FOR THE COMMON GOOD,

we build communities, platforms, and ecosystems that maximize the potential of data to drive discovery and improve human health.

2021–2022

Since its inception, the Pediatric Cancer Data Commons has experienced significant growth every year. In 2021–22 we continued to raise the bar, producing some particularly exciting milestones.

Several years of preparatory work came to fruition as we launched two new major platforms. The PCDC Data Portal now brings multiple types of clinical data together in one place, offering new analysis tools and creating unprecedented opportunities for cross-disease research and interoperability with other data commons. In collaboration with The Leukemia & Lymphoma Society, we also piloted and launched GEARBOX, a clinician and patient/family support tool that facilitates matching relapsed or refractory AML patients to clinical trials.

We can see the impact of our work growing, with new research collaborations and publications, important roles in national data sharing initiatives, and enthusiasm in the scientific community around the possibilities created by access to high-quality, interoperable data. As we continue to invest in building and expanding the PCDC, we are also beginning to find opportunities to apply our unique approach to data sharing in other rare diseases beyond pediatric cancer. We are excited to build on our current successes to contribute the most we can to improving health and bettering lives.

The progress in this report is the result of the hard work of hundreds of people, both within the PCDC team and beyond. We are grateful to the consortium members, collaborators, and data commons users who continue to drive science forward, to the funders who believe in and support our vision, and to the patients and families whose lived experience is the basis for everything we do. Thank you for joining us in working toward a world where access to high-quality data is never a barrier to improving human health.

Sam Volchenboum, MD, PhD
2021-22 by the numbers

**ENGAGEMENT AND COLLABORATION**

- **10** pediatric cancer consortia
- **113** members of 9 consortium leadership committees
- Committed data contributors with data from 46 countries across 6 continents

**DATA STANDARDS**

- **10** finalized data dictionaries
- **1** PCDC Master Data Dictionary
- **9** disease-specific dictionaries
- **3** more in progress

See all our data dictionaries: [sam.am/datadictionaries](sam.am/datadictionaries)

**THE PCDC DATA PORTAL**

- **31,651** total cases in the PCDC Data Portal
- **4,344** cases newly added this year
- **2** cancer types
- **6,969** rhabdomyosarcoma
- **24,682** neuroblastoma
- **+3** more cancer types now being harmonized

**IMPACT**

- **124** users of PCDC data exploration tools
- **11** new publications
- **9** new projects approved
- **28** total projects now in progress
- 34 pilot users of GEARBOx clinical trial matching tool

**SUSTAINABILITY**

- **20** active sources of funding
- **14%** gifts
- **38%** grants
- **48%** contracts

The next generation: we worked with 18 students and trainees

- **6** undergraduates
- **5** masters students
- **4** medical students
- **3** fellows
The PCDC continues to grow

Each of the ten consortia involved in building the PCDC is focused on a different area of pediatric cancer, and each is in its own stage of development. From bringing together stakeholders to developing consensus policies and harmonizing data, the hard work and expertise of our collaborators combined with the PCDC’s streamlined approach continues to result in fruitful relationships and more data available for research. To see our latest progress and learn more about what each of these milestones entails, visit sam.am/milestones.

Contributing to national data standards

Since 2020, the PCDC has collaborated with the National Cancer Institute (NCI) to support the use of consistent data standards on a national level. The NCI’s data thesaurus, known as NCIt, is the reference terminology used by the NCI and a growing number of other systems to enable data sharing by ensuring that data elements are defined and named in consistent ways. All the terms in PCDC data dictionaries are mapped to NCIt codes so that our data can be used interoperably with other sources. Through this process, we help to build and improve NCIt by acting as a standards authority and contributing missing terms related to pediatric cancer. Thus far, five of our data dictionaries have been fully integrated with NCIt, with more in progress.

A cancer predisposition data dictionary was established. As all PCDC data dictionaries continue to be improved and expanded, a v2 dictionary was established for osteosarcoma.

Initial data contributors were committed for ALL, CNS, cancer predisposition, and retinoblastoma, and one additional data contributor was committed for bone tumors.

INSPIRE (CNS tumors) and C3P (cancer predisposition) signed memoranda of understanding formally establishing their consortia.

Additional neuroblastoma and soft-tissue sarcoma cases were integrated into the PCDC.

NODAL began the process of preparing Hodgkin lymphoma data for the Commons.

New research projects were initiated from INRG, INSTRuCT, and MaGIC.

New papers were published by researchers from INRG, INSTRuCT, and MaGIC, some using data from the Commons and others expressing consensus opinions developed by consortium experts.
Since the PCDC’s earliest days, a global approach has been fundamental to our success. Collaborating internationally allows us to build larger datasets with better genetic and geographical diversity within the available data. As we have developed streamlined approaches to the logistical, legal, and regulatory work that this demands, the geographical scope of the PCDC has continued to grow.

This map illustrates the range of countries from which our ten disease-specific consortia are collecting data. Darker shades indicate countries with data being collected by more consortia.

The PCDC strives to expand our reach further, particularly in parts of the world currently underrepresented in cancer research. In February 2022, a PCDC-led group published “Mapping Pediatric Oncology Clinical Trial Collaborative Groups on the Global Stage” in *JCO Global Oncology*. This paper describes pediatric cancer clinical trial groups around the world with the aim of understanding their structure and function, as well as the clinical data sources they collect, to reveal possible opportunities for collaborative efforts.

The key takeaways of this work suggest that a ‘one size fits all’ approach to partnership is not possible: there is immense variation among existing groups and how they operate, and successful collaboration will require nuanced understanding of the specific stakeholders and needs in each region.
The PCDC Data Portal

A major highlight of 2021-22 was the launch of the PCDC Data Portal. Through the work of our technology team and collaborators in INSTRuCT and INRG, PCDC data are now more interoperable and easily accessible through our new web-based platform. Rhabdomyosarcoma data was the first in the portal for our pilot launch in autumn 2021, with neuroblastoma data integrated in 2022 and more cancer types to come soon.

The PCDC Data Portal utilizes Gen3, the open-source platform that powers the NCI’s Genomic Data Commons. This platform streamlines the work of adding data from additional disease groups and fosters interoperability with commons outside the PCDC.

Potential cohorts can be explored and filtered by many different variables using AND/OR operators.

Participant line-level data for research can be requested through the PCDC project request process.

In June we introduced a pilot of new self-service analysis tools, including a Kaplan-Meier survival curve generator.

There are currently 31,651 cases in the PCDC, an increase of 4,344 since last year.

With multiple cancer types now housed in one place, they may be searched individually or together. The possibilities for cross-disease research will grow further as more cancer types are integrated in the coming years.

Create an account and explore the PCDC Data Portal: portal.pedscommons.org

User guide, tutorial videos, and other documentation: docs.pedscommons.org

Definitions for all elements are easily available in the PCDC data dictionary.

Anyone can create an account and explore data dictionaries, perform cohort exploration, and generate data visualizations for patient cohorts based on data filters.
Collaborations for research and care

Improving outcomes for childhood cancer is a shared goal that requires shared efforts. In addition to the many collaborative relationships involved in building the PCDC, we partner with organizations that share our mission to build tools and resources that drive discovery within pediatric cancer. This year, we launched an important new clinical tool with The Leukemia & Lymphoma Society, and we continue to play a key role in national data sharing initiatives by working with the National Cancer Institute (NCI).

NCI Childhood Cancer Data Initiative

The PCDC is an important part of a national data sharing ecosystem through the NCI Childhood Cancer Data Initiative (CCDI). We engage in major projects centered on data harmonization, and our PI Sam Volchenboum serves in several advisory roles. We are proud to contribute to the CCDI’s efforts to improve cancer prevention, treatment, quality of life, and survivorship and to ensure that researchers learn from every child with cancer.

PCDC-H

In 2022 we concluded an 18-month project which established the foundations for integrating the PCDC with the NCI Cancer Research Data Commons (CRDC). By developing and mapping data to a common PCDC Harmonization (PCDC-H) data model aligned with the NCI Thesaurus, we made it possible to link PCDC data with data in other CRDC nodes across the country, creating a robust, integrated resource for pediatric cancer research.

C3DC

We are participating in developing the Childhood Clinical Data Commons (C3DC), a data node of the CCDI that will act as the primary source of individual-level data describing participants’ demographic and clinical characteristics. C3DC will interoperate with other CCDI data type-specific nodes such as genomics, imaging, and proteomics.”

*We have received $519,688 in funding for this project with an Option Period to extend for additional funding in an amount of $1,311,602. The total anticipated budget for this project is $1,863,270, 100% of which is financed with federal money.

GEARBOx

As part of The Leukemia & Lymphoma Society (LLS) PedAL Master Clinical Trial, we recently launched a clinical trials matching tool for clinicians to rapidly and accurately match children with relapsed acute myeloid leukemia to targeted treatments. GEARBOx (Genomic Eligibility Algorithm at Relapse for Better Outcomes) is a web-based tool that uses a matching algorithm to identify potentially appropriate clinical trials based on COG eligibility criteria and the patient’s clinical data, immunophenotype, and genomic profile.

GEARBOx users can enter information about patients, including demographics, disease characteristics, and biomarkers, that will help determine their eligibility for trials.

A list of currently open clinical trials that may be a potential match is returned, with details on location and enrollment.

Leukemia patients can be connected to an LLS Nurse Navigator for one-on-one support.

While the current version of GEARBOx is specific to AML, we will be extending GEARBOx to be used for additional types of cancer.

Explore GEARBOx: gearbox.pedscommons.org
Building community

The PCDC is one part of a larger ecosystem. In addition to the collaborations highlighted in the previous pages, this year we deepened our connections with the global scientific community through new advisory relationships and a guest speaker series.

Guest speaker webinar series

In October 2021 we launched The Common Good, a new guest speaker webinar series, with the goal of broadening our audience and having important conversations about opportunities for progress. For our first year of talks, we invited thought leaders from across the fields of pediatric oncology, clinical trials, and big data to share ideas that can redefine how we think about improving human health. The more we learn from each other, the better we can all work toward the common good.

- **Greg Simon**, JD, former President of Biden Cancer Initiative, on how clinical trials can be made more accessible and equitable
- **Sam Blackman**, MD, PhD, founder of Day One Biopharmaceuticals, on new drug development models that put children first
- **Adam Resnick**, PhD, Principal Investigator of the Gabriella Miller Kids First Data Resource Center, on large-scale resources to accelerate pediatric cancer research
- **Carlos Rodriguez-Galindo**, MD, Director of St. Jude Global, on addressing global disparities in childhood cancer

Members of the SAC and EAB are listed on page 21 of this report.

Watch the talks: sam.am/commongood

A new advisory structure

A new step in the PCDC's growth is the establishment of two advisory groups. Our Scientific Advisory Committee (SAC) of consortium representatives and domain experts provides strategic and operational guidance for the PCDC, while each consortium maintains autonomy in its own operations and scientific direction. In parallel, an External Advisory Board (EAB) made up of an outside group of clinical, data science, and digital health experts will launch in late 2022 to share their experiences and offer strategic advice on PCDC initiatives.

First year of Common Good guests

- **Greg Simon**, JD
- **Sam Blackman**, MD, PhD
- **Adam Resnick**, PhD
- **Carlos Rodriguez-Galindo**, MD

SAC geographical representation

- **41% United States**
- **59% outside US**

Spotlight on the SAC

The PCDC Scientific Advisory Committee launched in November 2021 and meets quarterly. The group includes representatives from ten countries and four continents, with leaders from each PCDC consortium and international oncology cooperative groups, statisticians, an ethicist, a patient advocate, and members at large, including early career investigators. The group has thus far advised on safeguards for new data exploration tools, EAB membership and discussion topics, and more.

Members of the SAC and EAB are listed on page 21 of this report.

Watch the talks: sam.am/commongood
Beyond pediatric cancer

The foundation of our work is reducing barriers to sharing useful, high-quality data between institutions, groups, and countries in order to increase opportunities for discovery. With our established success applying this approach to pediatric cancer, we have begun to find opportunities where similar efforts in other rare diseases could be paradigm shifting.

The streamlined and scalable infrastructure and processes that we have developed for pediatric cancer data are now enabling us to begin extending our work to new areas where we can make a difference. With our unique approach to building data commons, which prioritizes relationship-building, data quality, and sustainability, we are thoughtfully expanding our scope and working in partnership with researchers, clinicians, and patients to drive new science and improve lives.

The social context of disease

The Sociome refers to the wide array of factors beyond the biological mechanisms of disease that are involved in human health. These social contexts—everything from environmental pollution to economic stability to access to education and health care—play critical roles in individual health outcomes. However, unlike clinical and genomic data, most of these social contributing factors have not been comprehensively collected in a way that is suitable for large-scale analysis. A new partnership among many groups and researchers headquartered at the University of Chicago aims to address this challenge. The PCDC is playing a key role by contributing our expertise in data standardization, harmonization, and infrastructure to develop a data commons for the Sociome, a resource that will allow researchers to integrate the social context of disease with clinical and genomic data to better understand, predict, and treat numerous conditions and improve human health.

A rare disease: monogenic diabetes

Monogenic diabetes is an atypical form of the disease caused by changes to a single gene, representing 1-4 percent of cases of diabetes in the US. Due to its rarity, a single source for patient data would be a critical resource for researchers to advance science and clinical practice. Data-generating research initiatives already exist but are not coordinated, making this subtype of diabetes a strong fit for the PCDC approach to data commons development. Working closely with the UChicago Kovler Diabetes Center, PREDICT (PREcision DIabetes ConsorTium) will bring together international stakeholders to build a commons that will include clinical data, patient-reported outcomes, and data from wearable devices such as continuous glucose monitors. The consortium is currently applying PCDC methods to building data dictionaries and developing governance structures.
Driving discovery

As we continue to build and grow the PCDC with data for additional cancers, our consortia that have already made data available for research continue to drive new science and publish their discoveries. This year’s scientific impact includes research based on data from that have already made data available for research continue to drive new science and publications from the PCDC on our tools and methods.

Publishing papers on our methods is an important part of our commitment to driving progress in our field. In “Pediatric Cancer Data Commons: Federating and Democratizing Data for Childhood Cancer Research,” the PCDC team and collaborators present our experience constructing the PCDC to highlight the significance of developing a rich and robust data ecosystem for pediatric oncology and to provide essential information to those creating resources in other disease areas. We identify six critical features of successful data commons design and implementation that can be extended to any clinical area.

The paper “Immunogenomic determinants of tumor microenvironment correlate with superior survival in high-risk neuroblastoma” demonstrates the value of interoperable data by connecting PCDC clinical data with genomic data. The researchers used universal system identification numbers to link clinical data from INRG with tumor RNAseq data from the Gabriella Miller Kids First Data Resource Center to create a validation cohort, which was used to study the prognostic strength of T cell-inflamed gene expression and neoantigen load in high-risk neuroblastoma.

Selected Presentations


Frazier AL. The magic of MaGIC. Presented at the Annual Meeting on Women’s Cancer; March 2022.


Nicholson J. Germ Cell Tumours: What we have learned from MaGIC and how this model of collaboration can apply to other cancers. Presented at the 53rd Congress of the International Society of Paediatric Oncology; October 2021.


For the Common Good

Pediatric Cancer Data Commons 2021-22
Toward a better future

The PCDC is founded on community and collaboration, and the progress detailed in this report would not be possible without the generous support of those who believe in our work. We are inspired by our funders’ commitments to fighting childhood cancer and improving human health, and grateful for their investment in our vision. These funders supported us during the 2021-22 fiscal year.

Bridge to a Cure Foundation provided support for the INSPIRE consortium and the development of a data dictionary for CNS tumors.

Cancer Research Foundation enabled the PCDC to expand our international outreach and collaboration, including funding for incorporating retinoblastoma clinical trial data from Latin America into the Commons.

Children's Cancer Research Fund continued to enable our work building a retinoblastoma consortium (Global REACH) and data dictionary.

Children’s Research Foundation provided continued support for strengthening and expanding PCDC collaborations.

Comer Development Board helped to sustain the PCDC’s centralized project management, governance, and engagement operations across all our consortia.

The William and Evelyn Fuchs Family Foundation supported our ongoing efforts to integrate new disease areas and collaborators into the PCDC.

The Fund for Innovation in Cancer Informatics enabled us to begin working toward extending the GEARBOx clinical trials matching tool to additional types of cancer.

Gray Foundation supported our work extending the PCDC approach to data to rare diseases beyond pediatric cancer as we began developing a consortium and data commons for monogenic diabetes.

The Leona M. and Harry B. Helmsley Charitable Trust provided essential support for the newly created monogenic diabetes consortium PREDICT and our work to begin building a data commons for this disease.

The Leukemia & Lymphoma Society continued to support our AML consortium INTERACT as they focused on harmonizing and preparing data for the Commons, as well as enabling the development, pilot, and launch of the GEARBOx clinician support tool for the LLS PedAL initiative.

The Neuroblastoma Children’s Cancer Society provided funding for our neuroblastoma consortium INRG, with a special focus on efforts to gather and analyze data to better understand and treat secondary illness and side effects in neuroblastoma survivors.

Rally Foundation for Childhood Cancer Research and their partners Infinite Love for Kids Fighting Cancer and The Truth 365 continued to support our bone tumor consortium HIBiSCus as they created work groups, executed data contributor agreements, and finalized a second version of the osteosarcoma data dictionary.

St. Baldrick’s Foundation provided essential support for our development and launch of the PCDC Data Portal and analysis tools, as well as enabling engagement, governance, and project management efforts across all ten of our disease group consortia.

Another grant from St. Baldrick’s Foundation enabled the PCDC to continue working with the Consortium for Childhood Cancer Predisposition (C3P) to develop governance policies and a data dictionary.

Team Bright Side continued to enable our efforts to build an acute lymphoblastic leukemia consortium, develop an ALL data dictionary, secure data contributors, and begin harmonizing data.

A contract with Leidos and the National Cancer Institute enabled us to participate in developing the Childhood Clinical Data Commons, a data node of the CCDI housing participants’ demographic and clinical characteristics.

The Neuroblastoma Children’s Cancer Society provided funding for our neuroblastoma consortium INRG, with a special focus on efforts to gather and analyze data to better understand and treat secondary illness and side effects in neuroblastoma survivors.

Rally Foundation for Childhood Cancer Research and their partners Infinite Love for Kids Fighting Cancer and The Truth 365 continued to support our bone tumor consortium HIBiSCus as they created work groups, executed data contributor agreements, and finalized a second version of the osteosarcoma data dictionary.

St. Baldrick’s Foundation provided essential support for our development and launch of the PCDC Data Portal and analysis tools, as well as enabling engagement, governance, and project management efforts across all ten of our disease group consortia.

Another grant from St. Baldrick’s Foundation enabled the PCDC to continue working with the Consortium for Childhood Cancer Predisposition (C3P) to develop governance policies and a data dictionary.

Team Bright Side continued to enable our efforts to build an acute lymphoblastic leukemia consortium, develop an ALL data dictionary, secure data contributors, and begin harmonizing data.

A contract with the US Department of the Interior enabled us to continue integrating PCDC data standards with the NCI Cancer Research Data Commons.

An anonymous foundation supported our development of consensus data dictionaries.
The PCDC team

As the scope and impact of the PCDC continued to grow this year, our team grew as well. With experience in medicine, data standards, technology, governance, project management, and education, our expertise is diverse, but our vision is shared. Together we work to be leaders in how data are collected, connected, and used for good.

Maya Maric
Project Manager

Rolando Palacios
Technical Project Manager

Enal Hindi, MS
Sr. Project Manager

Bobae Kang
Front End Web Developer

Steve Kraskinsky
Full Stack Developer

Jooho Lee, PhD, MSLIS
Semantic Engineer

Mei Li, MS
Clinical Data Standards Analyst

Suzi Birz, MScMI, FHIMSS
Regulatory and Data Governance Consultant

Kat Bouzein, MS
Program Manager

Brad Carlson
Clinical Data Specialist

Seong Choi
Sr. Software and Data Integration Engineer

Ellen Cohen, MPP
Sr. Director of Academic Operations

Nicholas Ferraz, MS
Full Stack Developer

Brian Furner, MS
Sr. Director of Data and Technology

Luca Graglia, MS
Software Engineering Manager

Maya Maric
Project Manager

Rolando Palacios
Technical Project Manager

Enal Hindi, MS
Sr. Project Manager

Bobae Kang
Front End Web Developer

Steve Kraskinsky
Full Stack Developer

Jooho Lee, PhD, MSLIS
Semantic Engineer

Mei Li, MS
Clinical Data Standards Analyst

Scientific Advisory Committee

Todd Alonso, PhD
Maja Beck-Popovic, MD
Gianni Bisogno, MD, PhD
Susan L. Cohn, MD
Jamie Flattage, MD, MS
Brenda Galia, MD, FRCSc, CM, OCist
Ajay Gupta, MD, MS
Daren Hargrave, MB, CNB, MD
Stefanie Hecker Nolting, MD
Katie Janeway, MD
Pamela Kearns, MD, PhD
Kara Kelly, MD
Sarah Leary, MD, MS

Visit our website
commons.uchicago.edu

Explore PCDC tools
pedscommons.org

Follow us on Twitter and ResearchGate
twitter.com/PedsDataCommons
sam.am/researchgate

Questions about getting involved or supporting the PCDC?
Email Kat Bouzein at kblumhardt@chicago.edu

Abbreviation Key

ALL: acute lymphoblastic leukemia
AML: acute myeloid leukemia
CNS: central nervous system tumors
EWS: Ewing sarcoma
GCT: germ cell tumors
HL: Hodgkin lymphoma
MDS: non-malignant stem cell disorders
NBL: neuroblastoma
OS: osteosarcoma
RMS: rhabdomyosarcoma
STS: soft-tissue sarcoma