

NIH NATIONAL CANCER INSTITUTE  
Childhood Cancer Clinical Data Commons

search C3DC Search

Home Explore Studies Data Model Resources About

Resources

The Childhood Cancer Data Initiative (CCDI Hub)

C3DC Data Model

The GDC Data Portal

GitHub Harmonization Repository

The Cancer Data Standards Repository

A GROWING RESOURCE NETWORK

Welcome to the C3DC Resources page, your gateway to essential tools, repositories, and platforms! Here, we have curated a selection of key resources to support your work in data modeling, harmonization, and analysis.

**C3DC Data Model**

The C3DC model is a conceptual and structural representation of the harmonized data from CCDI and other studies. Developed in collaboration with the PCDC at the University of Chicago, the C3DC data model will initially focus on a limited set of data elements to establish processes for data ingestion, harmonization and storage.  
<https://github.com/CBIRT/c3dc-model>

**GitHub Harmonization Repository**

The GitHub hosted, open source, repository where the harmonization scripts, translation files and other resources are located.  
<https://github.com/chicagopcdc/c3dc-etl>

**CCDI Hub**

The Childhood Cancer Data Initiative (CCDI) Hub is an entry point for researchers, data scientists, and citizen scientists looking to use and connect with CCDI-related data. It provides information about available tools and applications that support the CCDI vision, along with descriptions of resources, each of which targets specific aspects of childhood cancer research.  
<https://ccdi.cancer.gov/>

**GDC**

The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis.  
<https://portal.gdc.cancer.gov/>

**caDSR**

The Cancer Data Standards Repository is one of the largest CDE registries developed by the National Cancer Institute (NCI) and contains over 60,000 CDEs covering many aspects of cancer research.  
<https://cadsr.cancer.gov/onedata/Home.jspx>




Figure 13: The Resource page

## C3DC About Page

Users can navigate to the About Page by clicking the link on the Home Page menu bar, where you will find more information about the content of C3DC. This includes details such as dataset and data model information as well as links to useful resources. There are dropdown menus to access the following documents:

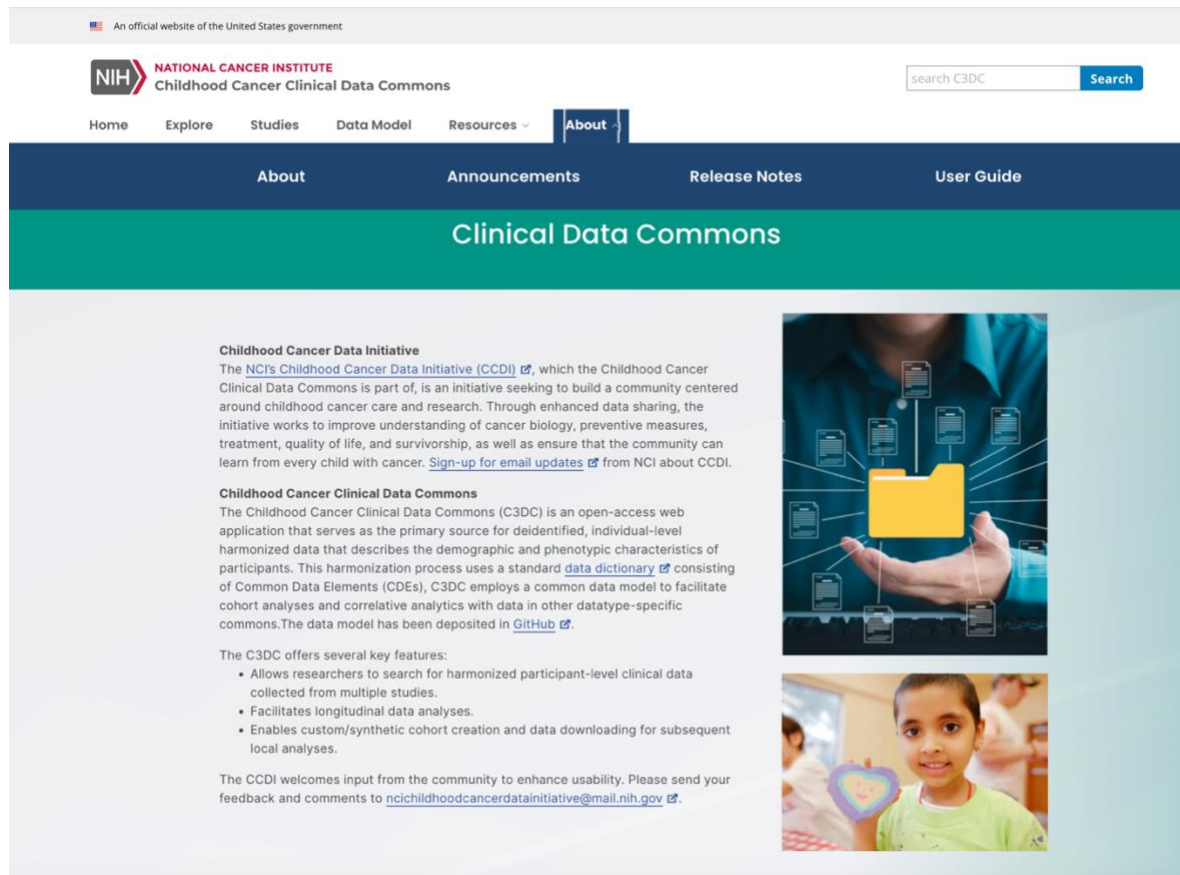


Figure 14: The About page

## Announcements

The announcement page contains all C3DC updates, both data and application, with the newest release information at the top of the list.

### C3DC Announcements

[All](#) [Data Update](#) [Application Update](#)

#### Datasets Update

November 7, 2024

This release includes three new harmonized data sets (phs003432, phs003519, and phs000466) and updates to existing data sets (phs002517, and phs002790) on the C3DC Explore page. All the data have been aligned to the most recent version of the data model. For more detailed information about these changes and to review the full list of updates, please refer to our data release update

[Read More](#)



#### Data Model Update

November 7, 2024

In this release, we are introducing enhancements to our data model. Specifically, we have added two new nodes: treatment and treatment response. These additions will allow for more detailed tracking and analysis of treatment regimens and their outcomes, providing valuable insights for researchers and clinicians.



Figure 15: The Announcements page

## Release Note

The Release Note page contains the change logs of all releases, noting the differences between the previous versions of both the data and application.

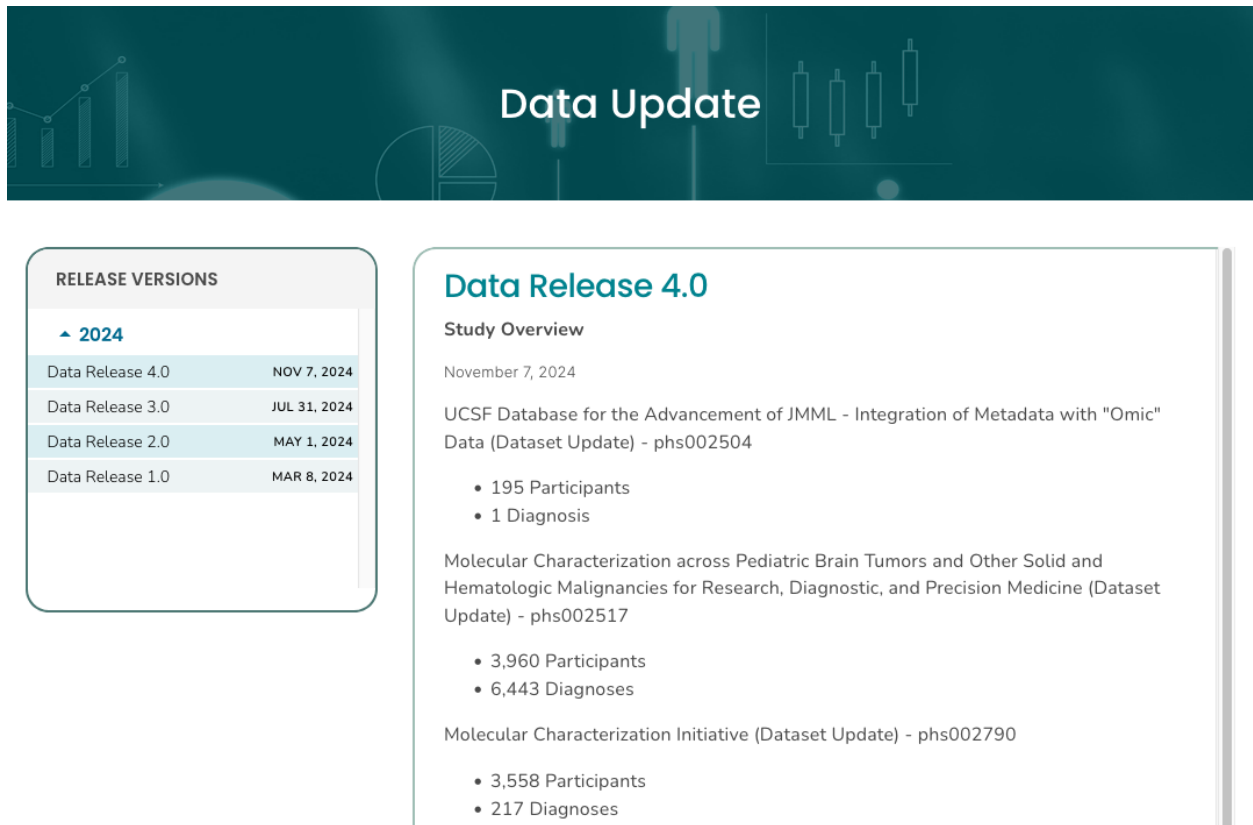


Figure 16: The Release Notes page



## User Guide

The User Guide page contains the mission statements for CCDI and C3DC, as well as useful resources related to the C3DC project.

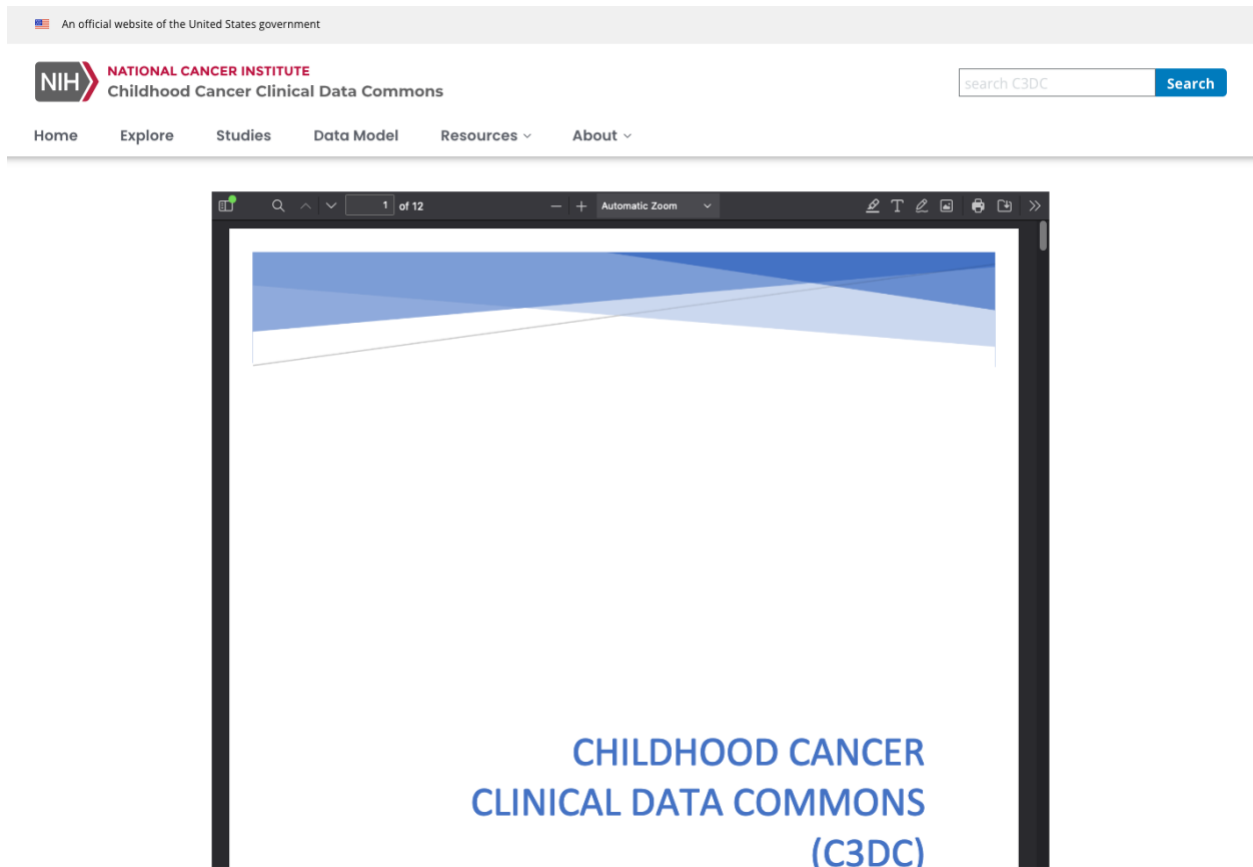


Figure 17: The User Guide