



CHILDHOOD CANCER DATA CATALOG (CCDC)

User Guide

5/15/2022

Version	Date	Description	Author
1.0.0	04/15/2022	Initial Version	CCDC Team
1.1.0	05/15/2022	Updated information about result filtering and display. Added resources to the site. Replaced some images with improved contrast versions.	CCDC Team

Contents

Introduction and Overview	3
Home page	3
Search Catalog Page	4
Dataset Results Views	5
Card View	6
Table View	6
Search	7
Export	8
Dataset Detail Page	8
Administrative data	9
Metadata	10
Core Data Elements	10
Additional Data Elements	11
Participating Resource Page	11
Resource Filtering	12
Resource Card View	13
Card View	13
Participating Resource Detail Page	14
Administrative data	16
Descriptive Data	16
Resource Description	16
Resource Tools	17
Data Content Type	17
Data Access	17
Dataset Summaries	17
About CCDI Data Catalog Page	18
Contributing to the CCDI Data Catalog	18
Glossary Page	19
User Guide	20
Documentation Search Page	20
Additional Support	22
Appendix A: Participating Resources	23

Introduction and Overview

The Childhood Cancer Data Catalog (CCDC) is an inventory of pediatric oncology data resources developed as part of the National Cancer Institute's (NCI) [Childhood Cancer Data Initiative](#) (CCDI). The CCDC includes listing of childhood cancer repositories, registries, programs, knowledgebases, as well as other catalogs that either manage or refer to pediatric cancer data. The aim of the CCDC is to help researchers, clinicians, and citizen scientist learn about existing pediatric resources for the development of new biomedical hypothesis, formation of new secondary analysis of data within existing resources, and to provide connections to existing pediatric oncology research sites.

This document describes the purpose and the general use of the CCDC site. By design, the CCDC only collects and holds the metadata from a specific resource and not the primary participant or sample data of that resource. Information on understanding this metadata is defined below, but users interested in specific resource data are encouraged to reach out to content holders using the point of contact (POC) information available on the CCDC site.

Home page

The CCDC homepage is the landing page for users navigating to the CCDC site. It provides a brief introduction to the site, allows users to perform a search of the cataloged data resources, and shares a

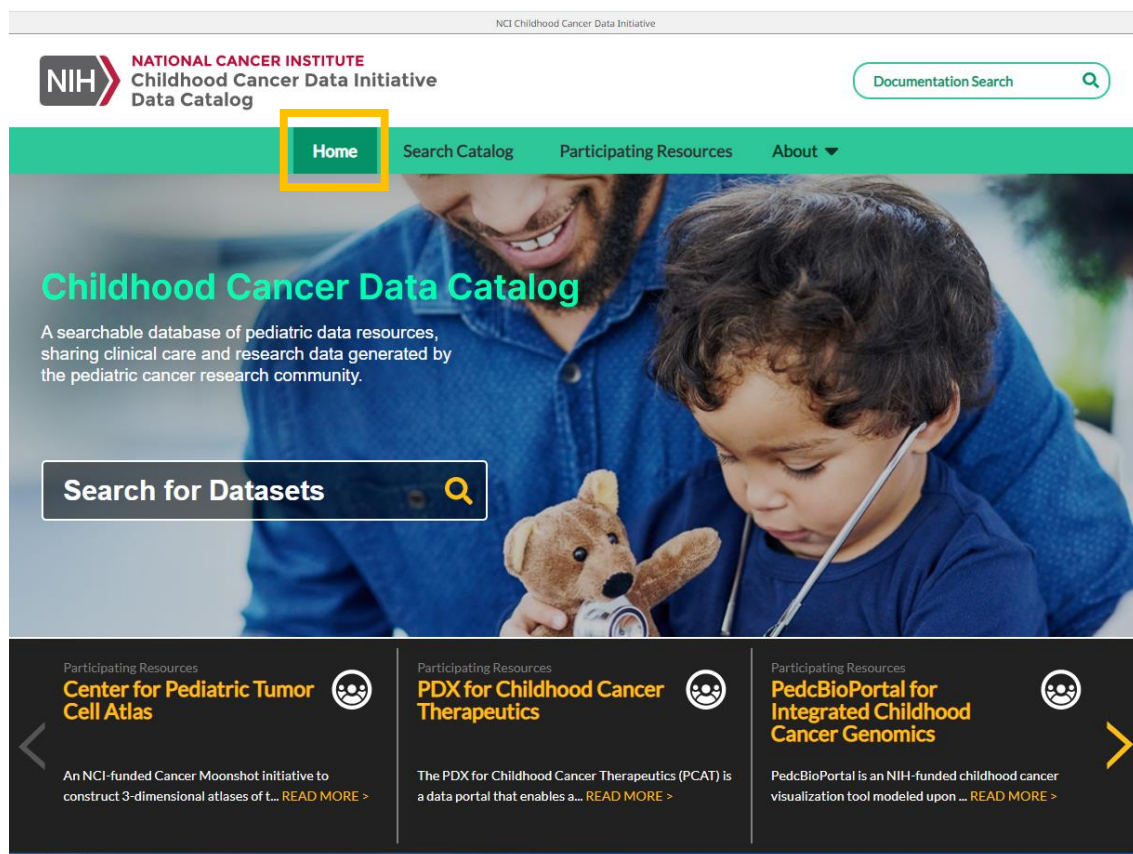


Figure 1: CCDC Home Page

limited number of participating resources available on the site (Figure 1). This page is also reachable by selecting 'Home' on the menu bar from any page in the site.

Users wishing to search the cataloged data resources can use the "Search for Datasets" field (Fig. 1) to perform a search that will automatically navigate a user to the Search Catalog page. More information on the Search Catalog page and the catalog search parameters is available below.

The home page also features a rotating carousel of participating resources available on the site to which users can navigate. This carousel contains example resources that will periodically change. Each example lists the resource name and a brief introduction to the resource. Users can either select the resource name or the 'READ MORE >' links to navigate to the resource's details page to learn more about the resource. More information about the participating resource details page is available below.

Search Catalog Page

The CCDC Search Catalog page allows users to search the individual datasets available from CCDC participating resources (Figure 2). You can navigate to the page by selecting 'Search Catalog' on the menu bar. Introductory information is available for each dataset along with a link to a more detailed dataset page. The Search Catalog page features a free text search that allows for filtering datasets.

The screenshot displays the NIH National Cancer Institute Childhood Cancer Data Initiative Search Catalog page. The header includes the NIH logo and the text "NATIONAL CANCER INSTITUTE Childhood Cancer Data Initiative Data Catalog". A search bar labeled "Documentation Search" is in the top right. The navigation bar has links for "Home", "Search Catalog" (highlighted with a yellow box), "Participating Resources", and "About".

The main section is titled "Search Results" with a help icon. Below it is a search bar labeled "Search the Catalog" with a "SUBMIT" button. The results are displayed in a table view, with options for "Card View" and "Table View". The table shows search results for "Acute Lymphoblastic Leukemia (ALL) Expansion Phase 2" and "Acute Lymphoblastic Leukemia (ALL) Expansion Phase 3 Acute Leukemias of Ambiguous Lineage (ALAL)".

Resources

- ☐ CBTRUS
- ☐ CCDI
- ☐ CCSS
- ☐ CGCI
- ☒ CIVIC
- ☐ COG
- ☐ GDC
- ☐ HCMI
- ☐ HTAN
- ☐ JAX PDX
- ☒ Kids First
- ☐ MTP
- ☐ MyPART
- ☐ NCCR
- ☐ Oncogenomics

Search Results

Acute Lymphoblastic Leukemia (ALL) Expansion Phase 2 Project

TARGET

Case Disease Diagnosis: Acute Lymphocytic Leukemia

Case Count: 51

Description: In the expansion effort (Phase 2), TARGET investigators analyzed tumors from pediatric patients, most who experienced an early bone marrow relapse (within 4 years of initial diagnosis), to identify new therapeutic approaches and/or biomarkers that correlate with poor clinical outcome to treat childhood pre-cursor B-cell ALL. The tissues used in this study were collected from patients enrolled in Children's Oncology Group (COG) biology studies and clinical trials.

Acute Lymphoblastic Leukemia (ALL) Expansion Phase 3 Acute Leukemias of Ambiguous Lineage (ALAL) Project

TARGET

Case Disease Diagnosis: Acute Lymphocytic Leukemia

Case Count: 115

Description: Leukemia is a cancer of blood cells and can arise within distinct lineages, either lymphoid or myeloid. On occasion, patients present with an acute leukemia for which a specific lineage cannot be clearly determined. These cases are classified as acute leukemias of ambiguous lineage (ALAL), and they account for less than 4% of all acute leukemias across age groups.

Acute Lymphoblastic Leukemia (ALL) Pilot Phase 1 Project

TARGET

Figure 2: CCDC Search Catalog Page

The catalog currently includes the following types of datasets.

- **Collection** - A group of datasets collected for any reason by an organization of researchers, stewards, or stakeholders either pertaining to a common theme or for a common purpose.
- **Program** - A coherent assembly of plans, project activities, and supporting resources contained within an administrative framework, the purpose of which is to implement an organization's mission or some specific program-related aspect of that mission.
- **Project** - Any specifically defined piece of work that is undertaken or attempted to meet the goals of a program and that involves one or more case studies. Also known as a Study or Trial.

Please note, the datasets available on the site are periodically updated and the types of datasets available may change.

Dataset Results Views

Information about available datasets is displayed in the results pane on the right-hand side of the Search Catalog page. The results pane can be configured in two ways with either a Card View (Figure 3) or a Table View (Figure 4). Each view can be toggled back and forth without altering the results of a performed search.

In both views, users can select a dataset name to navigate to a page with more detailed information about the dataset. Within each result, selecting the participating resource name/abbreviation (i.e., Kids First) will navigate to a more detailed page about that participating resource.

Card View Table View EXPORT

SORT BY Dataset Results per Page: 10 Showing 1-9 of 9

Burkitt Lymphoma Genome Sequencing Project (BLGSP) Project

CGCI

Case Disease Diagnosis: Burkitt Lymphoma

Case Count: 126

Sample Assay Method: RNA Sequencing Whole Genome Sequencing

Sample Count: 386

Description: The goal of the Burkitt Lymphoma Genome Sequencing Project (BLGSP) is to explore genetic changes in patients with Burkitt lymphoma (BL) that could lead to better prevention, detection, and treatment of this rare and aggressive cancer. The Office of Cancer Genomics (OCG) at the National Cancer Institute (NCI) initiated BLGSP in collaboration with the Foundation for Burkitt Lymphoma Research(link is external). The molecular characterization data from Burkitt Lymphoma pa ...

Figure 3: Search Catalog Results Pane Card View

Card View Table View EXPORT

Results per Page: 10 Showing 1-9 of 9

Dataset ^	Cases	Samples	Resource	Primary Dataset Scope
Burkitt Lymphoma Genome Sequencing Project (BLGSP)	126	386	CGCI	Project
Cancer Genome Characterization Initiative (CGCI)	112	2016	GDC	Program

Figure 4: Search Catalog Results Pane Table View

The information in the results pane from the Card View can be sorted with the 'Sort By' filter (Figure 5). Filtered results are displayed in ascending order by default but can be changed by selecting the descending order icon. In the Table View, table columns can be sorted by clicking the column headers.

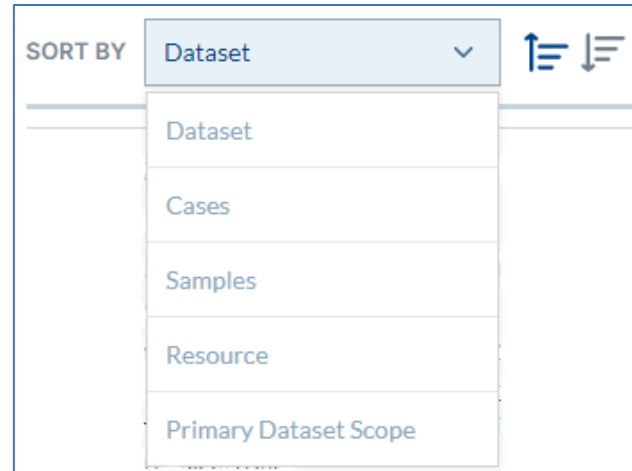


Figure 5: Search Catalog Sort By Filter

The result pane displays 10 search results by default. If a search returns more than 10 results, the subsequent results are available for review on subsequent pages. These pages can be accessed using a '<' or '>' buttons or a specific page number on the pagination bar at the top and bottom of the results pane (Figure 6). If returned results contain fewer than 10 results, the pagination buttons are disabled. User can select the 'Results per Page' dropdown to modify the number of displayed results.

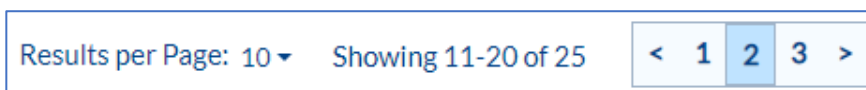


Figure 6: Search Catalog Pagination Bar

Card View

The Card View (Figure 3) is meant to provide a brief highlight of the datasets available in the catalog. Each card provides some abbreviated of selected metadata about a dataset for an initial assessment of relevancy. In the Card View, information for the following data elements is available for display:

- **Dataset Name**
- **Participating Resource Name**
- **Case Disease Diagnosis**
- **Case Count**
- **Sample Assay Method**
- **Sample Count**
- **Description**
- **Primary Dataset Scope**

If a dataset does not have information for a data element, the data element will not be displayed. The card view may also display 'Other Match' and an additional data element if information from the search text is contained within another dataset data element. 'Other Match' information will only appear when performing a search.

Table View

The Table View (Figure 4) is meant to provide a selection of information on the datasets available in the catalog. The Table View displays the following name and count data elements:

- **Dataset Name**

- **Participating Resource Name**
- **Case Count**
- **Sample Count**
- **Primary Dataset Scope**

Search

There is a text field at the top of the page (Figure 7) to search the catalog datasets. The search will perform a full and partial word search for any terms or text entered that meet the search parameters. The parameters, defined below, are also available on the site by selecting the '?' tooltip next to the search field.

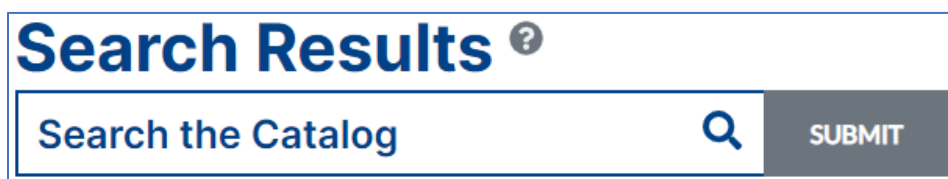


Figure 7: Search Catalog Search Field

Parameters for searching the catalog:

- A minimum of 3 characters are needed for a term search.
- Searches will return both full and partial word results (i.e., leuk will return results for leukemia).
- If you search multiple terms (i.e., lymphocytic survivors) the search results will return sources that contain all the specified terms (AND Boolean operator).
- Results can be filtered by Participating Resources by checking a resource in the resource column. Selecting multiple Resources will filter as an OR Boolean operator.
- Anatomical site searches leverage the NCI Thesaurus to display Case Tumor Site synonym matches. For example, a search for 'eye' will also return results for 'orbit.'

To execute a search, enter the text of interest and either select the 'Submit' button to the right of the search field or hit enter/return on your keyboard. This will execute a search and you should see a prompt with the searched text appearing to the right of the search field as a 'Search Text' bubble (Figure 8). With an executed search, the results pane will be updated to display the datasets that contain a match for the defined text. Additionally, the Resource column, to the left of the results pane, will update to display the participating resources that have datasets that match the search. To reset the search, you can select the 'x' within the Search Text bubble.



Figure 8: Search Catalog Valid Search Execution

The catalog search can also be filtered by selecting a resource (i.e., CCSS) from the Resource column (Figure 9). Selecting a Resource will automatically populate a prompt with the searched text appearing to the right of the search field as a 'Resources' bubble (Figure 9). A Resource filter can be applied alone

or in conjunction with a text search. To reset the search, you can select the 'x' within the Resource bubble or select the circular back arrow in the Resource column.

Search Results ?

Search the Catalog Resources: CCSS x

Resources

☐ CBTRUS

☐ CCDI

☒ CCSS

☐ CGCI

Card View Table View

Results per Page: 100 Showing 1-1 of 1 < 1 >

Dataset	Cases	Samples ^	Resource	Primary Dataset Scope
Childhood Cancer Survivor Study	25665		CCSS	Project

Figure 9: Search Catalog Resource Pane Filter

Export


The Search Catalog features an export that allows the results of a search to be exported for local manipulation or record keeping. To export results, there is an 'Export' link in the upper right of the results pane (Figure 3). This link will create an export of the results shown in the results pane. The file is exported as a comma-separated values (.csv) file that can be opened in a local text or spreadsheet tool.

The file contains all the information and data elements for a dataset that has been gathered for the CCDC. Each row is an individual dataset and includes information not found on the catalog search page. Please note that the information found on each row is unique to each dataset and that datasets, even within a single participating resource, may contain varying amounts of content per data element.

Dataset Detail Page

The Dataset Details page (Figure 10) provides detailed information about that dataset. The Dataset Details page is available from either the Search Catalog page or from a participating resources details page. The Dataset Details page contains a listing of all metadata the CCDC site has collected about a dataset. The page is broken up into two parts, a header with administrative information and a body that contains count and summary data.

The data on the Dataset Details page is only updated periodically and data on the CCDC site may not be as current as would be found on a participating resource's own data sharing site. For the most up to date information about a dataset, it is recommended that you verify any data of interest with the resource owner's website.



NATIONAL CANCER INSTITUTE

Childhood Cancer Data Initiative

Data Catalog

Documentation Search

Home

Search Catalog

Participating Resources

About

Home


Search Catalog

Childhood Cancer Survivor Study

Childhood Cancer Survivor Study

Data Resource: [CCSS](#)

Point of Contact: Gregory T. Armstrong, M.D., MSCE., greg.armstrong@stjude.org



PROJECT

ABOUT THIS DATASET

The Childhood Cancer Survivor Study (CCSS), a component of the Long-Term Follow Up Study that began in 1994 and was funded by a grant from the National Cancer Institute, includes all participants with a confirmed diagnosis of cancer and 5 year survival, a cohort of 35,923 childhood cancer survivors diagnosed between 1970 and 1999. It also includes over 5,000 siblings of survivors who serve as the comparison group for the study. The CCSS cohort has been assembled through the efforts of 31 participating centers in the United States and Canada. CCSS is a resource in which to investigate current and future questions regarding consequences of therapy, genetic associations, disease processes and causation, interventions, and quality of life among childhood cancer survivors. The Childhood Cancer Survivor Study (CCSS) includes all participants with a confirmed diagnosis of cancer and 5 year survival, a cohort of 35,923 childhood cancer survivors diagnosed between 1970 and 1999. The CCSS cohort has been assembled through the efforts of 31 participating centers in the United States and Canada. CCSS is a resource in which to investigate current and future questions regarding consequences of therapy, genetic associations, disease processes and causation, interventions, and quality of life among childhood cancer survivors. This dataset outlines the demographic, primary cancer and treatment characteristics of the Overall CCSS Cohort (the Expansion Cohort as of January 2021 plus the Original Cohort).

Core Data Elements

NUMBER OF CASES

25,665

Additional Data Elements

DATA REPOSITORY

<https://ccss.stjude.org/about-the-ccss.html>

Figure 10: Dataset Details Page


Administrative data

At the top of each Dataset Details page is an administrative header (Figure 11). The administrative header contains the name of the dataset, a link to the participating ‘Data Resource’ on the CCDC site to which the dataset belongs, and Point of Contact information for the dataset. When possible, the point of contact contains both a name and an email address for a resource owner.

Childhood Cancer Survivor Study

Data Resource: [CCSS](#)

Point of Contact: Gregory T. Armstrong, M.D., MSCE., greg.armstrong@stjude.org



PROJECT

Figure 11: Dataset Details Page Administrative Data

Metadata

The main body of the Dataset Details page contains the collected dataset metadata (Figure 12). In addition to the dataset description, the dataset metadata describes the 'Core Data Element' and the 'Additional Data Elements' of the dataset. These data elements are designed to provide a high-level overview of the data available within the dataset and provide supporting information about the provenance of the dataset.

ABOUT THIS DATASET	
<p>The Childhood Cancer Survivor Study (CCSS), a component of the Long-Term Follow Up Study that began in 1994 and was funded by a grant from the National Cancer Institute, includes all participants with a confirmed diagnosis of cancer and 5 year survival, a cohort of 35,923 childhood cancer survivors diagnosed between 1970 and 1999. It also includes over 5,000 siblings of survivors who serve as the comparison group for the study. The CCSS cohort has been assembled through the efforts of 31 participating centers in the United States and Canada. CCSS is a resource in which to investigate current and future questions regarding consequences of therapy, genetic associations, disease processes and causation, interventions, and quality of life among childhood cancer survivors. The Childhood Cancer Survivor Study (CCSS) includes all participants with a confirmed diagnosis of cancer and 5 year survival, a cohort of 35,923 childhood cancer survivors diagnosed between 1970 and 1999. The CCSS cohort has been assembled through the efforts of 31 participating centers in the United States and Canada. CCSS is a resource in which to investigate current and future questions regarding consequences of therapy, genetic associations, disease processes and causation, interventions, and quality of life among childhood cancer survivors. This dataset outlines the demographic, primary cancer and treatment characteristics of the Overall CCSS Cohort (the Expansion Cohort as of January 2021 plus the Original Cohort).</p>	
Core Data Elements	Additional Data Elements
<p>NUMBER OF CASES 25,665</p> <p>CASE SEX Female (11,940); Male (13,725)</p>	<p>DATA REPOSITORY https://ccss.stjude.org/about-the-ccss.html</p> <p>GRANT ID U24CA055727</p>

Figure 12: Dataset Details Page Metadata

For both Core and Additional Data Elements, if information was not available for a data element, that data element will not appear on the Dataset Details page. You should expect some variation in available data between datasets.

Core Data Elements

Core data elements describe and provide count information about data within the dataset. These data elements describe the participants, related disease information, and information about sample, if samples are part of the dataset. For each numeric data elements, such as Number of Cases, the numeric count value is of that element displayed. For response type data elements, such as Case Sex, the response is displayed along with the count for that response (i.e., Female (11,940)).

The following data elements are displayed as the core data element:

- **Number of Cases**
- **Case Sex**
- **Case Age at Diagnosis**
- **Case Race**
- **Case Ethnicity**
- **Case Disease Diagnosis**
- **Case Tumor Site**
- **Case Treatment Administered**
- **Case Treatment Outcome**

- **Number of Samples**
- **Sample Assay Method**
- **Sample Analyte Type**

Additional Data Elements

Additional Data Elements are generally the data elements that provide supporting information about a dataset such as grant identifier. However, some datasets do include additional information that describes the data within a dataset, such as the data element 'Has Mutation Data'.

The following are the standard data elements are displayed in the additional data element column:

- **Clinical Trial**
- **Data Repository**
- **Grant ID**
- **Grant Name**
- **Projects**
- **Project ID**
- **Published In**

Participating Resource Page

The CCDC Participating Resource page allows users to view and filter the resources that CCDC datasets are drawn from (Figure 13). The Participating Resources page can be navigated to by selecting

The screenshot displays the 'Participating Resources' page of the National Cancer Institute Childhood Cancer Data Initiative Data Catalog. The top navigation bar is green with links for 'Home', 'Search Catalog', 'Participating Resources' (highlighted with a yellow box), and 'About'. Below the navigation bar, the page title 'Participating Resources' is shown with '29 Results'. A banner image depicts a healthcare professional interacting with a child. On the left, a 'Research Description' sidebar lists filters for 'RESOURCE TYPE' (Catalog, Knowledgebase, Program, Registry, Repository) and 'DATA CONTENT TYPE' (Cell Lines, Clinical, Genomics/Omics, Imaging, Xenograft). The main content area features the 'Cancer Genome Characterization Initiative (CGCI)' with a description, a 'Point of Contact' link, and a 'DATASET SUMMARIES (2)' button.

Figure 13: Participation Resource Page

'Participating Resource' on the menu bar. The resource page includes introductory information for each resource available on the site along with a link to view a more detailed resource page. Resources on the page can be filtered by 'Resource Type' and by 'Data Content Type' to view specific resources.

The catalog currently includes the following types of resources.

- **Catalog** - A data catalog is not a data repository but rather a place where data is described with an index to what is available. A collection of digests and references (e.g., URL or POC) to corresponding research artifacts. There is a consistent structure across the collection of digests to facilitate filtering and identifying research artifacts of interest. A catalog contains some combination of Summary Research Data, Summary Clinical Data, Data Overview, and Resource Metadata.
- **Knowledgebase** - Biomedical knowledgebases extract, accumulate, organize, annotate, and link the growing body of information that is related to and relies on core datasets.
- **Program** - A coherent assembly of plans, project activities, and supporting resources contained within an administrative framework, the purpose of which is to implement an organization's mission or some specific program-related aspect of that mission.
- **Registry** - A cancer registry is an information system designed for the collection, storage, and management of data on persons with cancer. An inventory of individuals or samples, usually focused on a specific diagnosis or condition. In some cases, public health laws require collecting information in registries about individuals who have a specific disease or condition. In other cases, individuals provide information about themselves to these registries voluntarily. Thus, a registry contains Individual Clinical Data, but not Individual Research Data
- **Repository** - Biomedical data repositories accept submission of relevant data from the community to store, organize, validate, archive, preserve and distribute the data, in compliance with the FAIR Data Principles. A system for storing multiple research artifacts, provided at least some of the research artifacts contain Individual Research Data. A data repository often contains artifacts from multiple studies. Some data repositories accept research datasets irrespective of the structure of those datasets; other data repositories require all research datasets to conform to a standard reference model.

Resources available on the site are periodically updated and the types of resources available may change. A complete listing of current resources is available in Appendix A.

Resource Filtering

To filter the catalog resources there is a collapsible filter menu on left-hand side of the page (Figure 14). Within the filter menu there are two available filters— Resource Type and Data Content Type. For either filter, a selection can be made by checking the boxes to the left of the filter option. There is no limit on the number of filter selections that can be made.

When filtering, if selecting multiple options within a single filter the returned results will show using the OR operator for the selected options (Figure 14B). In this example, the results will display all catalog and program resources in the CCDC. If selecting multiple options across the two filters, the results will return with the AND operator for the selected options (Figure 14C). In this example, the results will display only catalogs with clinical data. To reset the filters either un-select any selected options or refresh the page.

Figure 14 consists of three panels, A, B, and C, each showing a 'Research Description' filter interface. The interface is divided into two sections: 'RESOURCE TYPE' and 'DATA CONTENT TYPE', each with a list of options and a selection checkbox.

- Panel A (No selections):** All checkboxes are unchecked. Under 'RESOURCE TYPE', the options are Catalog, Program, Registry, and Research Data Repository. Under 'DATA CONTENT TYPE', the options are Cell Lines, Clinical, Genomics/Omics, Imaging, and Xenograft.
- Panel B (Single filter Selection):** The 'Catalog' checkbox under 'RESOURCE TYPE' is checked. All other checkboxes are unchecked.
- Panel C (Multi-filter Selection):** The 'Catalog' checkbox under 'RESOURCE TYPE' and the 'Clinical' checkbox under 'DATA CONTENT TYPE' are checked. All other checkboxes are unchecked.

Figure 14: Participating Resource Research Filter. A – No selections. B – Single filter Selection. C – Multi-filter Selection

Resource Card View

Information about participating resources is displayed in the results pane on the right-hand side of the Participating Resource page (Figure 13). The resources in the pane are displayed in a Card View. Within the resource card, select a resource's name to navigate to a page with more detailed information about the resource and its associated datasets.

The result pane displays 10 search results by default. If a filter returns more than 10 results, the subsequent results are available for review on subsequent pages. These pages can be accessed using a '<' or '>' buttons or a specific page number on the pagination bar at the top and bottom of the results pane (Figure 13). If returned results contain fewer than 10 results, the pagination buttons are disabled. User can select the 'Results per Page' dropdown to modify the number of displayed results.

Card View

The resource card view (Figure 14) is meant to provide a brief highlight of each resource available in the catalog. Each card provides a summary the resource, information about connecting directly with a resource, and information about the number of datasets from a resource listed on the CCDC site. Each card displays the following data elements for each resource:

- **Resource Name**
- **About this Resource**
- **Point of Contact Name**
- **Resource URL**
- **Resource Type**

- **Number of available Datasets**

Each resource card links to a page with additional detailed information about a participating resource. Selecting the 'Resource Name' will navigate to the top of the Participating Resource Details page. Selecting the 'Dataset Summaries' will navigate directly to the dataset summaries section of the Participating Resource Details page.

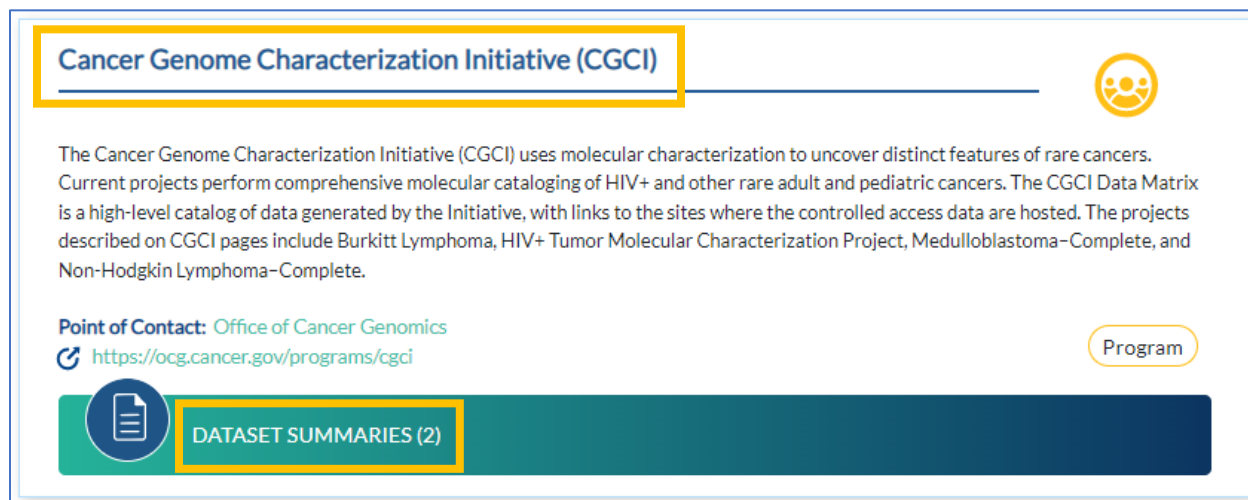
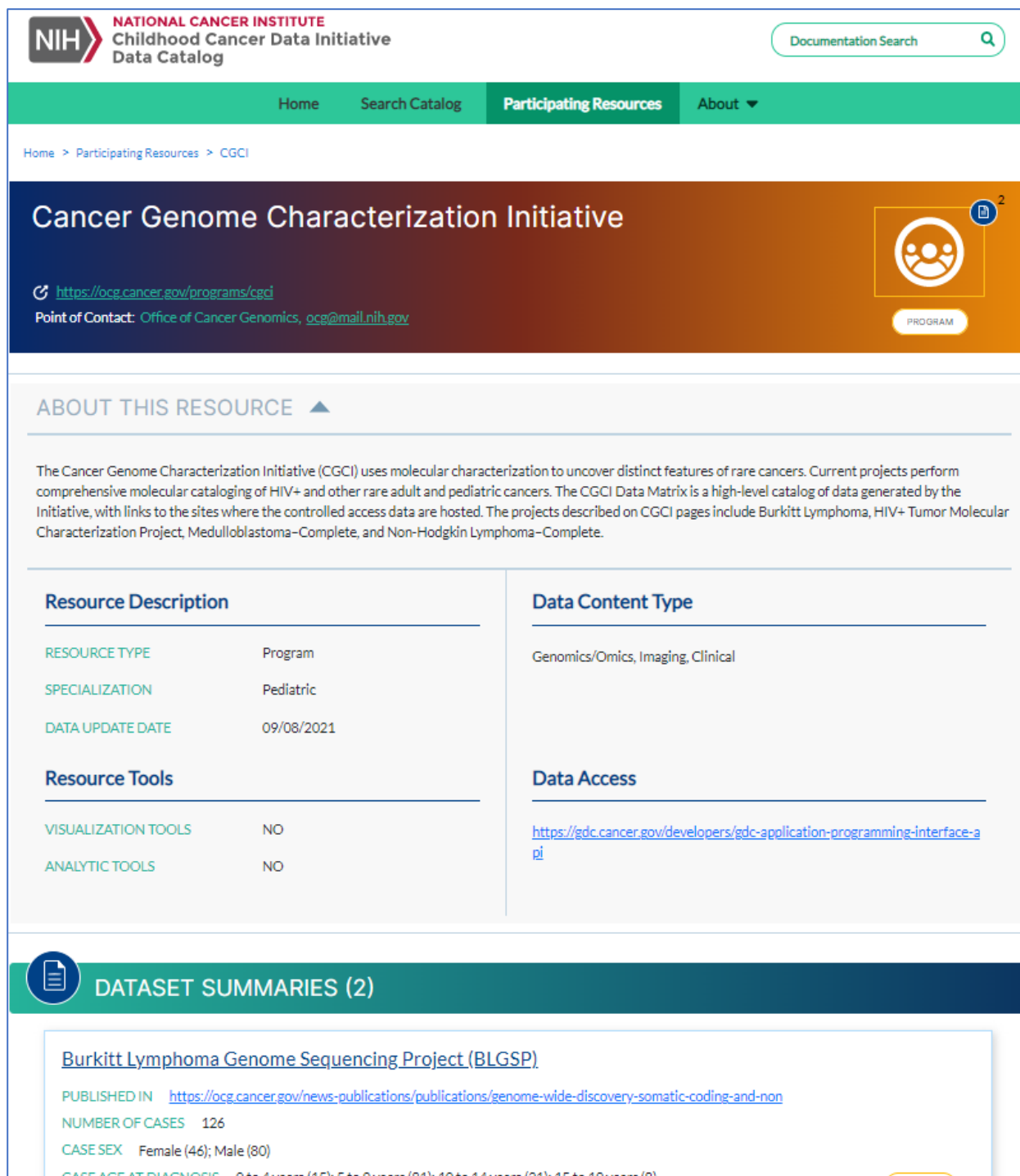


Figure 15: Participating Resource Card View

Participating Resource Detail Page

For each participating resource within the CCDC, there is a page of detailed information about that resource, named the Participating Resource Details page (Figure 16). The resource details can be navigated to by selecting a resource on either the Participating Resource page, the Search Catalog page, or the Dataset Details page. The Participating Resource Details page contains a listing of metadata the CCDC site has collected about a resource. The page is broken up into three parts, a header with

administrative information, a body section of descriptive information about the resource, and a body section listing the datasets associated with a resource on the CCDC site.



The screenshot shows the NIH National Cancer Institute Childhood Cancer Data Initiative Data Catalog. The header includes the NIH logo, the site name, and a search bar. The navigation bar has links for Home, Search Catalog, Participating Resources, and About. The breadcrumb trail shows Home > Participating Resources > CGCI.

Cancer Genome Characterization Initiative

<https://ocg.cancer.gov/programs/cgci>
Point of Contact: Office of Cancer Genomics, ocg@mail.nih.gov

ABOUT THIS RESOURCE

The Cancer Genome Characterization Initiative (CGCI) uses molecular characterization to uncover distinct features of rare cancers. Current projects perform comprehensive molecular cataloging of HIV+ and other rare adult and pediatric cancers. The CGCI Data Matrix is a high-level catalog of data generated by the Initiative, with links to the sites where the controlled access data are hosted. The projects described on CGCI pages include Burkitt Lymphoma, HIV+ Tumor Molecular Characterization Project, Medulloblastoma-Complete, and Non-Hodgkin Lymphoma-Complete.

Resource Description		Data Content Type
RESOURCE TYPE	Program	Genomics/Omics, Imaging, Clinical
SPECIALIZATION	Pediatric	
DATA UPDATE DATE	09/08/2021	

Resource Tools		Data Access
VISUALIZATION TOOLS	NO	https://gdc.cancer.gov/developers/gdc-application-programming-interface-api
ANALYTIC TOOLS	NO	

DATASET SUMMARIES (2)

Burkitt Lymphoma Genome Sequencing Project (BLGSP)

PUBLISHED IN <https://ocg.cancer.gov/news-publications/publications/genome-wide-discovery-somatic-coding-and-non-coding-mutations-in-burkitt-lymphoma>

NUMBER OF CASES 126

CASE SEX Female (46); Male (80)

CASE AGE AT DIAGNOSIS 0 to 4 years (15); 5 to 9 years (81); 10 to 14 years (21); 15 to 19 years (8)

Figure 16: Participating Resource Detail Page

The data on the Participating Resource Details page is only updated periodically and data on the CCDC site may not be as current as would be found on a participating resource's own data sharing site. For the most up to date information, it is recommended that you verify the data with a resource owner.

Administrative data

At the top of each Participating Resource Details page is an administrative header (Figure 17). The administrative header contains the name of the resources, a link to the resource's own data sharing website, Point of Contact information for the dataset, and the resource type defined by the CCDC. Point of contact information contains both a name and an email address for a resource owner whenever possible. The header also contains an icon in the upper providing a quick indication of the number of datasets associated with a resource.

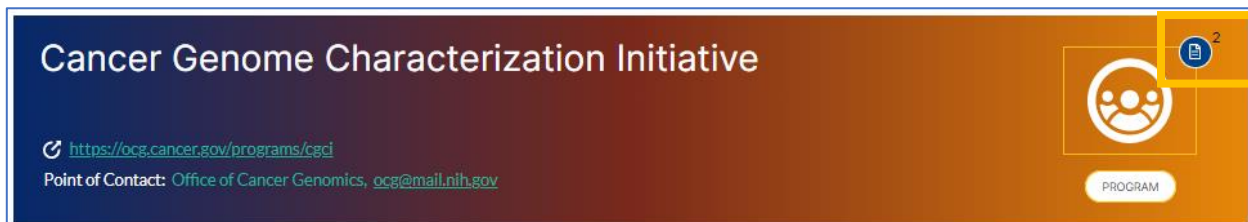


Figure 17: Participation Resource Details Administrative Data

Descriptive Data

The first section of the body of the Participating Resource Details page contains descriptive information about the resource (Figure 18). In addition to the 'About This Resource' description there are four sections of additional information – Resource Description, Resource Tools, Data Content Type, and Data Access.

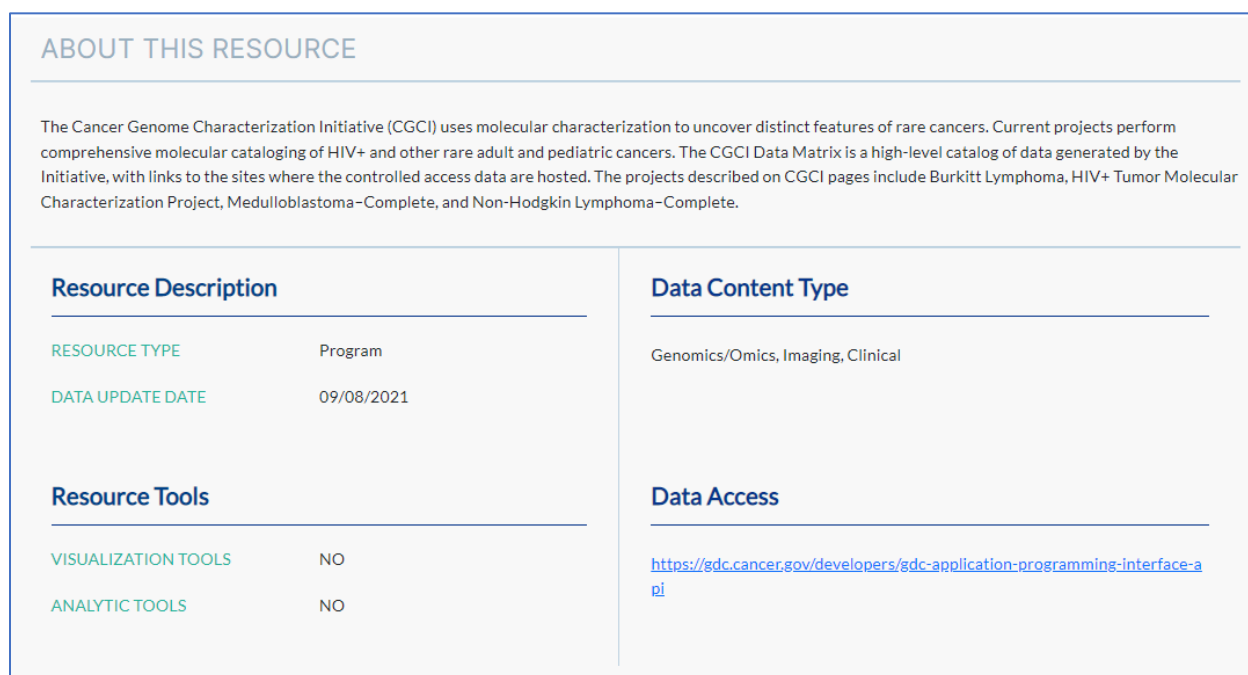


Figure 18: Participating Resource Details Descriptive Data

Resource Description

The Resource Description provides information about how the CCDC organizes the resource and when information on the CCDC site was last updated.

- **Resource Type** – CCDC categorization
- **Data Update Date** – Date the resource was last updated on the CCDC

Resource Tools

The Resource Tools provides information on if a participating resource's own data sharing site includes tools for additional dataset assessment. The CCDC currently identifies these types to assessment tools:

- **Visualization Tools** – A tool for visualization of available data
- **Analytic Tools** – A tool for analytic manipulation of available data

Data Content Type

The Data Content Type defines the type of data within a resource. The CCDC currently identifies these following content types:

- **Cell Lines**
- **Clinical**
- **Genomics/Omics**
- **Imaging**
- **Xenograft**

Data Access

The Data access section provides a link to the resource's own data sharing site where information or datasets are available.

Dataset Summaries

The second section of the body of the Participating Resource Details page contains a listing of the datasets, within the CCDC, associated with a participating resource (Figure 19). All participating resources will have at least one associated dataset and the number of associated datasets is identified within the section header. Each dataset is listed on a card view within the summary.

The screenshot displays a section titled "DATASET SUMMARIES (2)" with a document icon. It contains two dataset cards. The first card is for the "Burkitt Lymphoma Genome Sequencing Project (BLGSP)" and includes the following information: "PUBLISHED IN" with a link to <https://ocg.cancer.gov/news-publications/publications/genome-wide-discovery-somatic-coding-and-non>, "NUMBER OF CASES" as 126, "CASE SEX" as Female (46); Male (80), and "CASE AGE AT DIAGNOSIS" as 0 to 4 years (15); 5 to 9 years (81); 10 to 14 years (21); 15 to 19 years (8). A "PROJECT" button is located at the bottom right of the card. The second card is for the "Non-Hodgkin Lymphoma Project" and includes: "PUBLISHED IN" with a link to https://ocg.cancer.gov/programs/cgci/publications?field_publication_project_target_id_selective=33&items_per_page=10, "NUMBER OF CASES" as 6, "CASE SEX" as Female (1); Male (5), and "CASE AGE AT DIAGNOSIS" as 15 to 19 years (2); 20 to 24 years (2); 30 to 34 years (1); 35 to 39 years (1). A "PROJECT" button is also at the bottom right of this card.

Dataset Name	Published In	Number of Cases	Case Sex	Case Age at Diagnosis	Action
Burkitt Lymphoma Genome Sequencing Project (BLGSP)	https://ocg.cancer.gov/news-publications/publications/genome-wide-discovery-somatic-coding-and-non	126	Female (46); Male (80)	0 to 4 years (15); 5 to 9 years (81); 10 to 14 years (21); 15 to 19 years (8)	PROJECT
Non-Hodgkin Lymphoma Project	https://ocg.cancer.gov/programs/cgci/publications?field_publication_project_target_id_selective=33&items_per_page=10	6	Female (1); Male (5)	15 to 19 years (2); 20 to 24 years (2); 30 to 34 years (1); 35 to 39 years (1)	PROJECT

Figure 19: Participating Resource Details, Dataset Summaries

As the Participating Resource Details page already provides background information about the project, the individual dataset cards are more concise and quantitative than those found on the Search Catalog page. The dataset card views display the following data elements for each dataset:

- **Dataset Name**
- **Published In**
- **Number of Cases**
- **Primary Dataset Scope**

If available, the card view also displays:

- **Case Sex**
- **Case Age at Diagnosis**

Selecting the Dataset Name will navigate to the datasets Dataset Details page where you can learn more about a dataset. Datasets available on the site are periodically updated and the types of datasets available may change.

About CCDI Data Catalog Page

The CCDC About CCDI Data Catalog page describes the CCDC and provides information on engaging with CCDI and contributing to the CCDC (Figure 20). The About page can be reached by navigating to the 'About CCDI Data Catalog' submenu item within 'About' dropdown on the menu bar. Once on the page users can review information about the CCDC and how it fits within the CCDI. The About page also provides a link to join the CCDI mailing list. Joining the mailing list does not affect a user's experience on the CCDC site, but provides future news, content updates, and other information from the CCDI.



Figure 20: About CCDI Data Catalog Page

Contributing to the CCDI Data Catalog

The CCDC team is continually looking to expand and improve the inventory of pediatric oncology data resources on the CCDC site. Users with knowledge of additional resources are invited to share that

knowledge with the CCDI team and contribute to the advancement of the CCDC. Users with comments are invited to contact the CCDI directly using the email link provided on the About page, using the Contact Us link in the site footer, or directly emailing NCIChildhoodCancerDataInitiative@mail.nih.gov.

Users willing to share a new resource are invited to download the summary submission template. This template is the tool used to add new information. Instructions on filling out this template are available by contacting the NCI's Office of Data Sharing (ODS) at NCIChildhoodCancerDataInitiative@mail.nih.gov. Completed templates can be submitted to the CCDI.

Please note, all submitted templates will be reviewed by the CCDC team. Any accepted data submission or suggested changes will be incorporated as part of periodic releases of the CCDC site.

Glossary Page

The CCDC Glossary page provides a listing of terms and abbreviations used on the CCDC site (Figure 21). To reach the Glossary page users navigate to the 'Glossary' submenu item within 'About' dropdown on the menu bar.

The glossary page has two main components, the alphabetical ribbon for navigation and the term pane where information is listed and defined. The information in the glossary is arranged alphabetically and can be navigated by selecting letters on the alphabetic ribbon. Letters on the ribbon will appear in one of three ways. If a letter appears green with an underline it indicates that the letter is selected. Terms from a selected letter will appear in the term pane. If a letter appears black, then that letter contains glossary terms but is not selected. Black letters can be selected to update the term pane with terms of the newly selected letter. If a letter appears grey, there are no terms defined for that letter.

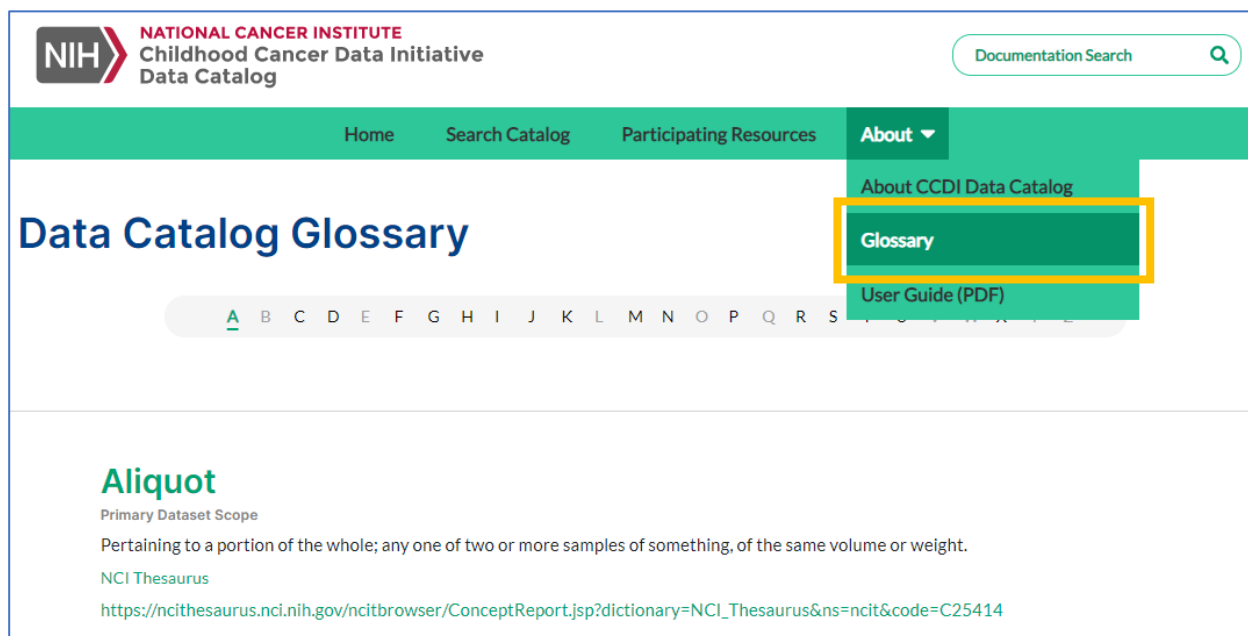


Figure 21: Data Catalog Glossary Page

Each glossary term will always have the following components:

- **Term** - The term being defined
- **Term Category** - A grouping of like terms by the CCDC

- **Definition** - The definition of the term

Terms may also have the following:

- **Reference** - The name of an external definition source if a definition is not defined by the CCDC team.
- **Reference URL** - A URL link to the external definition

User Guide

The CCDC User Guide link provides a link to a PDF version of the site user guide (Figure 22). Selecting the user guide link should open the guide in a new tab. Please note, for Safari user, clicking on the "User Guide" menu item will directly download the PDF format of the user guide.

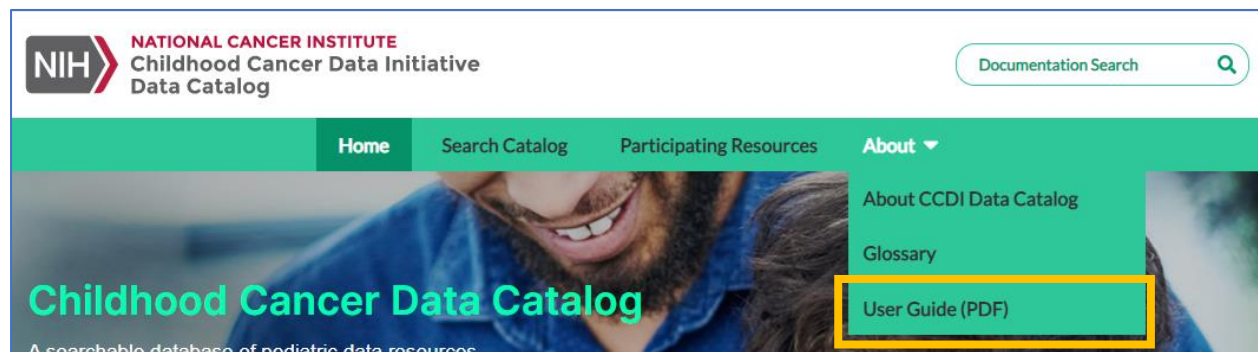


Figure 22: CCDC User Guide Link

Documentation Search Page

The Document Search field is a free-text search tool that searches across three CCDC site pages - Home page, About CCDI Data Catalog, and the Glossary. The search, located on the right-hand side of the CCDC site header, is accessible on any CCDC site page (Figure 23). To use the global search, type in a term or text of interest and press enter or click on the magnifier icon. A successful execution will take you to the documentation search page (Figure 24). This page is only accessible when using the Documentation Search field.

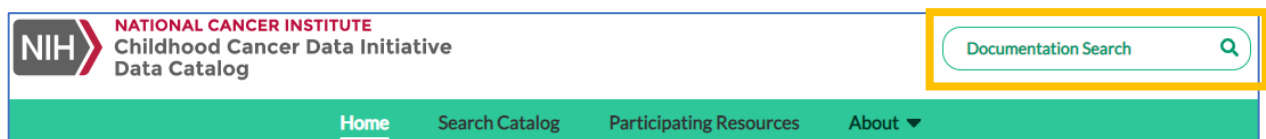


Figure 23: Documentation Search Field

On the Documentation Search page, searches can be refined or changed using the search bar in the header of the page (Figure 24).

NIH NATIONAL CANCER INSTITUTE
Childhood Cancer Data Initiative
Data Catalog

Home Search Catalog Participating Resources About ▼

Documentation Search

Type to search

3 Results

[About CCDC](#)

The Childhood Cancer Data Catalog is part of NCI's Childhood Cancer Data Initiative (CCDI) , which is building a community centered around childhood cancer care and research data. Through enhanced data sharing, we can improve understanding of cancer biology, preventive measures, treatment, quality of life, and survivorship, as well as ensure that researchers learn from every child with cancer. Sign-up for email updates from NCI about CCDI.

<https://ccdc-qa.bento-tools.org/about>

[CCDC Glossary](#)

Figure 24: Documentation Search Page

Parameters for the Document Search:

- A minimum of 3 characters are needed for a term search.
- Searches will return both full and partial word results (i.e., “can” returns results for cancer).
- If you search multiple terms (i.e., “Cancer Institute”) the search result returns sources that contain all the specified terms (AND Boolean operator).

The result of a search appears in the results pane on the main body of the page (Figure 25). Results only return if a static page includes the searched term. For each returned entry, the search displays the page title, the first block of page text, and the URL of the page. The number of returned results will be indicated on the pagination bar at the bottom of the results page.

3 Results

About CCDC

The Childhood Cancer Data Catalog is part of NCI's Childhood Cancer Data Initiative (CCDI) , which is building a community centered around childhood cancer care and research data. Through enhanced data sharing, we can improve understanding of cancer biology, preventive measures, treatment, quality of life, and survivorship, as well as ensure that researchers learn from every child with cancer. Sign-up for email updates from NCI about CCDI.

<https://datacatalog-dev.ccdi.cancer.gov/about>

CCDC Glossary

Describe CCDC Terms, definition and source of the definition.

<https://datacatalog-dev.ccdi.cancer.gov/glossary>

CCDI Data Catalog

The CCDI Childhood Cancer Data Catalog is a searchable database of National Cancer Institute and other pediatric cancer resources.

<https://datacatalog-dev.ccdi.cancer.gov/>

Results per Page: 10 ▾ Showing 1-3 of 3 < 1 >

Figure 25: Documentation Search Page Results Pane

Additional Support

Users requiring any additional support or with any questions about the CCDC can contact Childhood Cancer Data Initiative at NCIChildhoodCancerDataInitiative@mail.nih.gov. Please include a name and a reference to the CCDC with submitted any inquiry.

Appendix A: Participating Resources

Information as of 5/12/2022.

Participation Resource	Resource Type	CCDC Update Date
Children's Oncology Group	Catalog	12/6/2021
Human Cancer Model Initiative	Catalog	4/21/2022
Pediatric Genomic Data Inventory	Catalog	7/8/2021
Clinical Interpretations of Variants in Cancer (CIViC)	Knowledgebase	3/1/2011
OncoKB (OncoKB)	Knowledgebase	3/2/2022
Cancer Genome Characterization Initiative (CGCI)	Program	9/8/2021
Center for Pediatric Tumor Cell Atlas (HTAN)	Program	5/25/2021
Childhood Cancer Survivor Study	Program	5/25/2021
MyPART	Program	12/13/2021
Pediatric MATCH	Program	12/13/2021
Pediatric Preclinical Testing Consortium	Program	11/1/2021
Texas KidsCanSeq	Program	12/15/2021
Therapeutically Applicable Research to Generate Effective Treatments (TARGET)	Program	11/17/2021
Central Brain Tumor Registry of the United States (CBTRUS)	Registry	3/18/2022
National Childhood Cancer Registry Explorer	Registry	11/16/2021
Genomic Data Commons	Repository	9/3/2021
Kids First	Repository	12/3/2021
Oncogenomics	Repository	12/13/2021
Patient-Derived Xenograft and Advanced In Vivo Models (PDX-AIM)	Repository	5/11/2021
PDX Development and Trial Centers Research Network (PDXNet)	Repository	4/27/2022
PDX for Childhood Cancer Therapeutics (PCAT)	Repository	5/25/2021
PedcBioPortal for Integrated Childhood Cancer Genomics (PEDC)	Repository	9/14/2021
Pediatric Cancer Data Commons (PCDC)	Repository	7/8/2021
Sequence Read Archive	Repository	12/15/2021
St. Jude Cloud	Repository	5/9/2022
The Cancer Imaging Archive	Repository	5/4/2022
The Jackson Laboratory PDX Models (JAX PDX)	Repository	5/2/2022
Treehouse Childhood Cancer Initiative	Repository	9/10/2021