



# CHILDHOOD CANCER CLINICAL DATA COMMONS (C3DC) User Guide

07/31/2024

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

## Table of Contents

<b>INTRODUCTION AND OVERVIEW</b>	<b>3</b>
<b>C3DC HOME PAGE</b>	<b>4</b>
<b>C3DC EXPLORE PAGE</b>	<b>5</b>
FACET SEARCH	6
CHART VISUALIZATION	6
TABLE VISUALIZATION	7
DOWNLOAD HARMONIZED DATA	7
<b>C3DC STUDIES PAGE</b>	<b>8</b>
C3DC STUDIES DETAILS PAGE	8
<b>C3DC RESOURCE PAGE</b>	<b>10</b>
<b>C3DC ABOUT PAGE</b>	<b>11</b>
LATEST RELEASE	11
USER GUIDE	11

## 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.

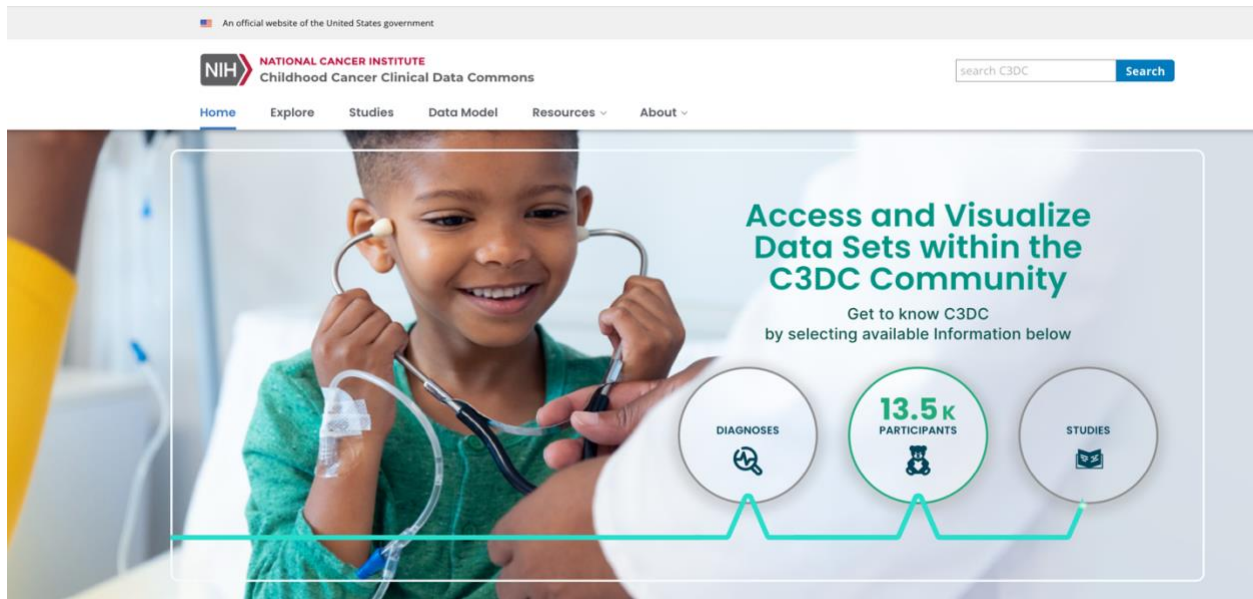


Figure 1: C3DC Home Page - top

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

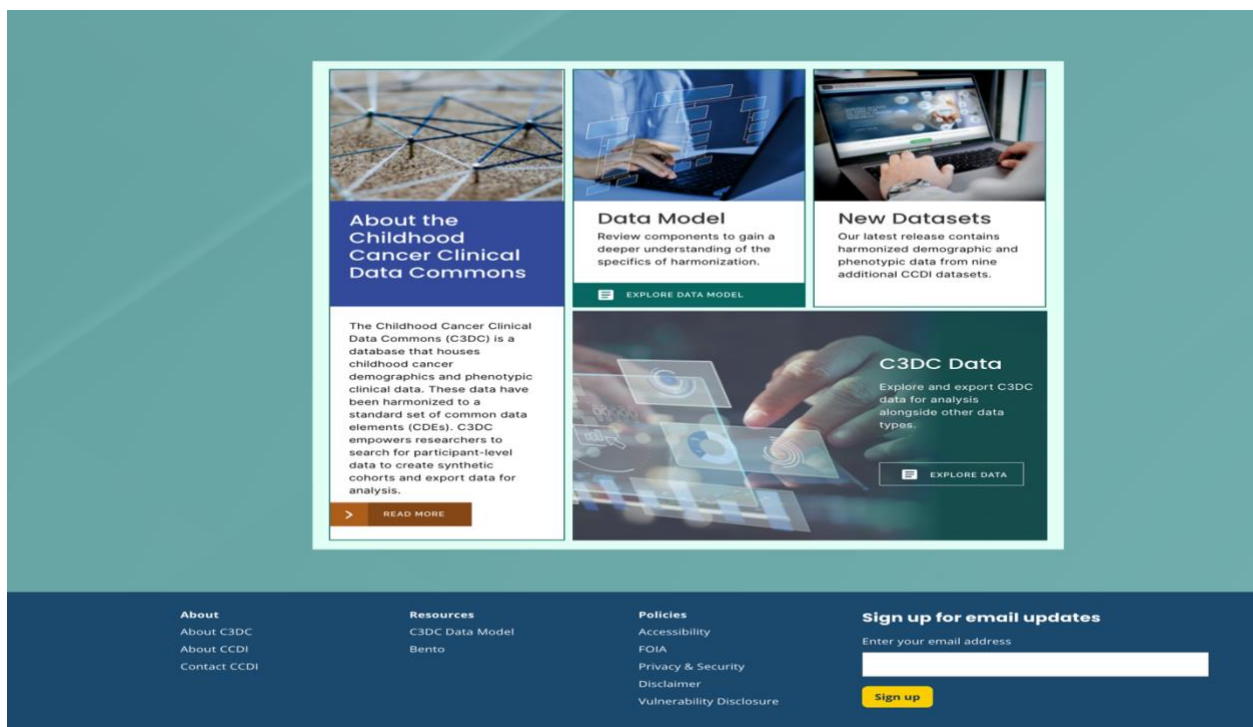


Figure 2: C3DC Home Page - bottom

## 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.

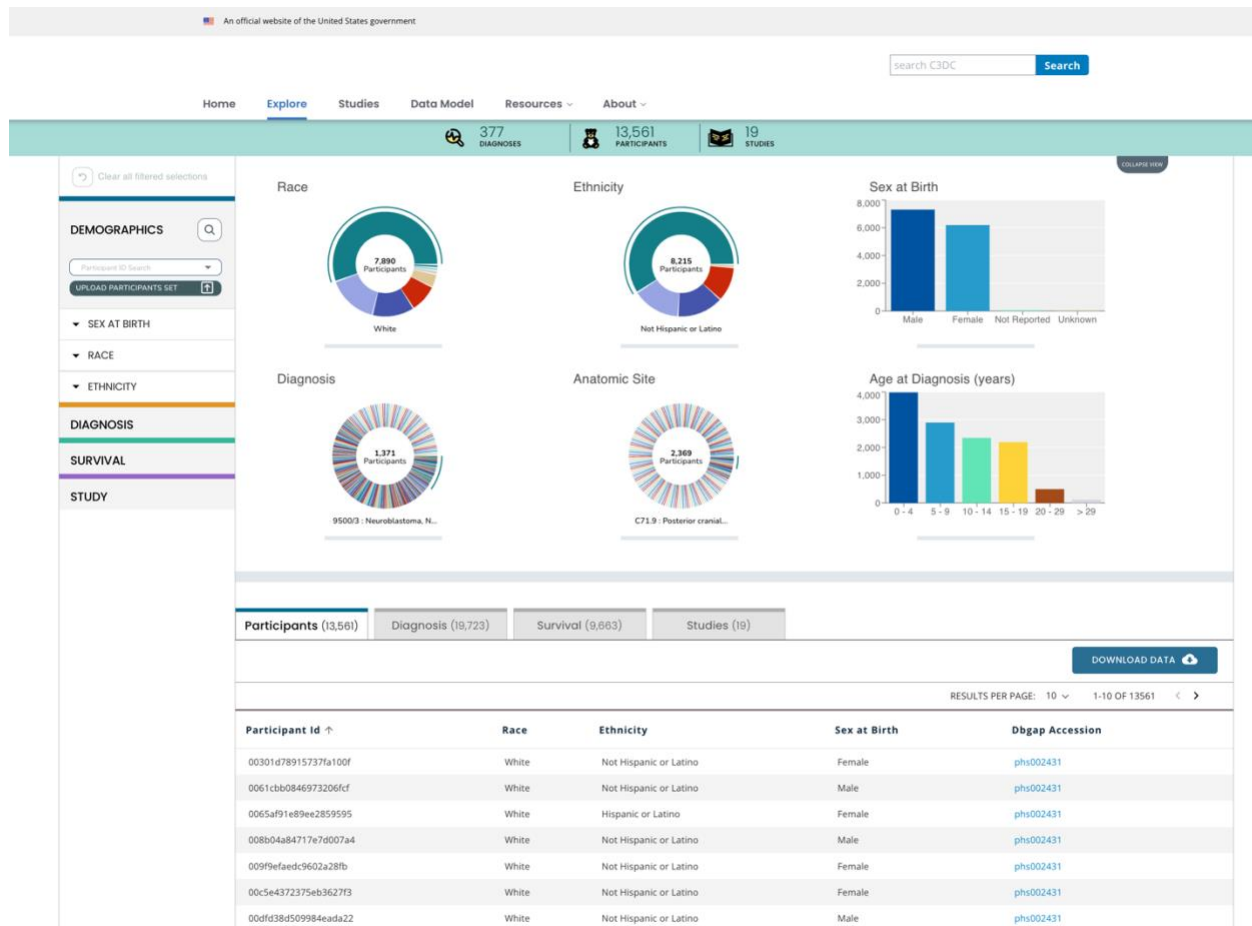


Figure 3: The Explore page – Faceted Search (left), Visualization (top), and Data Tables (bottom)

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

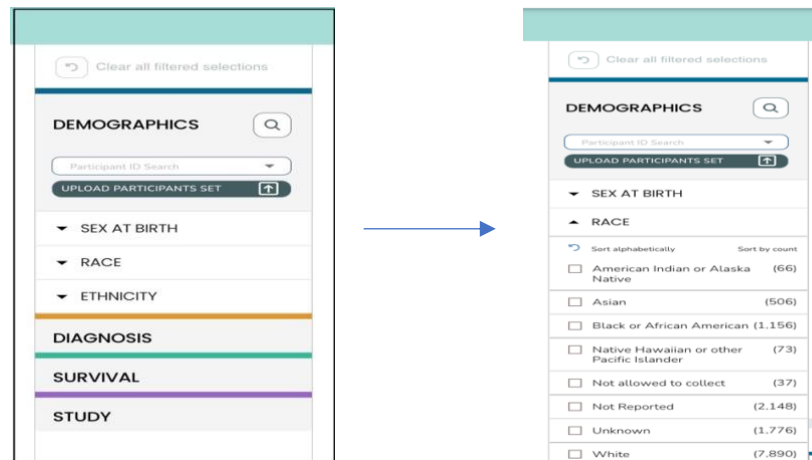


Figure 4: Performing a faceted search on the C3DC Explore Page

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

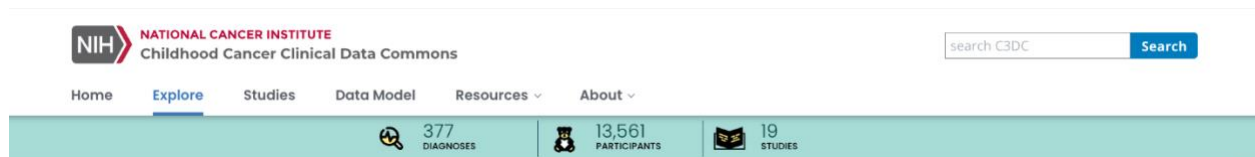


Figure 5: Results returned from a faceted search in the Stats bar

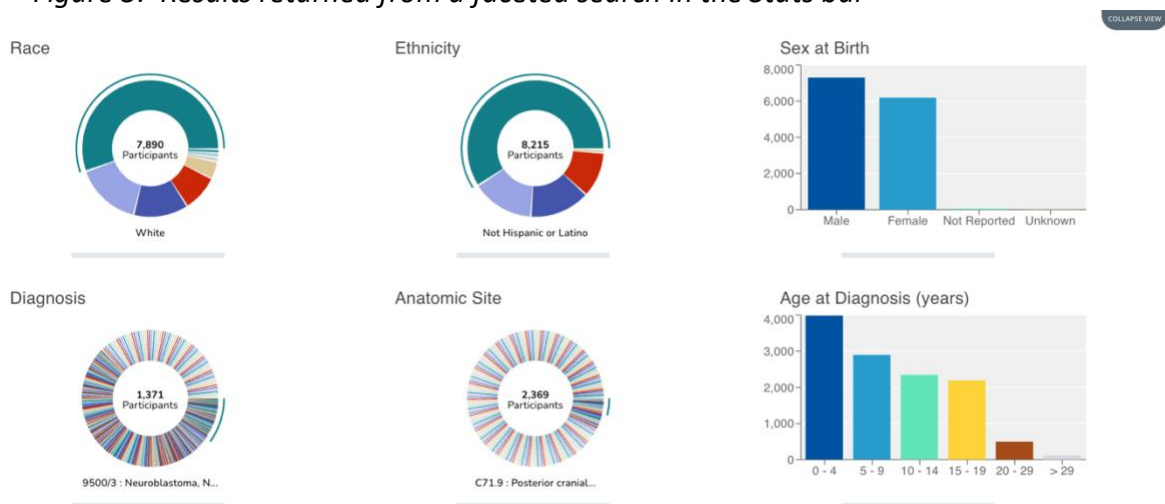


Figure 6: Results returned from a faceted search in six graphs in the Visualization section

Count of Participant Record				
Participants (13,532)	Diagnosis (19,694)	Survival (9,663)	Studies (18)	
<div> <div>Download Data</div> </div>				
<div> <div>RESULTS PER PAGE: 10</div> <div>1-10 OF 13532</div> <div> <div></div> <div></div> </div> </div>				
Participant Id ↑	Race	Ethnicity	Sex at Birth	dbGaP Accession
00301d78915737fa100f	White	Not Hispanic or Latino	Female	<a href="#">phs002431</a>
0061cbb0846973206cf	White	Not Hispanic or Latino	Male	<a href="#">phs002431</a>
0065af91e89ee2859595	White	Hispanic or Latino	Female	<a href="#">phs002431</a>
008b04a84717e7d007a4	White	Not Hispanic or Latino	Male	<a href="#">phs002431</a>
009f9eafedc9602a28fb	White	Not Hispanic or Latino	Female	<a href="#">phs002431</a>
00c5e4372375eb3627f3	White	Not Hispanic or Latino	Female	<a href="#">phs002431</a>
00dfd38d509984eada22	White	Not Hispanic or Latino	Male	<a href="#">phs002431</a>
00e3d1d383c8c08a25d2	White	Not Hispanic or Latino	Male	<a href="#">phs002431</a>
00eff50ac98a200da9f9	White	Not Hispanic or Latino	Male	<a href="#">phs002431</a>
01504e3b7031a9eec2b2	White	Not Hispanic or Latino	Female	<a href="#">phs002431</a>
<div> <div>RESULTS PER PAGE: 10</div> <div>1-10 OF 13532</div> <div> <div></div> <div></div> </div> </div>				

**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.

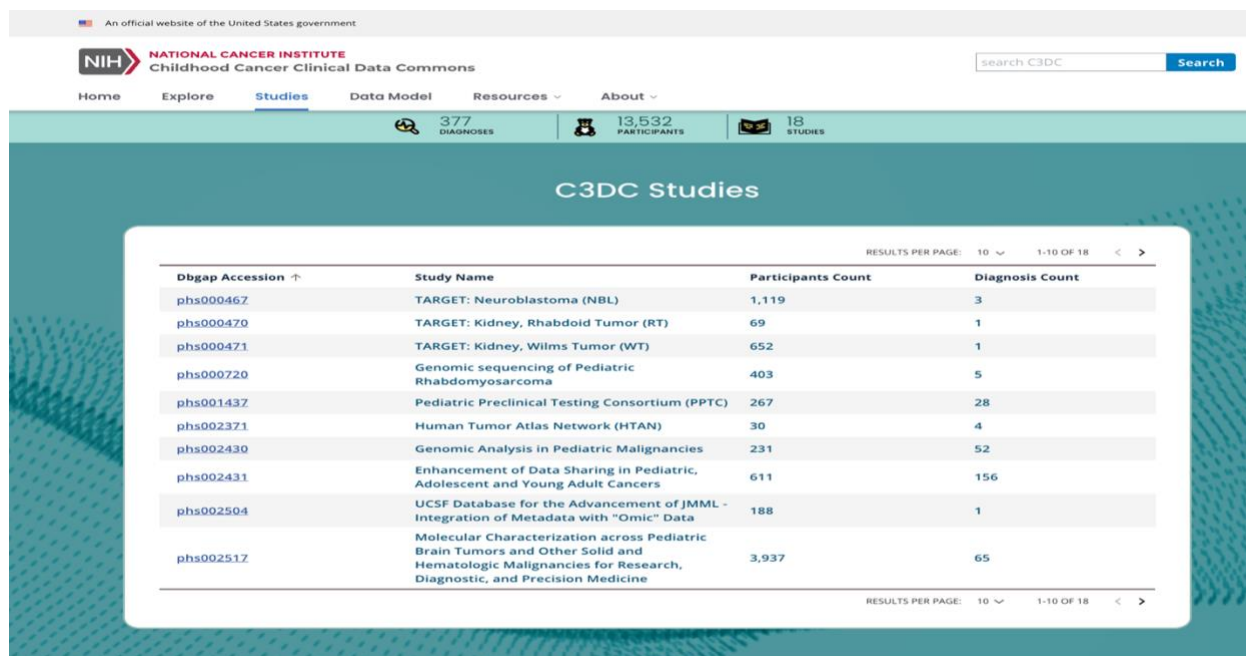
[illegible]

7



## 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.



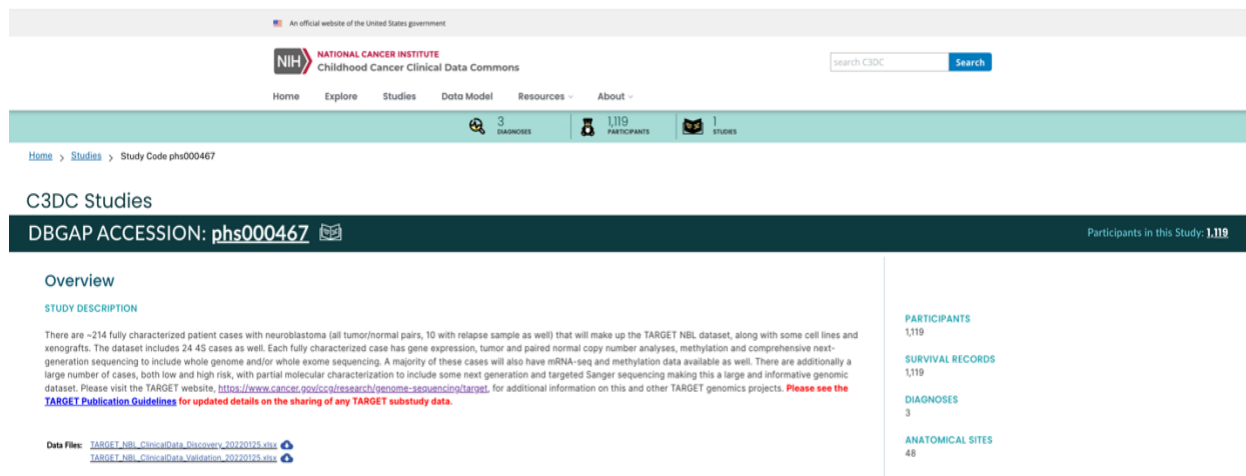
The screenshot shows the C3DC Studies page. At the top, there is a navigation bar with links: Home, Explore, Studies (active), Data Model, Resources, and About. Below the navigation bar, there are three summary cards: 377 DIAGNOSES, 13,532 PARTICIPANTS, and 18 STUDIES. The main content area is titled "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 18 studies, with the first 10 visible in the screenshot. The table is paginated, showing results 1-10 of 18.

Dbgap Accession	Study Name	Participants Count	Diagnosis Count
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	188	1
phs002517	Molecular Characterization across Pediatric Brain Tumors and Other Solid and Hematologic Malignancies for Research, Diagnostic, and Precision Medicine	3,937	65

Figure 9: 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 (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 green header with the text "C3DC Studies" and "DBGAP ACCESSION: phs000467". Below the header, there is a navigation bar with links: Home, Explore, Studies (active), Data Model, Resources, and About. The main content area is titled "Overview" and contains a "STUDY DESCRIPTION" section. The description states that 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 mRNA-seq 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/ccp/research/genome-sequencing/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 substudy data.

Below the description, there is a "Data Files" section with two links: [TARGET\\_NBL\\_ClinicalData\\_Discovery\\_20220215.xlsx](#) and [TARGET\\_NBL\\_ClinicalData\\_Validation\\_20220215.xlsx](#).

On the right side of the page, there is a sidebar with a list of statistics: PARTICIPANTS (1,119), SURVIVAL RECORDS (1,119), DIAGNOSES (3), and ANATOMICAL SITES (48).

Figure 10: 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 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.

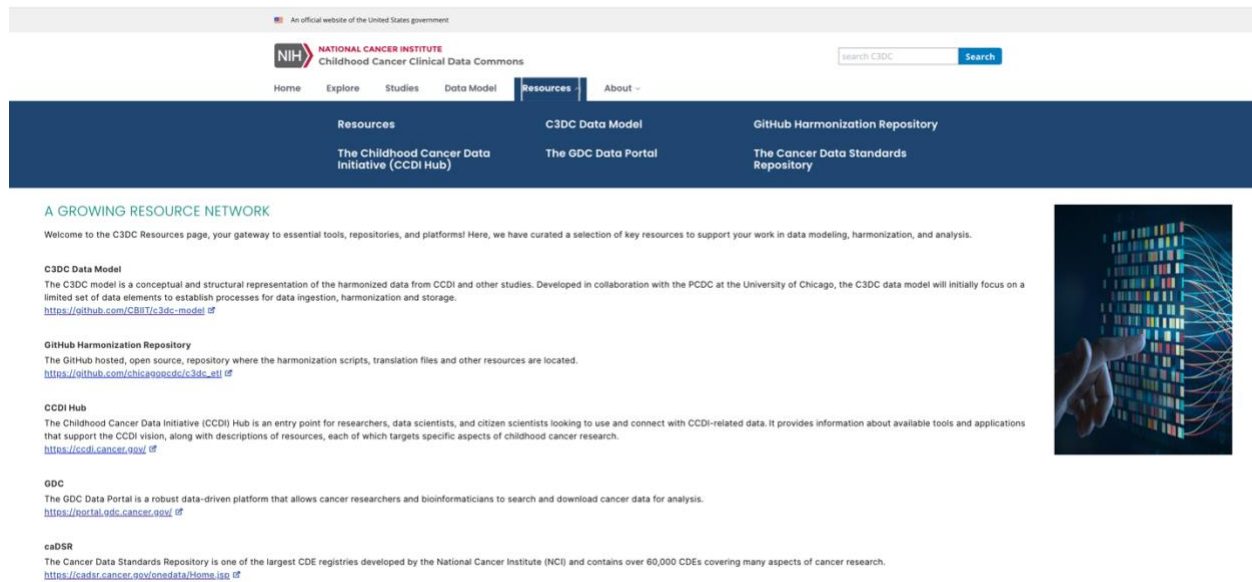
NCI encourages researchers to use CDEs to enhance data accuracy, consistency, and interoperability across diverse areas of health research. CDEs are pieces of data captured in multiple datasets across different studies. The caDSR, short for the [Cancer Data Standards Registry and Repository](#) is a 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.

```
graph TD
    Study --> Participant
    Study --> ReferenceFile[Reference File]
    Participant --> Treatment
    Participant --> TreatmentResponse
    Participant --> Diagnosis
    Participant --> Survival
    Participant --> Sample
```

Figure 11: 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 shows the C3DC Resource Page, an official website of the United States government. The page features a navigation bar with links to Home, Explore, Studies, Data Model, Resources, and About. Below the navigation bar, there are three main sections: Resources, C3DC Data Model, and GitHub Harmonization Repository. The Resources section includes links to The Childhood Cancer Data Initiative (CCDI Hub), The GDC Data Portal, and The Cancer Data Standards Repository. The C3DC Data Model section includes a link to The GDC Data Portal. The GitHub Harmonization Repository section includes a link to The Cancer Data Standards Repository. The page also features a search bar and a footer with the text "A GROWING RESOURCE NETWORK".

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 C3DC Data Model GitHub Harmonization Repository

The Childhood Cancer Data Initiative (CCDI Hub) The GDC Data Portal 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/C3DC/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/c3dcgpc/c3dc-gh>

**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>




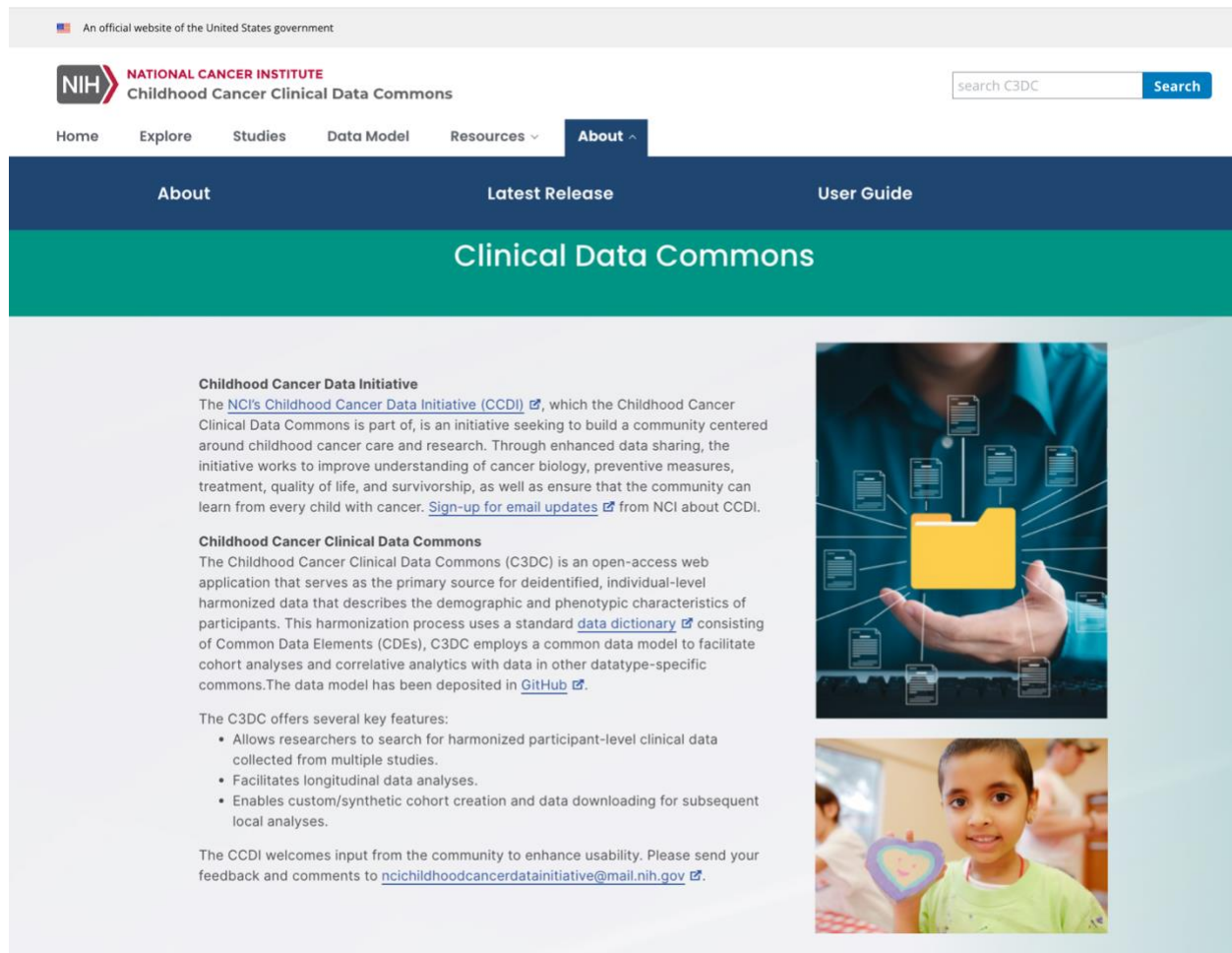
Figure 12: The Resource page

## 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:

[Latest Release](#)

[User Guide](#)



The screenshot shows the 'About' page of the Childhood Cancer Clinical Data Commons (C3DC). At the top, there is a header with the NIH logo and the text 'NATIONAL CANCER INSTITUTE Childhood Cancer Clinical Data Commons'. A search bar with the text 'search C3DC' and a 'Search' button is located on the right. Below the header is a navigation bar with links: 'Home', 'Explore', 'Studies', 'Data Model', 'Resources', and 'About'. The 'About' link is highlighted. Below the navigation bar is a dark blue banner with the text 'Clinical Data Commons'. The main content area is divided into two columns. The left column contains text about the 'Childhood Cancer Data Initiative' and the 'Childhood Cancer Clinical Data Commons'. The right column features two images: the top one shows a hand holding a yellow folder with various documents floating around it, and the bottom one shows a young child holding a heart-shaped object.

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

About Latest Release User Guide

### Clinical Data Commons

**Childhood Cancer Data Initiative**  
The NCI's [Childhood Cancer Data Initiative \(CCDI\)](#), which the Childhood Cancer Clinical Data Commons is part of, is an initiative seeking to build a community centered around childhood cancer care and research. Through enhanced data sharing, the initiative works to improve understanding of cancer biology, preventive measures, treatment, quality of life, and survivorship, as well as ensure that the community can learn from every child with cancer. [Sign-up for email updates](#) from NCI about CCDI.

**Childhood Cancer Clinical Data Commons**  
The Childhood Cancer Clinical Data Commons (C3DC) is an open-access web application that serves as the primary source for deidentified, individual-level harmonized data that describes the demographic and phenotypic characteristics of participants. This harmonization process uses a standard [data dictionary](#) consisting of Common Data Elements (CDEs), C3DC employs a common data model to facilitate cohort analyses and correlative analytics with data in other datatype-specific commons. The data model has been deposited in [GitHub](#).

The C3DC offers several key features:

- Allows researchers to search for harmonized participant-level clinical data collected from multiple studies.
- Facilitates longitudinal data analyses.
- Enables custom/synthetic cohort creation and data downloading for subsequent local analyses.

The CCDI welcomes input from the community to enhance usability. Please send your feedback and comments to [ncichildhoodcancerdatainitiative@mail.nih.gov](mailto:ncichildhoodcancerdatainitiative@mail.nih.gov).

Figure 14: The about page