CHILDHOOD CANCER CLINICAL DATA COMMONS (C3DC)

User Guide

03/05/2025

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
5.00	03/05/2025	Release 5 (version 1.4.0)	C3DC Team

Table of Contents

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 Home Page allows users to navigate through key sections, including Explore, Cohort Analyzer, Data Model, and About pages.

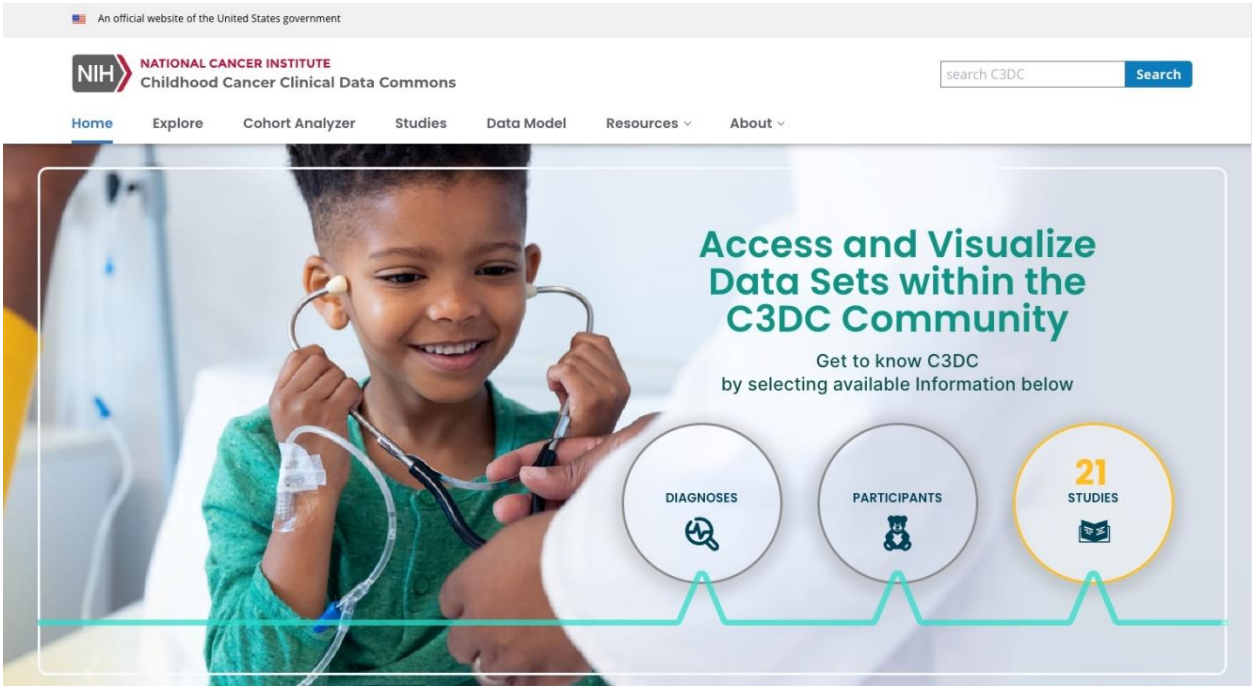

Figure 1: C3DC Home Page (top)

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

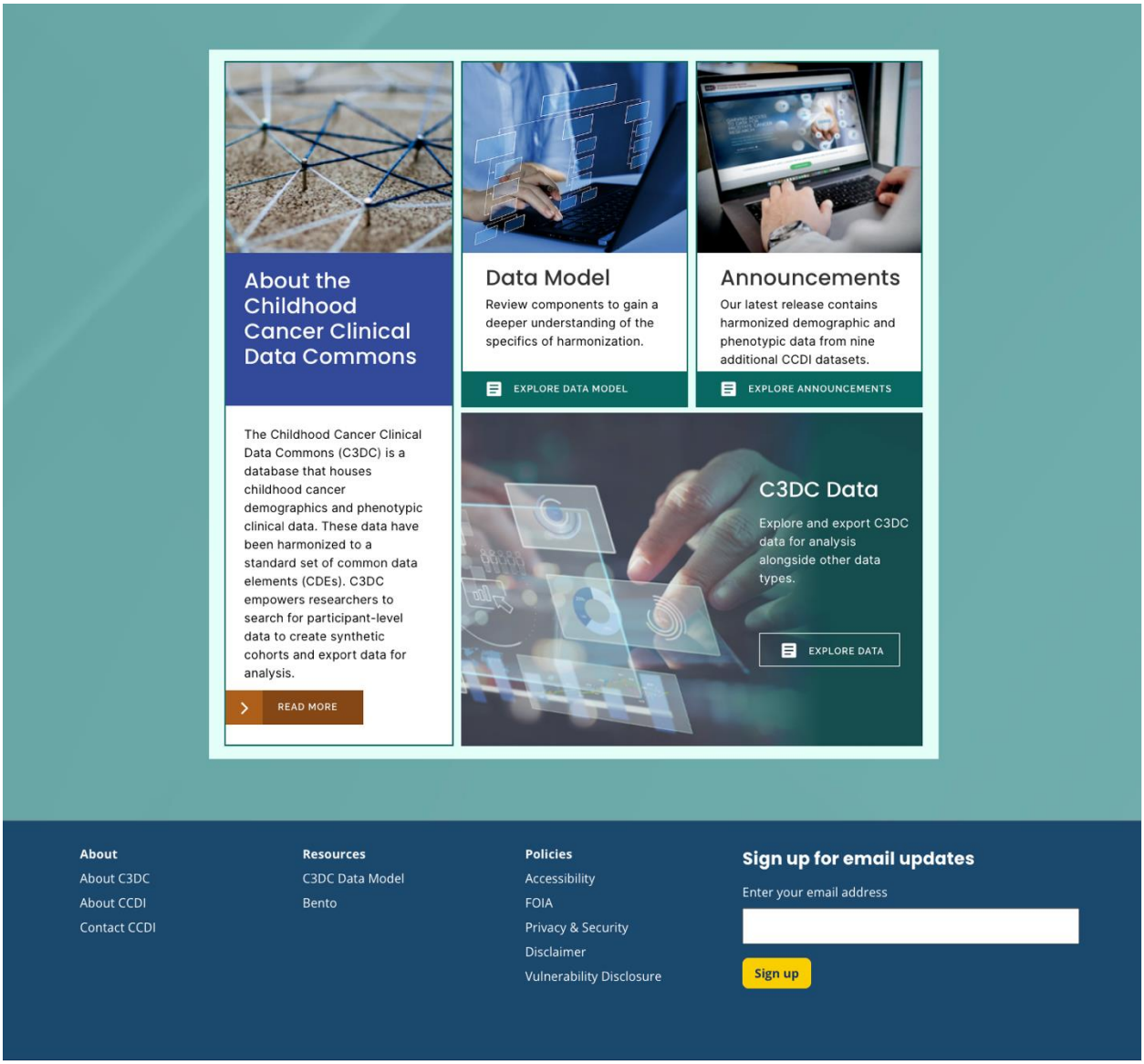

Figure 2: C3DC Home Page (bottom)

C3DC Explore Page

The Explore Page is the main interface for searching and visualizing data. Users can refine results using faceted search options and analyze participant-level details.

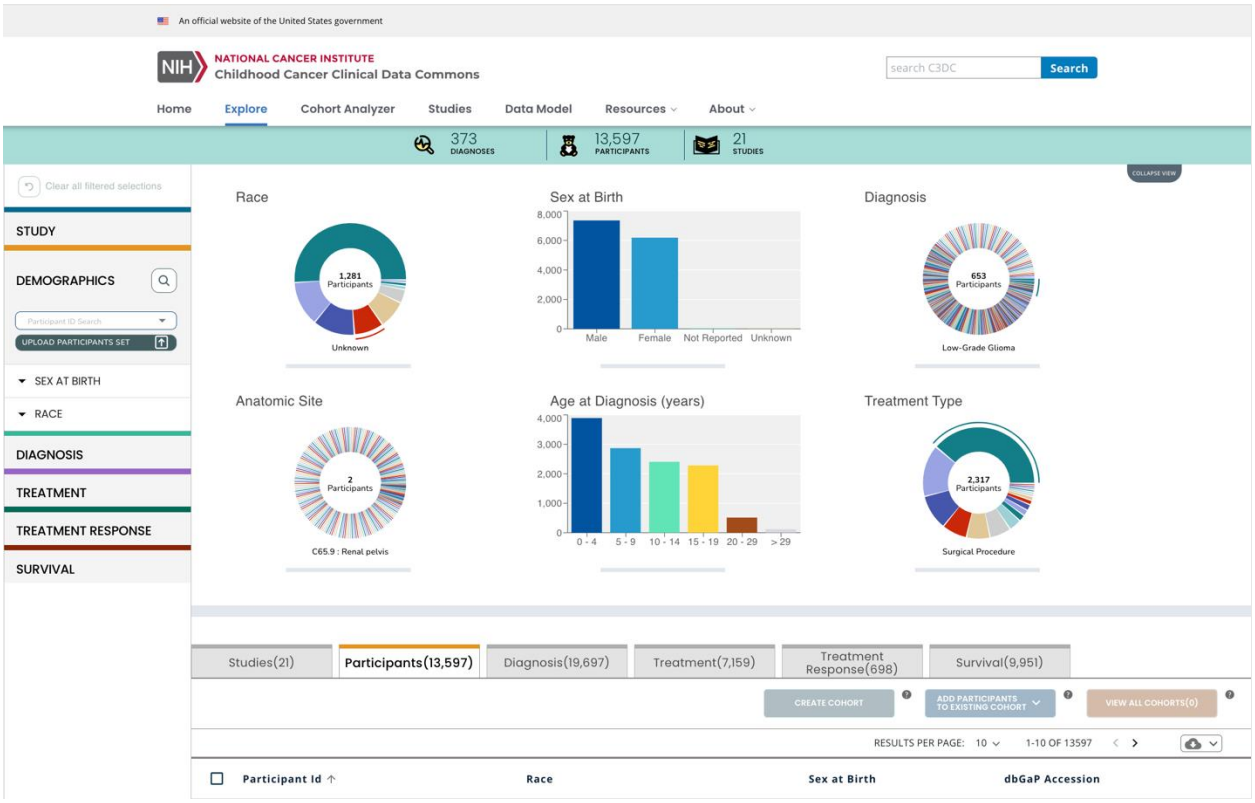

Figure 3: The Explore Page – Faceted Search (left), Visualization (top), and Data Tables (bottom)

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

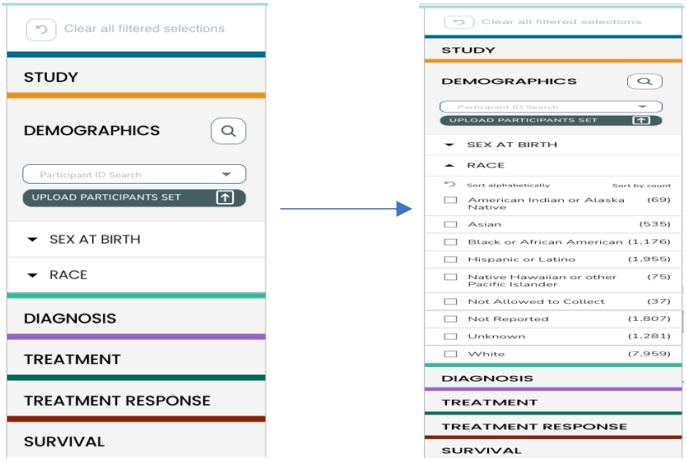

Figure 4: Performing a faceted search on the C3DC Explore Page

Chart Visualization: The **Stats Bar**, **Visualization Section**, and **Data Table** dynamically update based on applied filters.

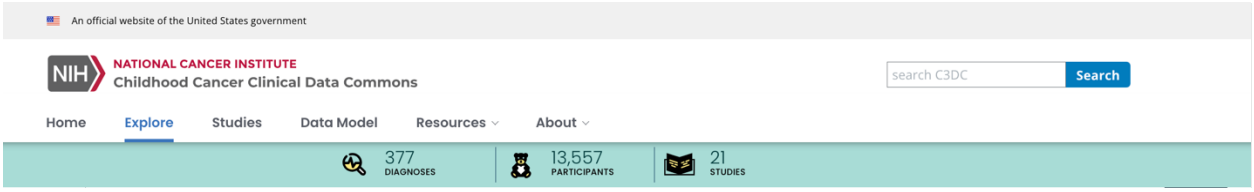

Figure 5: Results returned from a faceted search in the Stats bar

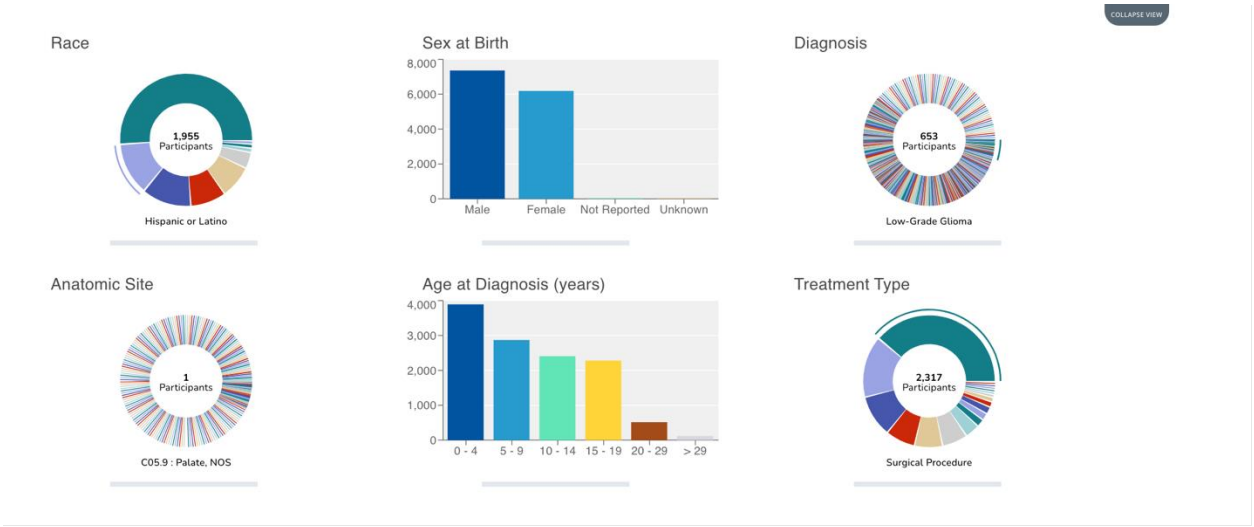

Figure 6: Results returned from a faceted search in six graphs in the Visualization section

Table Visualization: Displays participant details with **tooltips explaining table headers**.

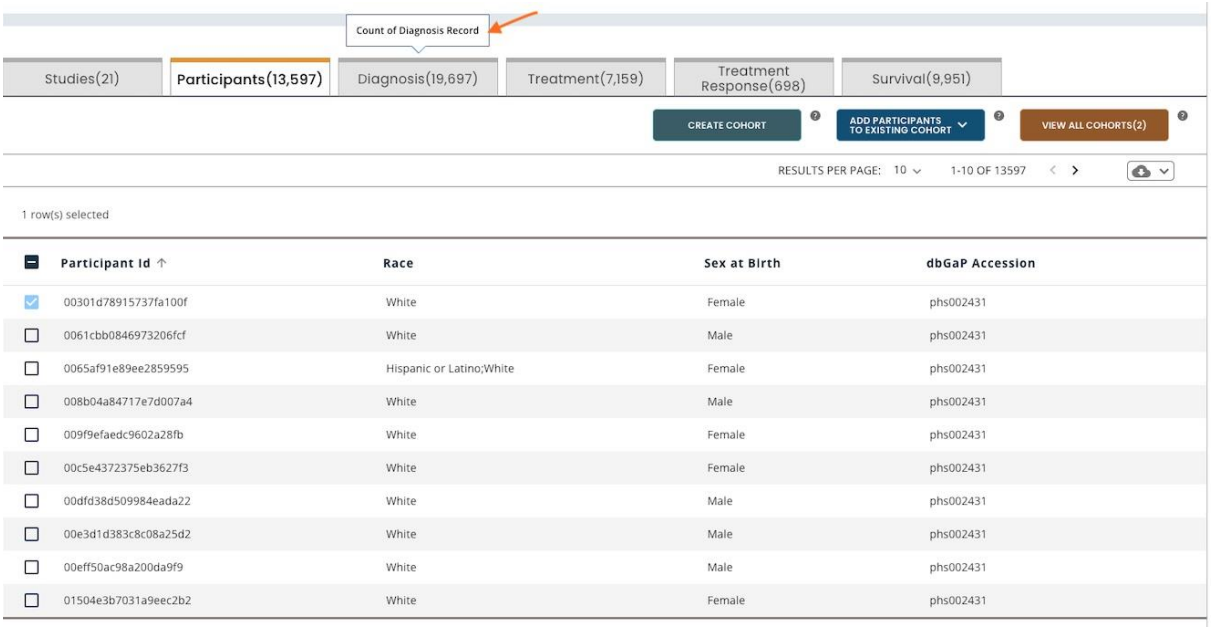

Figure 7: Results returned from a faceted search in the Table section

Download Harmonized Data: Users can download the contents of the Studies, Participants, Diagnosis, Treatment, Treatment Response, and Survival tabs by selecting the "Download Data" button under the table tab headers. Users can download filtered data in **CSV or JSON formats**.

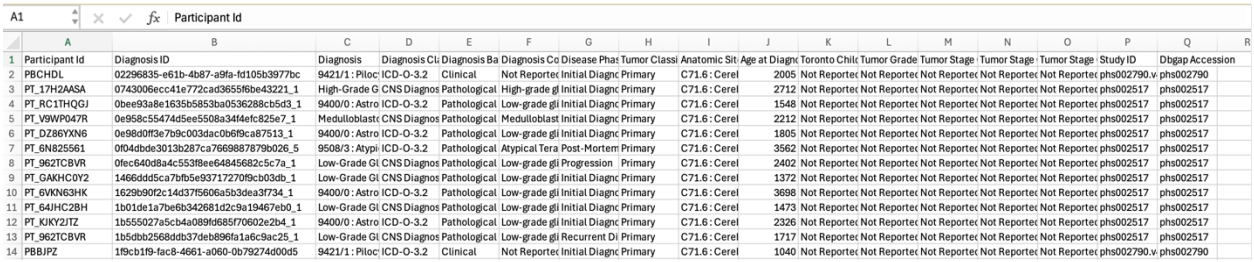

Figure 8: Download Diagnosis Table tab as CSV file format

Cohort Selector: The Cohort Selector enables users to create and manage up to 20 cohorts. This feature offers flexibility to researchers, allowing them to create cohort groups according to their specific requirements.

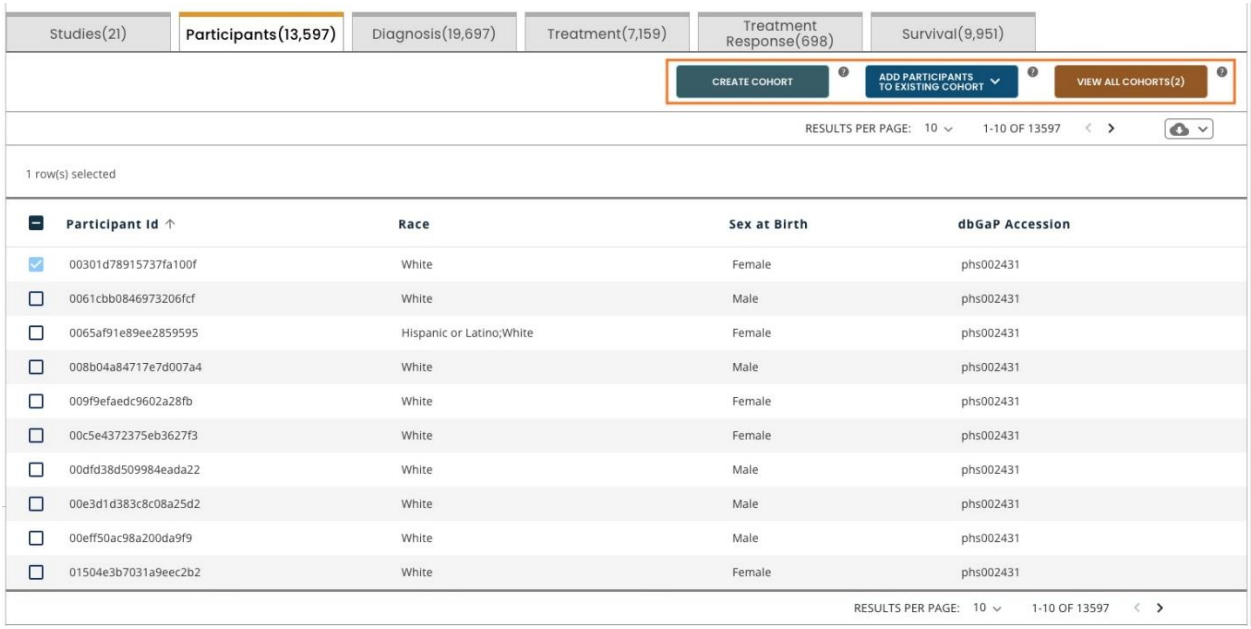

Figure 9: Cohort Selection features visible on the Explore page Table tabs (see choice buttons in orange rectangle)

Users can do the following:

- Create Cohort: Select participant IDs from the facet search and create a new cohort. Users can name and describe the cohort for easy reference.
- Add Participants to Existing Cohort: Add participants to existing cohorts or remove them. Entire cohorts can also be deleted as needed.
- View All Cohort(s): 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.

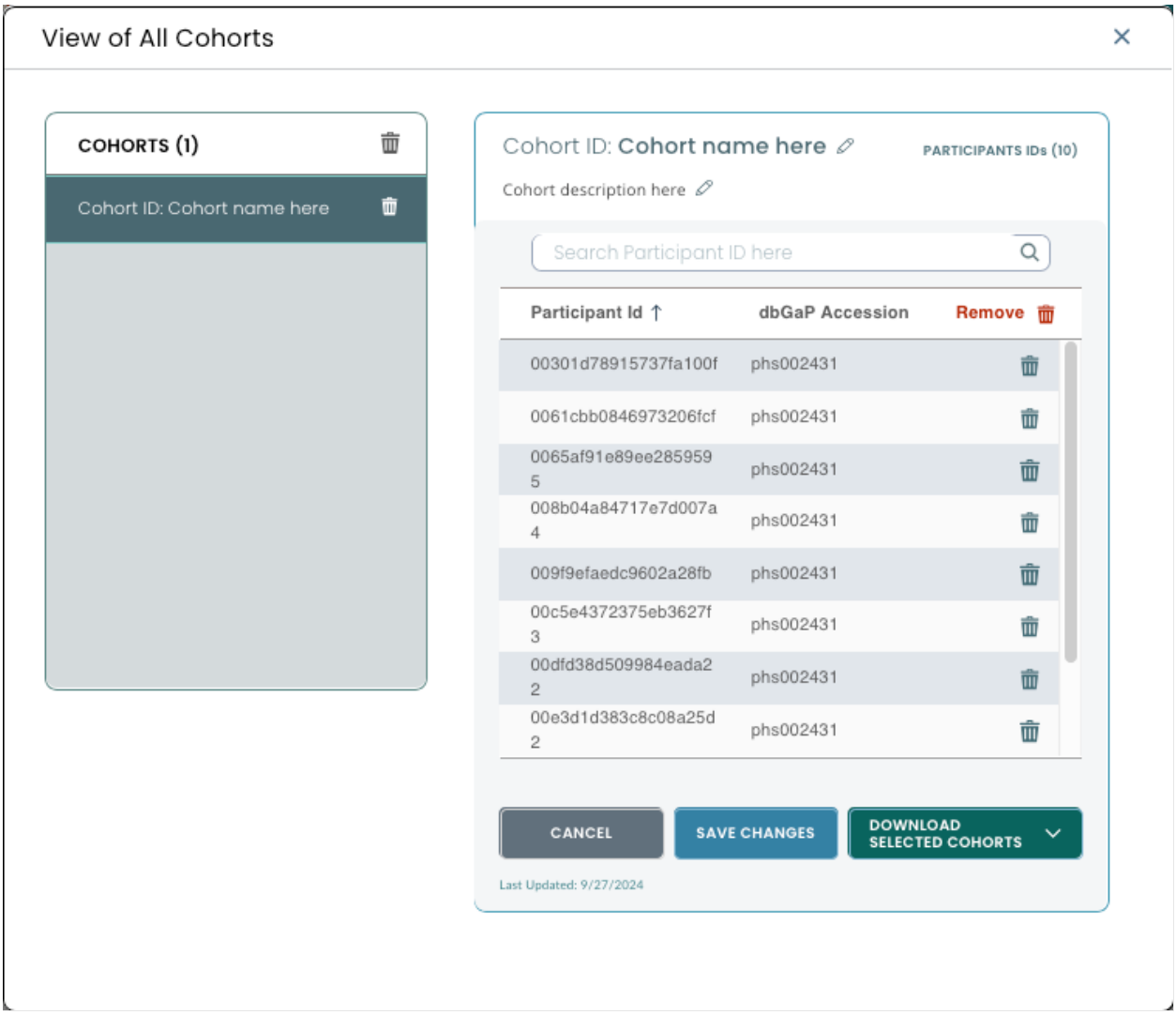

Figure 10: View All Cohorts and edit capabilities

C3DC Cohort Analyzer (New Key Feature)

The Cohort Analyzer is designed to compare up to three cohorts and visualize their intersections through an interactive Venn diagram and a corresponding data table. This feature leverages cohorts created on the Explore page, enabling users to analyze key relationships and distinctions based on unique values between datasets effectively.

Customizable Properties:

The radio buttons allow users to select more than one property for comparison. Available properties include:

- Participant ID
- Diagnosis
- Treatment

Enhanced Analytical Capabilities:

Users will be able to visualize overlaps and unique attributes within each cohort. In addition, users can:

- Investigate specific sections of the Venn diagram to view participant-level details from the corresponding table view
- Export results, including the data table and Venn diagram, for further analysis or integration into other platforms.
- Use advanced filters to refine cohort comparisons, such as narrowing by treatment or specific diagnosis.

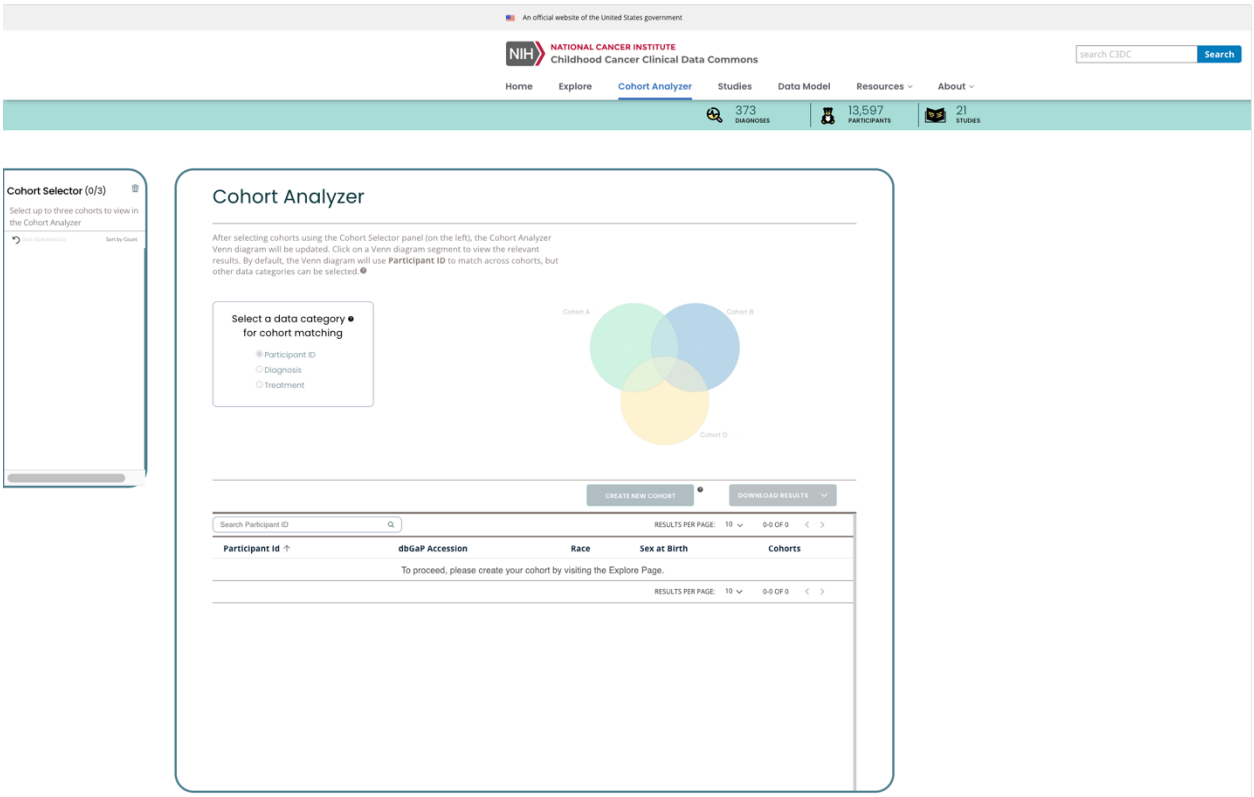

Figure 11: Cohort Analyzer landing page

To start using the Cohort Analyzer, you will first need to select the cohorts you want to analyze. As you add cohorts, the system will automatically keep track of your cohorts on the left side Cohort Selector. This tool's functionality adapts based on the number of selected cohorts, ensuring a customized analysis.

Select your first cohort by clicking the check box in the Cohort Selector sidebar. The Venn diagram and table will update to display the cohort information based on the participant ID, diagnosis or treatment based on the radio button selection. In this example, we are using the participant ID, thus the table will contain properties specific to the participant as well as show to what cohorts the participant belongs.

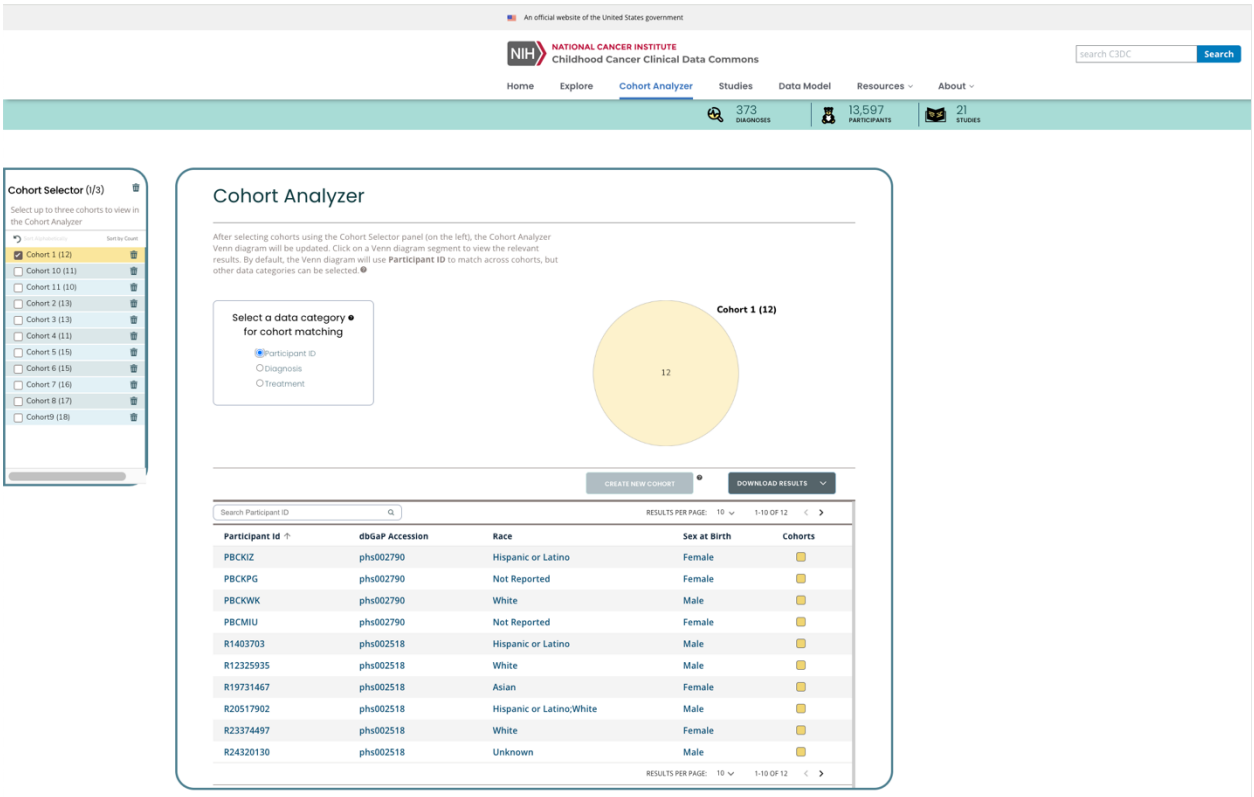

Figure 12: One Cohort Selected: You have created and selected one cohort.

Select another cohort in the Cohort Selector to see the Venn diagram and table update again. This time, if there are common participants between both cohorts, the diagram will show the shared participants in the intersection between the two. In the table, with none of the Venn diagram selected, it will display all participants and their respective cohort.

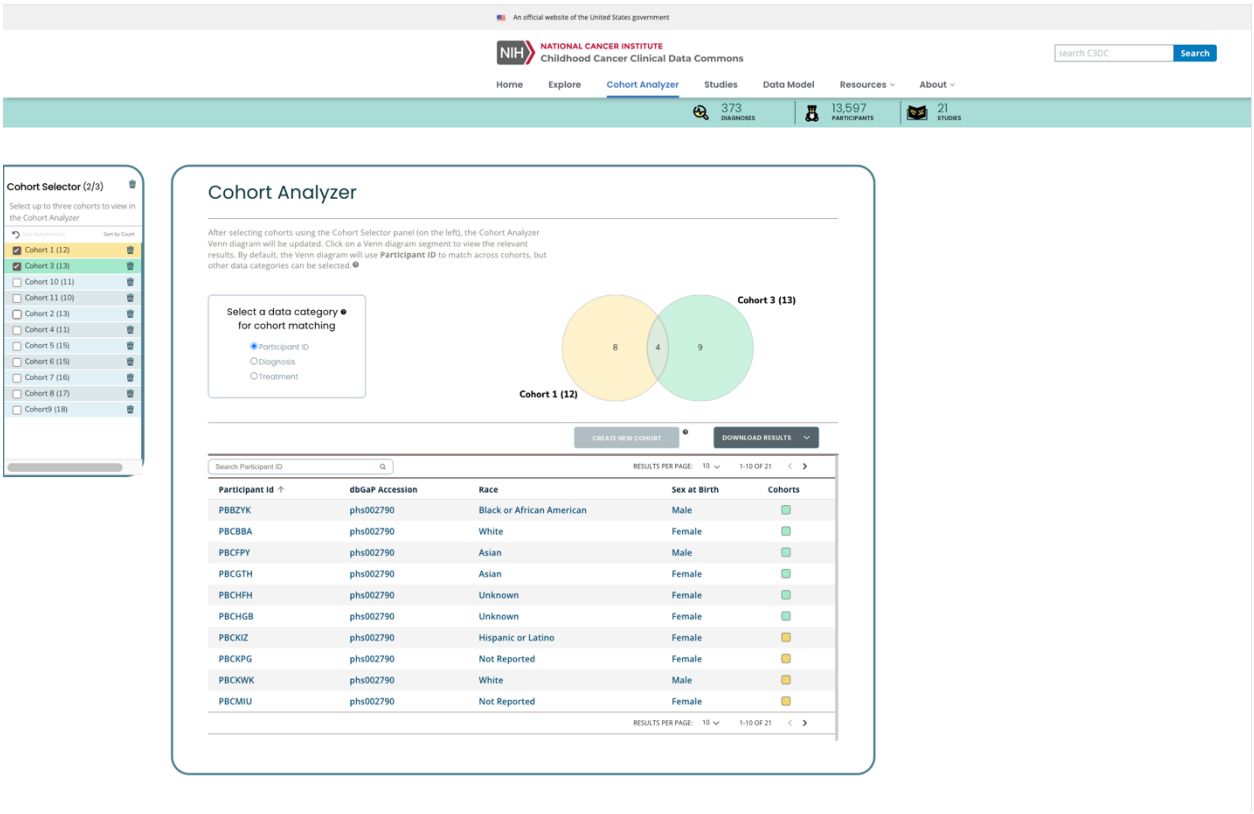

Figure 13: Two Cohorts Selected: You have selected two cohorts. Visualize shared and unique data points between these cohorts.

Select a third and final cohort. The Venn diagram and table will update once more with all the participant level data.

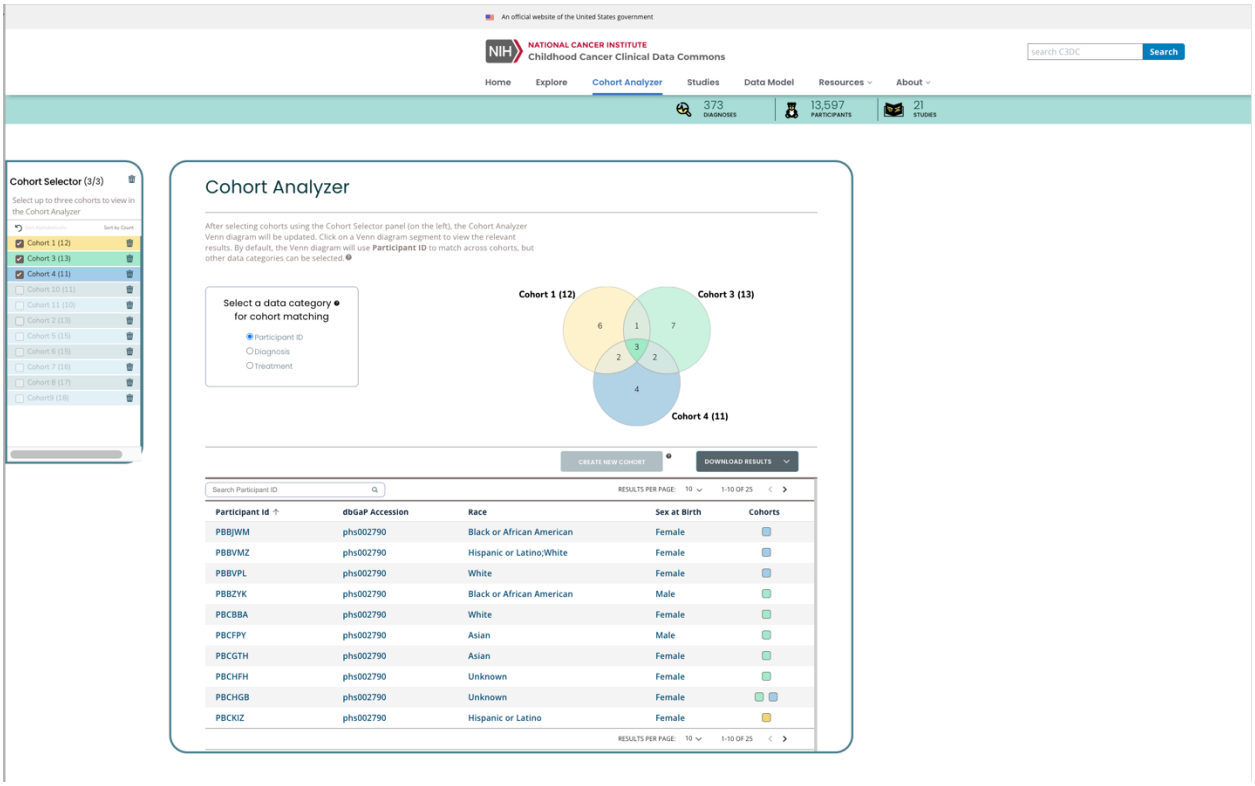

Figure 14: Three Cohorts Selected: You have selected three cohorts. Explore their intersections and unique attributes using the Venn diagram.

Please note that the number in parentheses by the cohort's name in the Venn diagram represents the count of unique records for that radio button selection. The number inside the Venn diagram sections are the count of unique values for that radio button selection. Finally, the count next to your cohort in the Cohort Selection side bar indicates the total participants in your cohort.

At this point, you can select one of these pieces on the Venn diagram to update the table to show only those participants and their respective data. In the example below, the center intersection was selected. The table updates showing only participants that are found in all three cohorts.

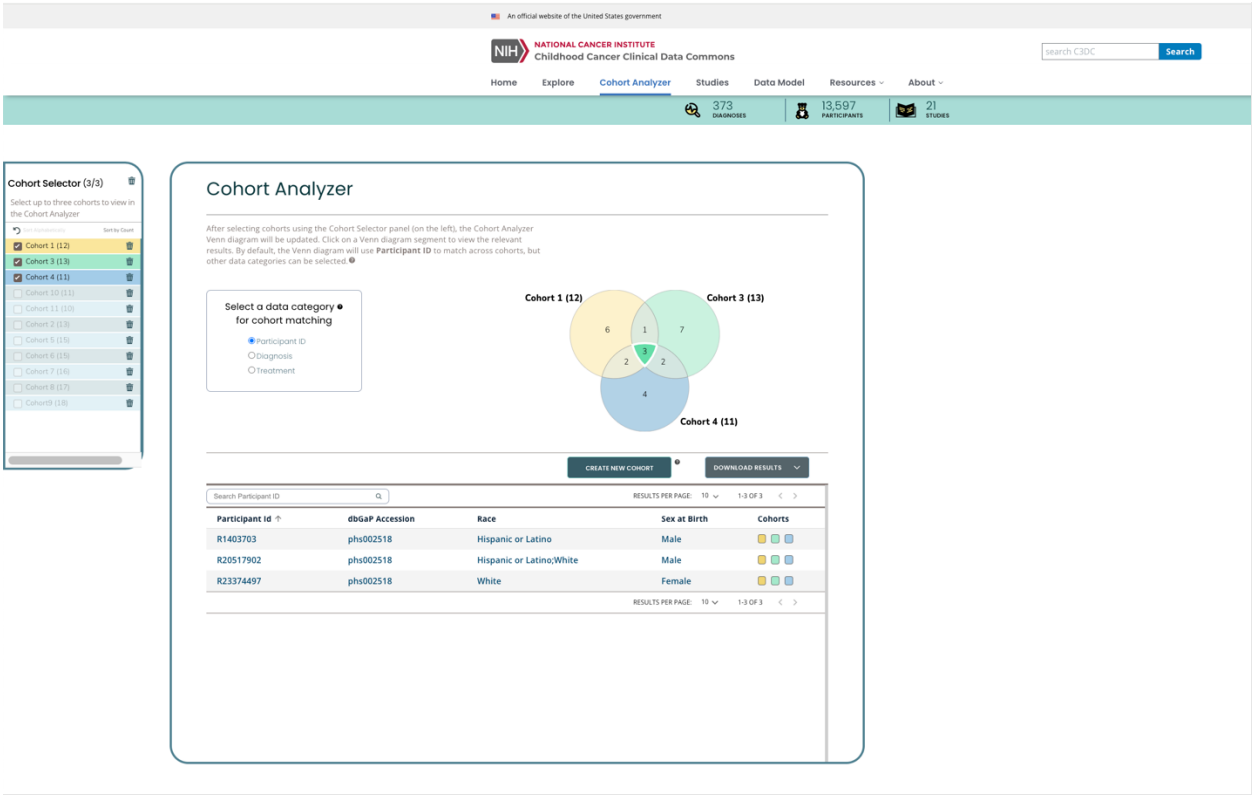

Figure 15: View the center intersection between all Cohort selected (see dark green highlighted region)

The user will see the intersections of all three cohort. Additionally, the user will also see intersections between two cohorts. Clicking on the desired intersection will result in the table being updated accordingly with metadata for those selected participants.

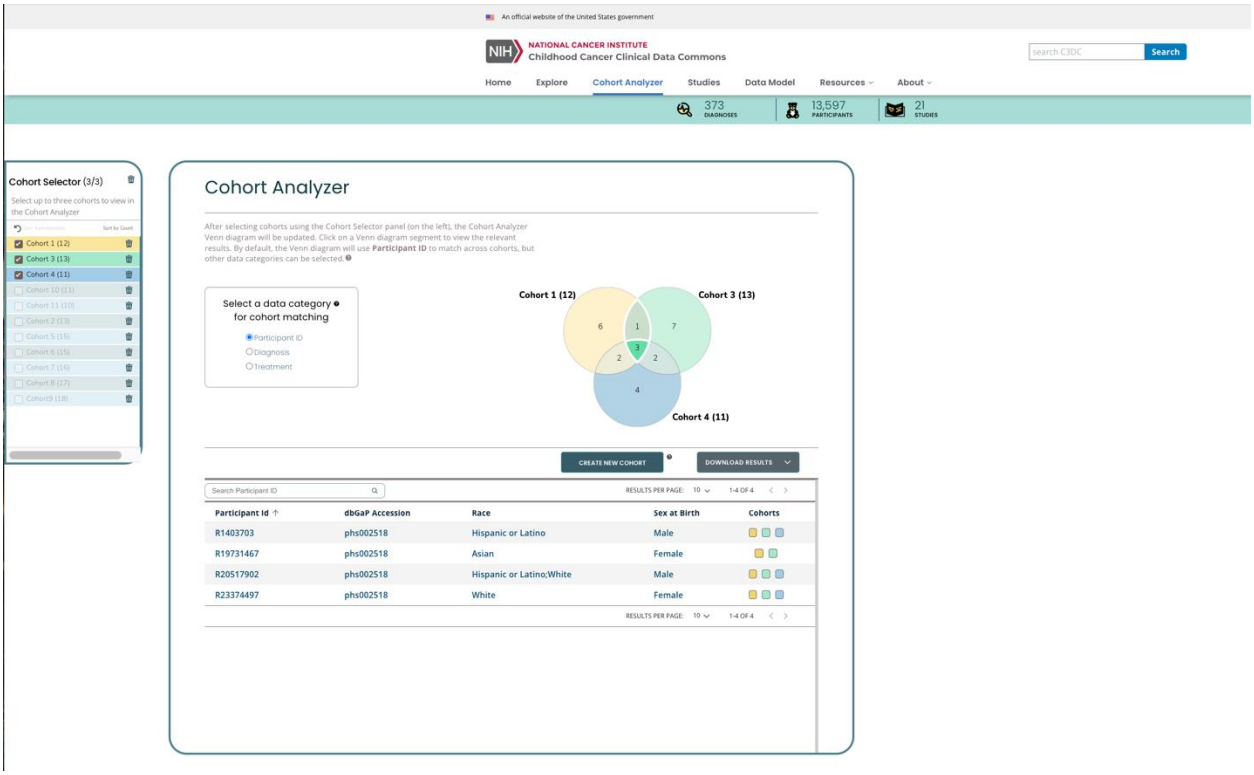

Figure 16: View the specific intersections between selected Cohorts

The Cohort analyzer offers a powerful method to explore how various clinical attributes overlap and differ across multiple groups. By visualizing the shared and unique data points using a Venn diagram, you can identify common patterns or variations in key clinical variables such as diagnosis, treatment, and participant characteristics. This analysis helps reveal underlying trends in the clinical data that may be crucial for research, such as identifying which treatment protocols are common across cohorts or exploring the presence of specific diagnoses.

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.

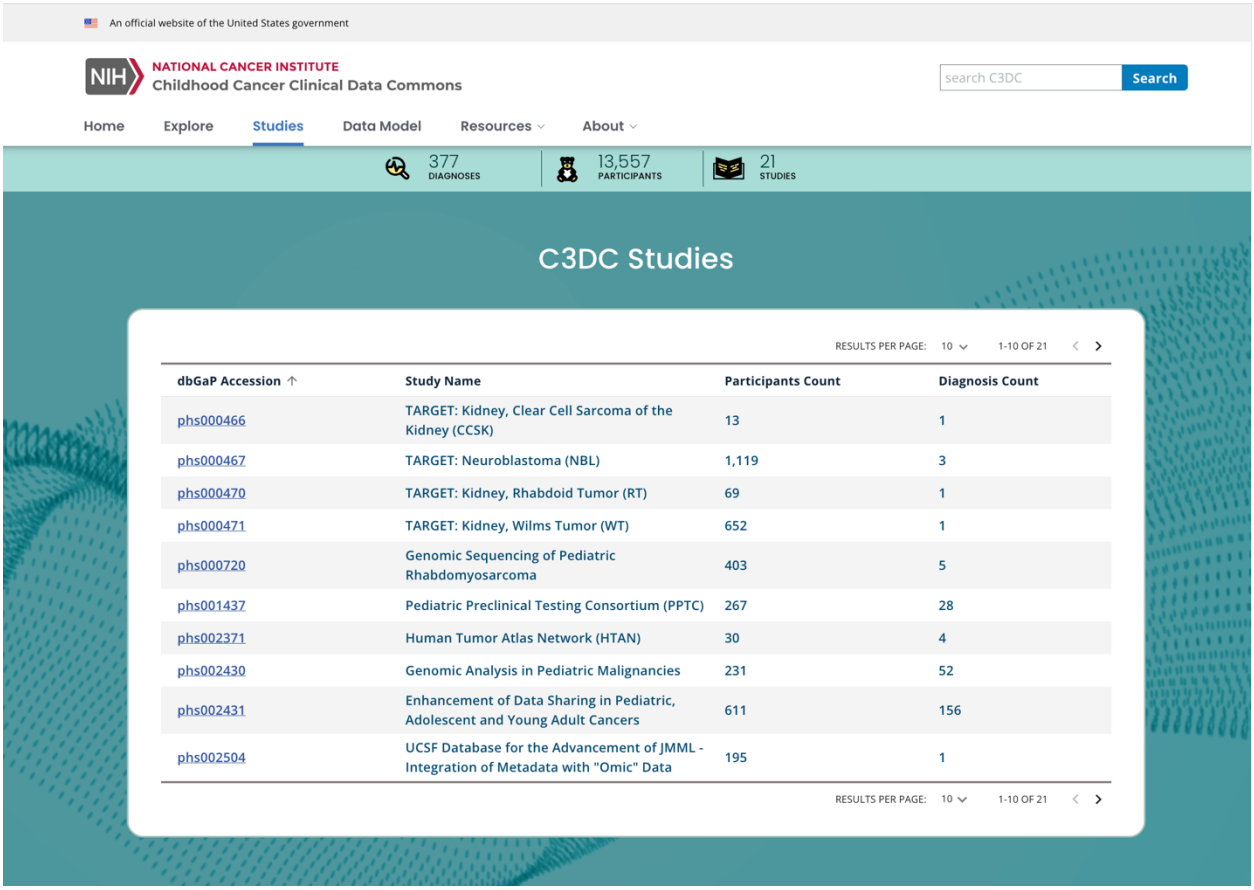

Figure 17: The Studies Page

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 (TARGET datasets phs000466, phs000467, phs000468, phs000469, phs000470, and phs000471) and manifest metadata is available for all other CCDI studies. 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.

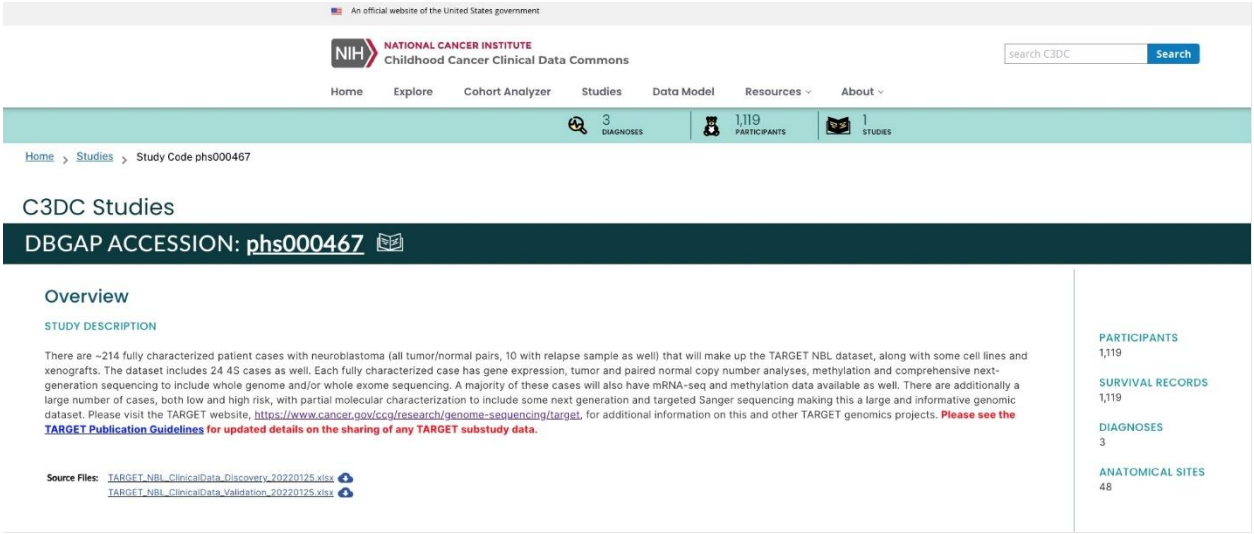

Figure 18A: The Studies Details page – downloadable Open Access Source file data for TARGET datasets

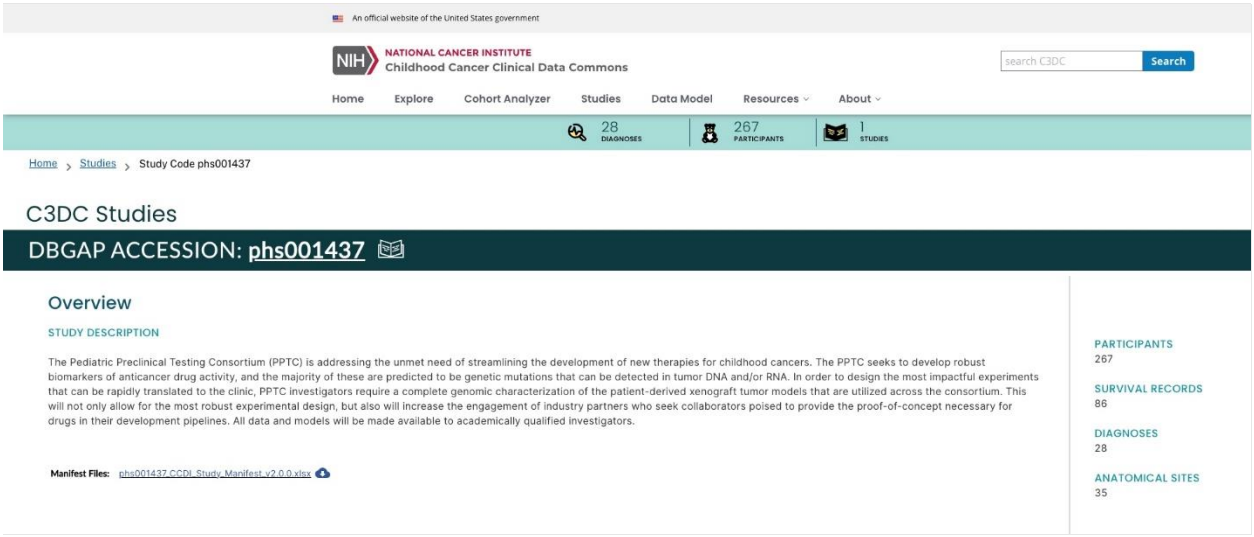

Figure 18B: The Studies Details page – downloadable manifest metadata files for all other CCDI studies

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.

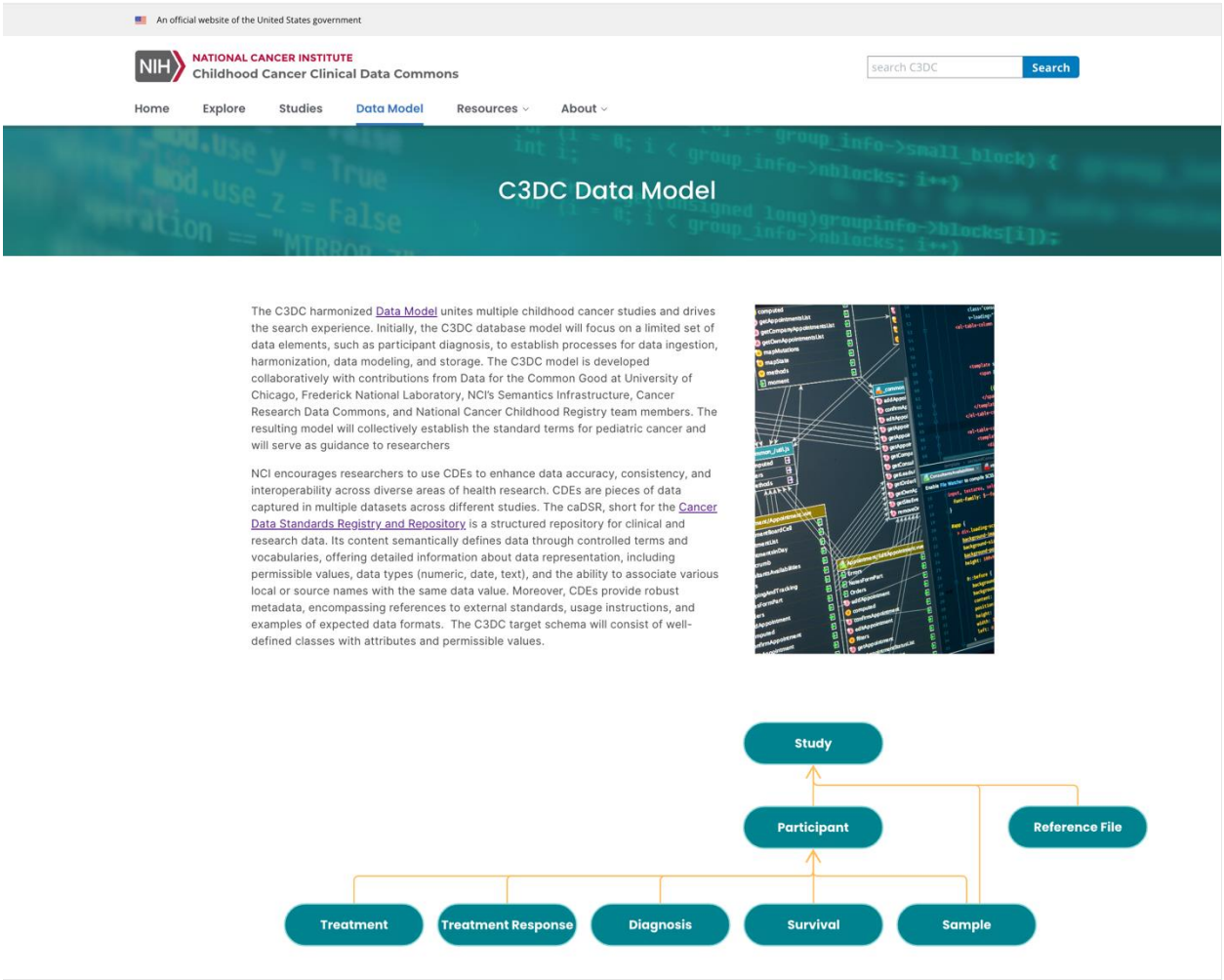

Figure 19: The Data Model page

C3DC Resource Page

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

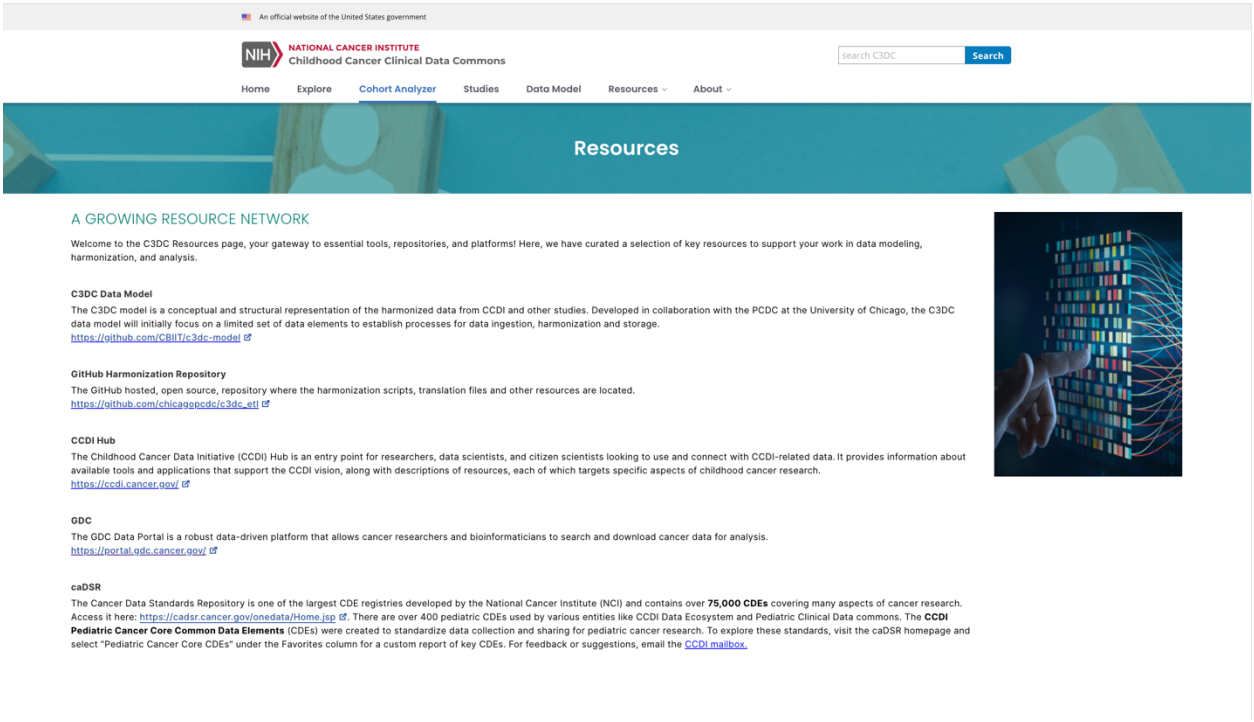

Figure 20: The Resource page

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:

- Announcements
- Release Notes
- User Guide

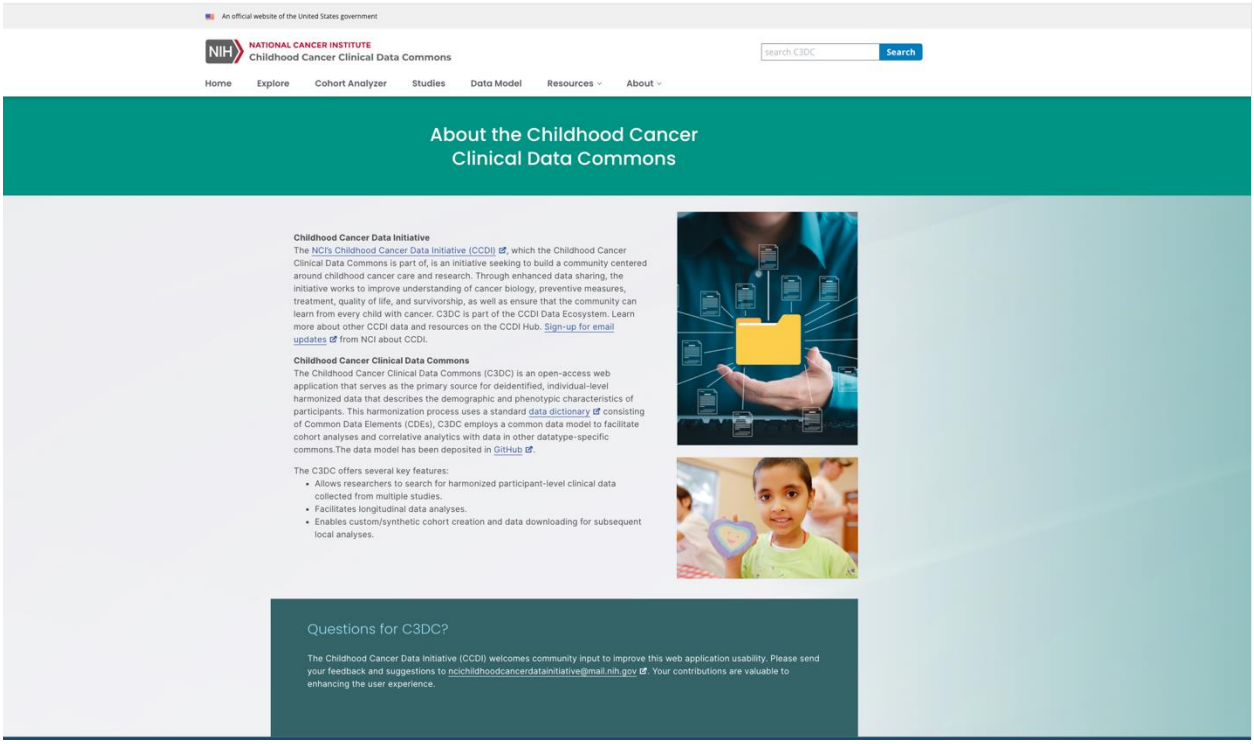

Figure 21: The About page

Announcements

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

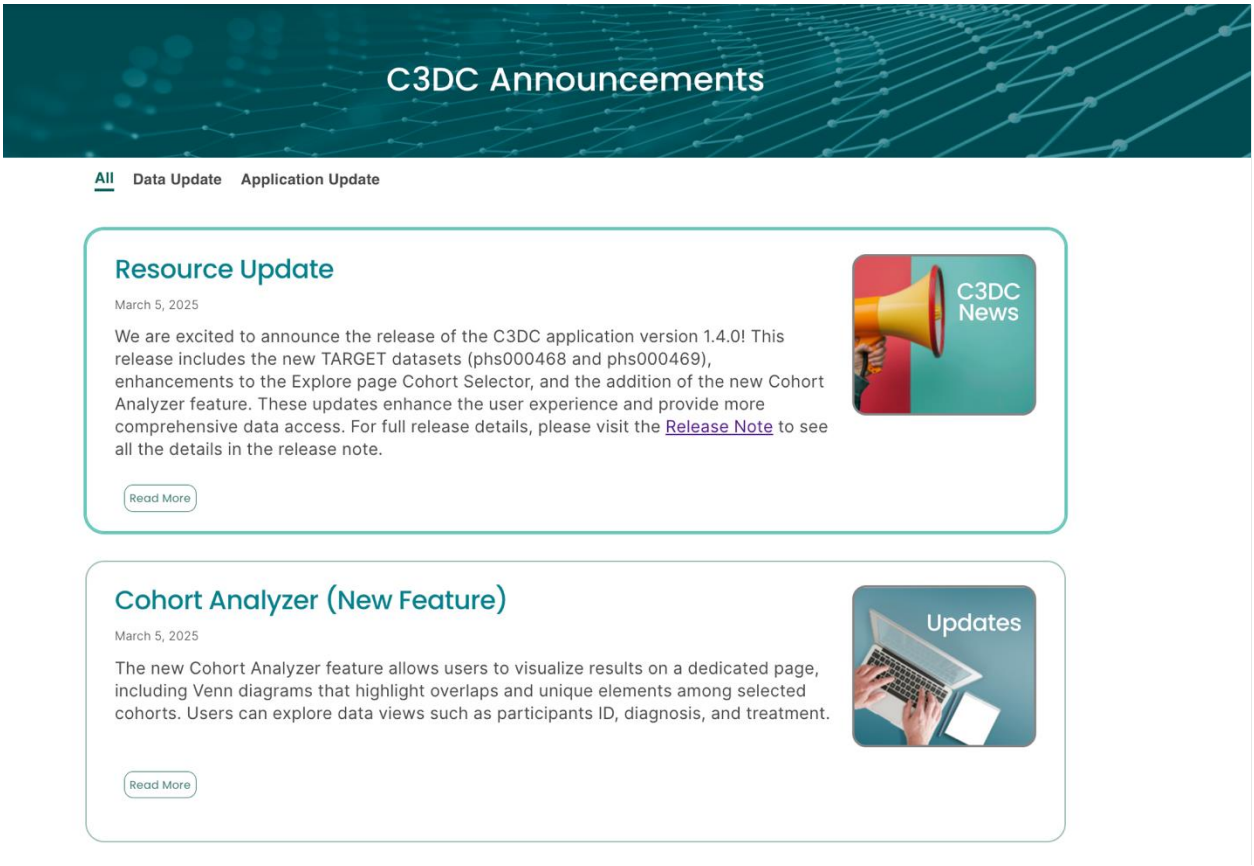

Figure 22: The Announcements page

Release Notes

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

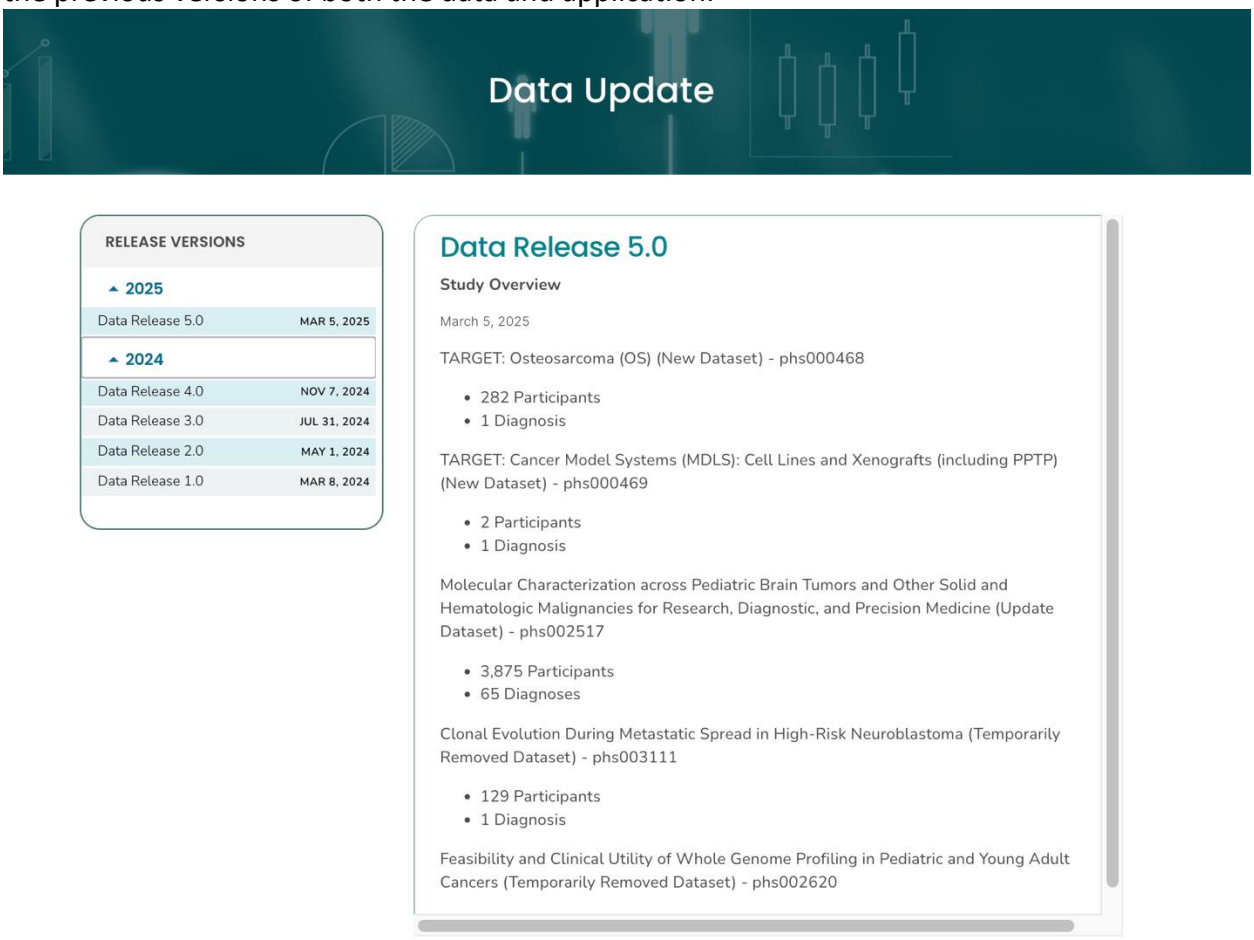

Figure 23: The Release Notes page

User Guide

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

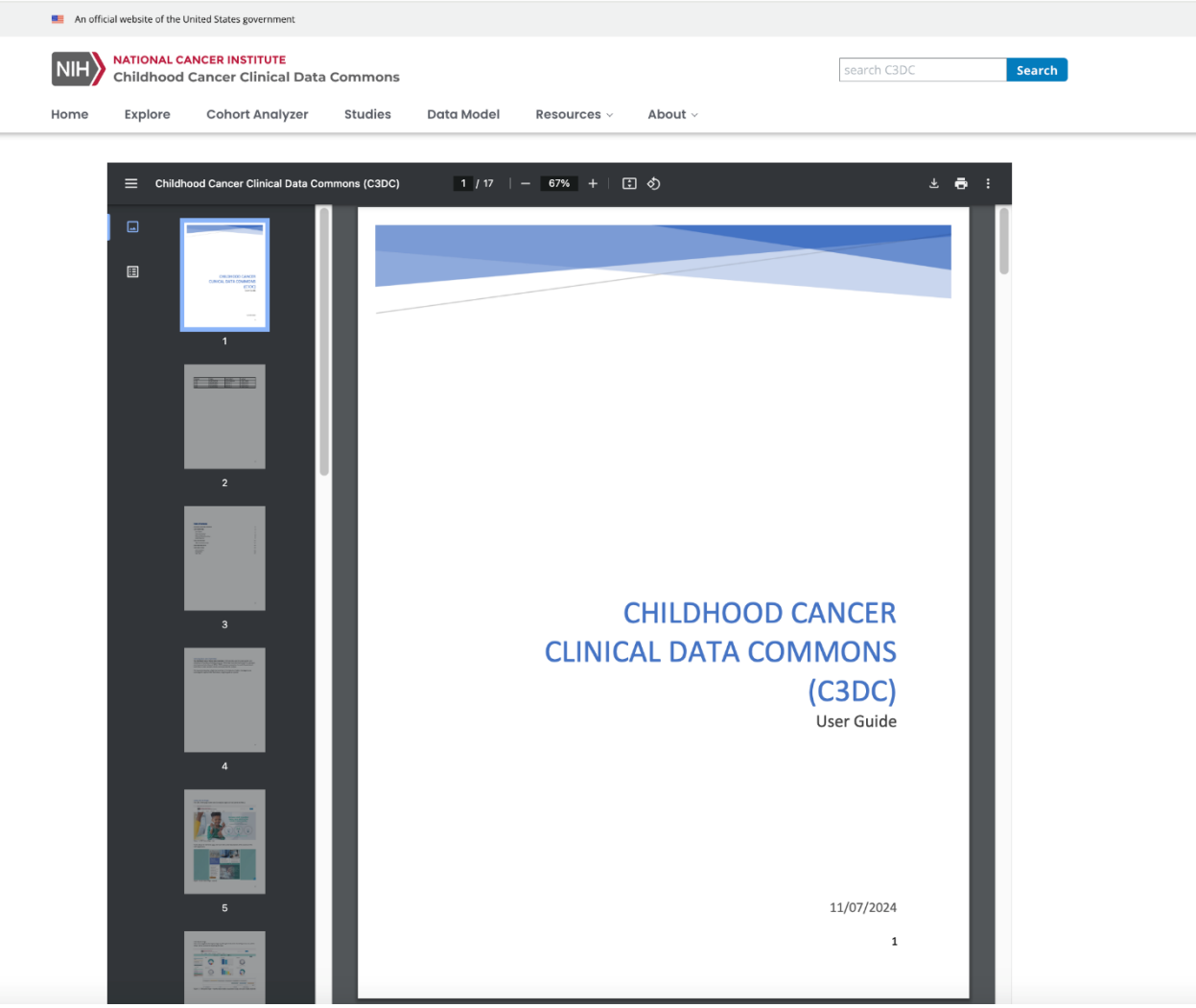

Figure 24: The User Guide