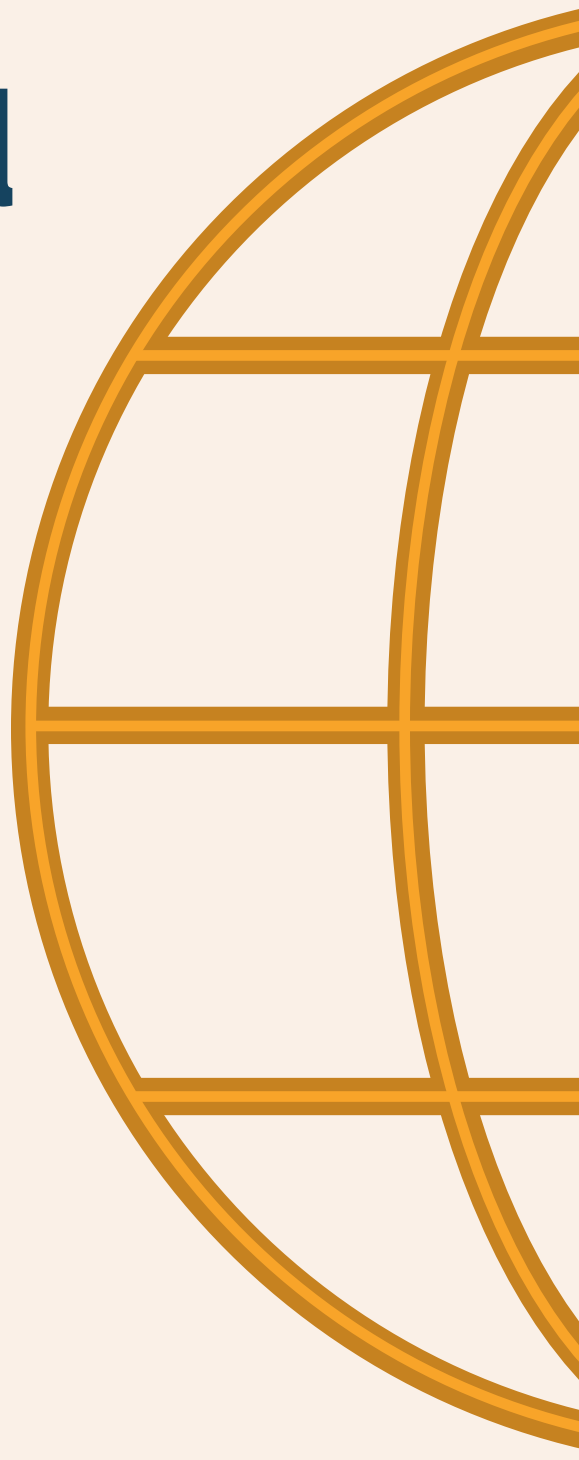
Connected Data

Collective Impact



Data for the Common Good
at the University of Chicago

July 2023–June 2024

D4CG

Working toward a world where access to high-quality data is never a barrier to improving human health.

Contents

Powering discovery with the PCDC 2

New horizons for research and care 11

Stronger together 16

2023–2024

As we reflect on another year of progress and impact, we're reminded that our greatest strength lies in the power of collaboration. At Data for the Common Good (D4CG), we are fortunate to work alongside a diverse and dedicated community of researchers, clinicians, patient advocates, and organizations who share our vision of a world where access to high-quality data is never a barrier to improving human health.

Our flagship program, the Pediatric Cancer Data Commons (PCDC), is thriving. We continue to engage with more groups and work to broaden and deepen the data available, creating a robust resource for studying cancer through the entire lifetime of a patient, including predisposition syndromes and post-treatment effects. This year we were honored to have the PCDC recognized as a Champion of Open Science by the White House Office of Science and Technology Policy—an achievement that underscores the value of an open, collaborative approach to scientific discovery, and a testament to the hard work of our partners and the commitment of patients and families who contribute their data.

We've continued to leverage the lessons learned from the PCDC as we apply our unique D4CG approach—inclusive, consensus-based, and strongly focused on data quality and usability—to more areas where we know data sharing can unlock discovery. We are thrilled to now be building an unprecedented international data commons for monogenic epilepsies. Our efforts in monogenic diabetes and the Sociome are growing, and our sights are set on bringing our approach to more rare diseases and difficult-to-study topics.

One more highlight of this year for us was the many opportunities we had to connect with our colleagues in person at meetings and conferences. Collaboration is truly at the heart of everything we do. Together, we will continue building a future where community-driven data sharing helps solve some of the most challenging health problems of our time.

Sam Volchenboum, MD, PhD

Powering discovery with the PCDC

The Pediatric Cancer Data Commons builds communities and harnesses data from around the world to create a single unified platform for cancer research.

OSTP Champion of Open Science

We were extremely honored this year to have the Pediatric Cancer Data Commons named a Champion of Open Science by the White House Office of Science and Technology Policy (OSTP). These awards highlight efforts that tackle problems by broadening access to research so that everyone can benefit from science and technology. The PCDC and the other four winners "exemplify what's possible with an open, equitable, and secure research enterprise." An open and collaborative approach is the foundation of the PCDC and all of our work at D4CG. This award is a testament to the value of this approach, the hard work of our collaborators and supporters, and the patients and families who have contributed data to the PCDC.

> PCDC's efforts to grow this vitally important knowledge base, while advancing the goals of the Cancer Moonshot, through a commitment to relationship-building, data quality, and sustainability will continue to enable researchers, clinicians, patients, and their loved ones to push the boundaries of research and improve health.
>
> - OSTP press release, March 2024

The PCDC Data Portal continues to grow

The PCDC Data Portal, the public platform for exploring and requesting data from the Pediatric Cancer Data Commons, continues to grow as our consortia collect and harmonize data. A highlight of 2023-24 was onboarding three new cancer types, bringing the number of diagnoses with data now available to six. The portal is enabling a growing number of researchers around the world to perform cohort discovery, build survival curves, and start projects with PCDC data.

Data for **6** cancer types

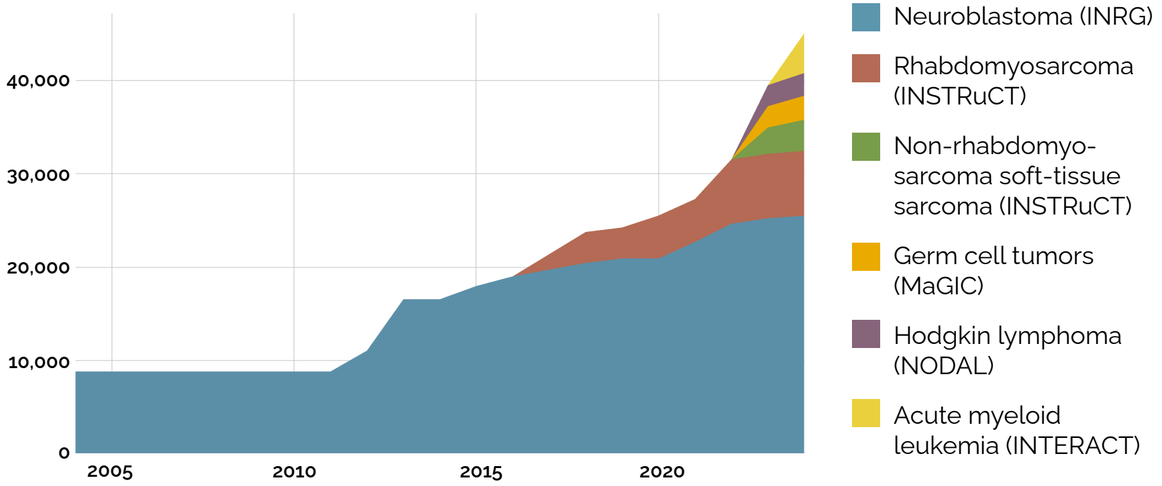
45,117 total participants in the portal

8,027 added in FY24

1,832 added in late 2024

28% increase since FY23

Growth in data available over time



We are thankful for all the hard work from consortia that goes into gathering, harmonizing, and making PCDC data available.

The PCDC Data Portal in action

1,108
registered data portal users
+89% since FY23

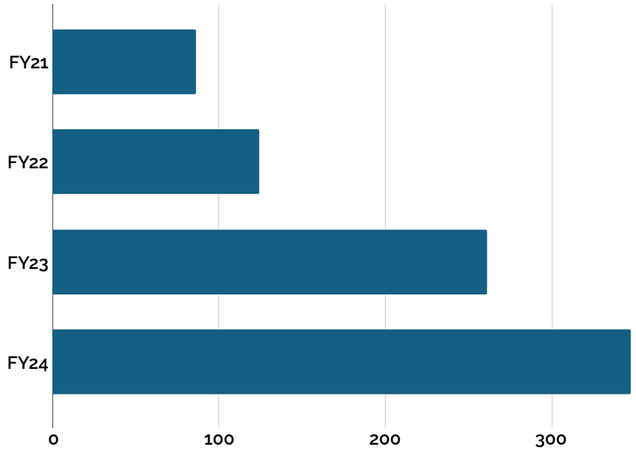
347
active in the last six months of FY24
+33% since FY23

110
survival curves generated
+22% since FY23

Portal users logged in from
48
countries

62% USA
38% other countries

PCDC Data Portal active users by year



33 research projects using PCDC data currently approved or in progress

See all PCDC publications and projects in progress in our online dashboard at **sam.am/pcdcresearch**

Explore the PCDC Data Portal:
portal.pedscommons.org

User guide, tutorial videos, and more:
docs.pedscommons.org

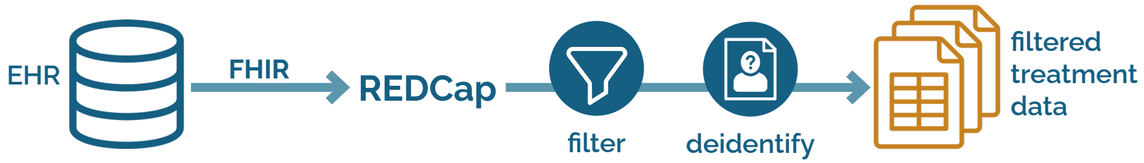
A pathway to enriching the PCDC with real-world treatment data

While the retrospective data already in the PCDC have enabled numerous discoveries, the inclusion of real-world, electronic health record (EHR) data would significantly increase the types of meaningful research that the commons can support, especially when it comes to understanding the impact of specific therapeutics on patient outcomes. These real-world data, which provide valuable insights into routine clinical care that are often missing from clinical trials, have traditionally only been obtainable through time-consuming and error-prone manual chart review.

Publication highlight

Extracting Electronic Health Record Neuroblastoma Treatment Data With High Fidelity Using the REDCap Clinical Data Interoperability Services Module. Furner B, Cheng A, Desai AV, Benedetti DJ, Friedman DL, Wyatt KD, Watkins M, Volchenboum SL, Cohn SL. *JCO Clin Cancer Inform*. 2024 May:8:e2400009. doi: 10.1200/CCI.24.00009

In a pilot study with INRG, we tested a method for the automated extraction of real-world neuroblastoma treatment data from EHR using the REDCap Clinical Data Interoperability Services (CDIS) module, leveraging the Fast Healthcare Interoperability Resource (FHIR) standard, which represents EHR data in an interoperable way. Results showed that this method performed with a high degree of accuracy, demonstrating the feasibility of automated extraction. We are now in the process of expanding our pilot to include additional clinical sites, EHR vendors, and eventually other data types and diseases. With wider implementation, this approach will make it easier for EHR data to be extracted and included in data resources, enrich the data in the PCDC, and facilitate research that could lead to new treatments and better outcomes for patients.



Studying cancer through a lifetime

We're working to make the PCDC a resource that will support studying cancer through the entire life cycle, including pre-cancer syndromes and post-treatment effects. Broadening the types of data in the PCDC can both serve the unique needs of specific patient communities and benefit all patients by enabling research across wider populations.

AYA cancer in the PCDC

Some cancer types in the PCDC, like germ cell tumors and Hodgkin lymphoma, frequently affect adolescent and young adult (AYA) patients. We've spent this year working with our AYA Research Council (ARC), which launched in November 2023, to prioritize bringing more data into the PCDC for this historically under-studied population.

More AYA data in the PCDC will help enable the study of:

- The unique cancer subtypes and biology that affect this population
- Differences in therapies between pediatrics and medical oncology
- Differences in adverse event profiles and differential toxicity

Pediatric Cancer Data Commons
AYA Research Council

Spotlight on the ARC

The AYA Research Council amplifies the perspective of the adolescent and young adult cancer population and provides strategic and operational guidance related to cross-disease AYA science to address the unique research needs of this community.

ARC co-chairs

Jennifer McNeer
United States

Mike Roth
United States

Dan Stark
United Kingdom

Gabriela Villanueva
Argentina

A pre-cancer syndrome: Fanconi anemia

Fanconi anemia (FA) is a rare genetic cancer predisposition disease. Individuals living with FA have a very high risk of developing bone marrow failure and solid tumors and often have many other systemic issues. In partnership with the Fanconi Cancer Foundation, the new consortium FRIENDS (Fanconi Research Initiative for Education, Networking, and Data Sharing) is working to engage the FA community, develop a data dictionary, and collect and harmonize previously-siloed FA data for inclusion in the PCDC, creating new opportunities for rare disease research and discovery.

FRIENDS presented about their work and held a ceremonial MOU signing at the 2024 FCF Scientific Symposium. From left: Isis Sroka, Kirk Wyatt, Laura Hefner

Upon entering the world of rare diseases, I encountered a profound quote: **"If you work with common diseases, you have to compete. If you work with rare diseases, you have to make friends."** More than two decades later, this sentiment gets even more real to me than ever. Thanks to the remarkable initiative by FCF, a collaboration with D4CG has been launched, aiming to register as many individuals with FA worldwide as possible.

-Eunike Velleuer-Carlberg, MD

2023-24 PCDC milestones

Our progress tracker on page 9 highlights some of the biggest accomplishments of our consortia this year. To learn more about what each of these steps entails, visit **sam.am/milestones**.

- A partnership was established with the Children's Hepatic Tumors International Collaboration (CHIC) to include liver tumors in the PCDC. Work began in late 2023 with strategic planning and a line-by-line review of the liver tumors data dictionary.
- The first version of the data dictionary for retinoblastoma was completed, and work began on the Fanconi anemia data dictionary.
- Reproductive HOPE signed their Memorandum of Understanding, officially establishing their consortium.
- Data from MaGIC and NODAL joined the PCDC Data Portal for the first time in September 2023 and INTERACT data in January 2024.
- New participants and updated data were added to the PCDC Data Portal for INRG and INSTRuCT.
- INRG, INSTRuCT, and MaGIC published new papers, and NODAL had its first publication, a paper introducing the consortium.

For full names of consortia, please see Appendix 1.

What's next?

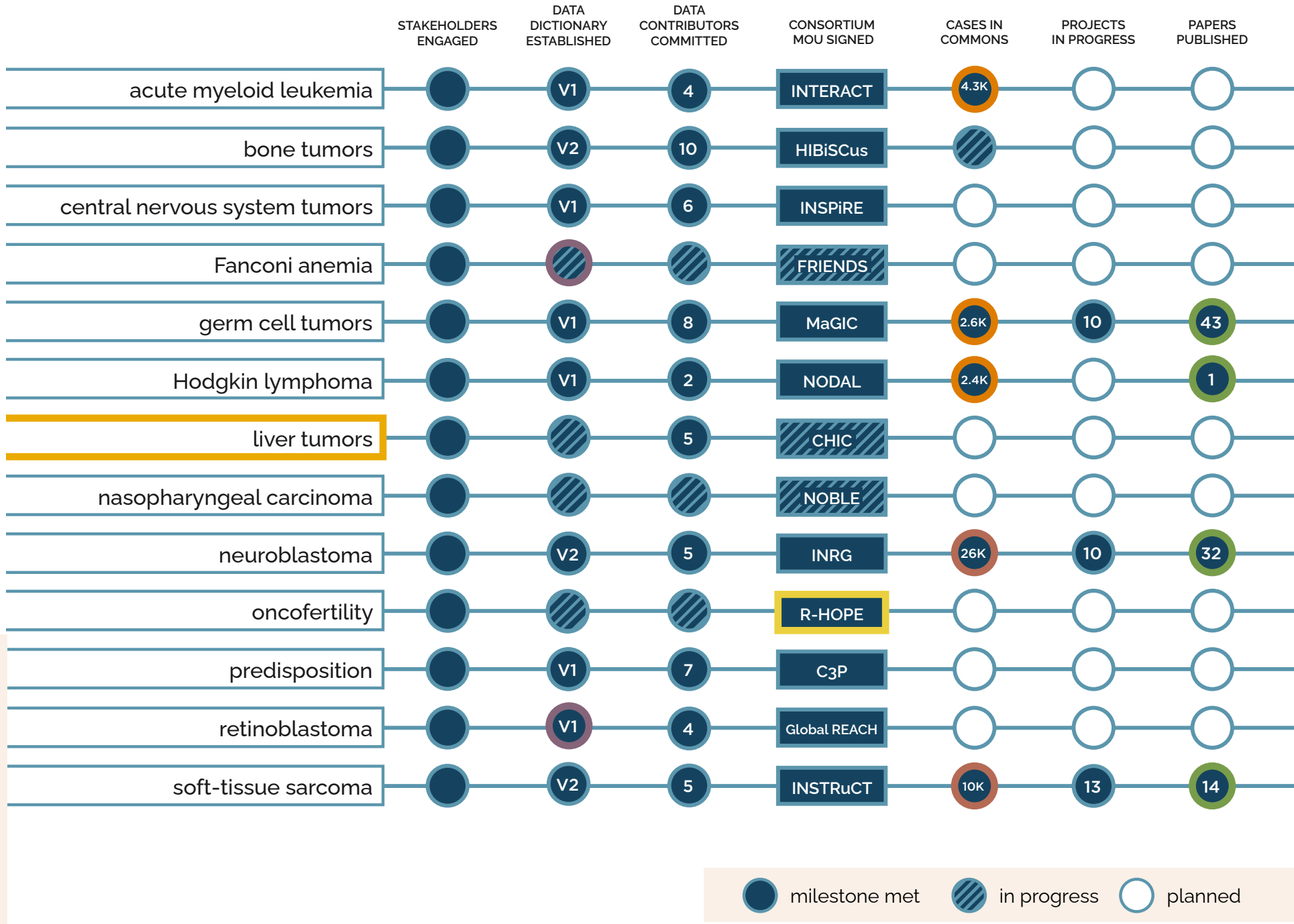
Consortia in development

We are always working to include more cancer types and related syndromes in the PCDC. While we do not yet have PCDC consortia for the following diagnoses, efforts are underway with researchers in these fields and we look forward to furthering our work together.

- Acute lymphoblastic leukemia
- Histiocytoses
- Lynch syndrome
- Non-Hodgkin lymphoma
- Pediatric myelodysplastic syndrome
- Post transplant lymphoproliferative disease
- Wilms tumor

Working behind the scenes

The hundreds of collaborators forming the PCDC's diagnosis-specific consortia continue to be the driving force behind the success of the PCDC, as they work to develop data dictionaries, share and harmonize data across institutions and borders, implement governance, and drive new science.



International advisory groups

Our international advisory groups are a key part of the PCDC ecosystem, providing diverse perspectives and experience to guide our priorities and growth. We are grateful for the individuals who contribute their time and expertise to the PCDC in this official capacity, as well as for the leadership of our consortium executive committees and the work of our many collaborators who are involved in consortia and data-contributing groups in other ways.

What's next?

Including more voices

PCDC LEAP — Pediatric Cancer Data Commons Lived Experience Advisory Panel

Our work could not exist without the partnership of patients and their loved ones. Their lived experience of pediatric cancer and treatment gives them unique insight into how we can prioritize and optimize our work to make a difference in more lives, and it's critical that their voices are included in our decision making. In addition to having patient advocates as members of our other three advisory groups and some consortium executive committees, we are now launching a fourth advisory group specifically for this purpose. The PCDC Lived Experience Advisory Panel (LEAP) will provide strategic and operational guidance for the PCDC and is composed of individuals who have a personal experience with pediatric cancer, either as a patient or family member of a patient, and who are passionate about the study of pediatric cancer. We received advice from patient advocates on the creation of the group at a planning workshop in June 2024, and the LEAP had its inaugural meeting in November 2024.

External Advisory Board (EAB)
Scientific Advisory Committee (SAC)
AYA Research Council (ARC)

86 members of
3 PCDC advisory groups in FY24

members represent
14 countries

1 group in development

In 2023-24:

- **4** SAC meetings
- **1** EAB meeting
- **2** ARC meetings
- **2** special workshops

Special workshop topics:
Artificial intelligence
Patient & family involvement

For a list of all advisory group members, please see Appendix 3.

New horizons for research and care

Innovation and impact

Leveraging the successes of the PCDC, we've continued to extend our expertise in data sharing, community-building, and technological infrastructure to other areas of impact, including both research and clinical practice. Our unique approach to building data commons is creating unprecedented opportunities to study human health, particularly for rare and difficult-to-study diseases. We are also building and expanding tools to transform clinical trial enrollment, make it easier for cancer survivors to access their health data, and enable healthcare providers to deliver more personalized care.

What's next?

More support for pediatric cancer survivors

We're now in the early stages of two projects that are focused on improving long-term follow-up care for pediatric cancer survivors.

Bright Side Navigator, a partnership with Team Bright Side, aims to address the challenges for childhood cancer survivors in their follow-up care throughout their lives. The goal of this research study is ultimately to create a widely-available ecosystem for patients, providers, and researchers to access a wide range of data and recommendations for care.

RESILIENT, a collaboration with Dr. Guergana Savova funded by Hyundai Hope on Wheels, is a project focused on using electronic health records (EHR) and natural language processing tools to address survivorship needs. We are working to extract and summarize oncology treatment exposure data from free-text notes in EHR and make the processed data available for survivorship tools like Passport for Care, reducing the need for manual data entry into these important downstream platforms.

GEARBOx

Clinical trials matching tool

GEARBOx (Genomic Eligibility AlgoRithm for Better Outcomes) is an innovative platform for clinical trials matching, developed by D4CG as part of The Leukemia & Lymphoma Society Dare to Dream Project and also supported by the Fund for Innovation in Cancer Informatics (ICI) and a grant from the NCI Informatics Technology for Cancer Research (ITCR) program.

GEARBOx is currently in use, and we are working extensively behind the scenes to further expand and improve it. This includes work to improve the accuracy of results, enhance usability, and extend the tool to additional cancer types and trials.

2023-24 highlights

- In December 2023, GEARBOx began to include studies in two new disease areas: neuroblastoma and germ cell tumors. A significant milestone, this represented the first time GEARBOx could be used for solid tumor trial matching.
- In order to make GEARBOx easier to use, we are working with clinician stakeholders to simplify the set of questions necessary to find a potential match without sacrificing accuracy.
- We are working with experts to identify and prioritize trials and matching features to add for relapse and refractory germ cell tumors, particularly for the AYA population. The scarcity of available studies for these rare cases makes this a particularly important area for matching assistance.
- We continue to refine our our methods for extracting criteria from clinical trials, testing multiple semi-automated ways to do this work. This includes a collaboration with UTHealth Houston to explore using Large Language Models (LLMs) for information extraction.

GEARBOx in action

220

registered GEARBOx users

90

active in the last six months of FY24

Explore GEARBOx: **gearbox.pedscommons.org**

PREDICT

Consortium and data commons for monogenic diabetes

Working closely with the University of Chicago Kovler Diabetes Center, PREDICT (PREcision DIabetes ConsorTium) has brought together stakeholders to build a consortium and data commons to study monogenic diabetes, a rare subtype of diabetes caused by changes to a single gene. PREDICT made significant progress this year, focused on the key steps of data dictionary development and executing data sharing agreements.

2023-24 highlights

- The Memorandum of Understanding officially establishing PREDICT was signed by ten institutional data contributors.
- With input from data standards experts and diabetes researchers, a monogenic diabetes data dictionary has been established that will be used to harmonize data for the PREDICT Data Commons.

Sociome Data Commons

Studying the social determinants of health

D4CG is part of a multidisciplinary consortium working to build the Sociome Data Commons (SDC), a research platform for large-scale data analysis including publicly-available geocoded datasets about non-clinical aspects of life affecting health, such as social, environmental, behavioral, psychological, and economic factors.

2023-24 highlights

- The SDC was expanded with the addition of 19 curated, themed datasets from 72 different data sources.
- A publication in the *Journal of Clinical and Translational Science* introduced the SDC and presented a pilot project.
- Results were presented at three national conferences.
- We expanded collaborative projects into head and neck cancer; idiopathic lung fibrosis; mortality and extreme weather events; pancreatic cancer; serous uterine cancer; lower extremity amputation; and identifying neighborhood disorder via Google Street View images and computer vision.

This work was supported by NIH award number UL1TR002389-07.

A data commons for monogenic epilepsies

Building on our successes in pediatric cancer and monogenic diabetes, we are excited to be extending D4CG data commons development to the field of pediatric monogenic epilepsies. With funding from the Chan Zuckerberg Initiative (CZI), we are poised to create the world's largest electroencephalogram (EEG) data commons for these rare and challenging epilepsies.

Why

Monogenic epilepsies, caused by individual gene mutations, are often severe and difficult to treat. A unified, high-quality data source has the potential to open new research avenues to find treatments and cures.

Who

D4CG has partnered with Dr. Doug Nordli, a leading pediatric epilepsy expert and co-director of the University of Chicago's Comprehensive Epilepsy Center, to lead the project. We have also partnered with Citizen Health as the first major data contributor to the program.

By following the example of collaborative data sharing used successfully by pediatric oncology over the past several decades we can quickly make significant strides in the care of children with monogenic epilepsies. Our focus is on EEG data, not only because of its value in classification and prognostication but also for its promise as an important biomarker that will aid in the evaluation of existing and future breakthrough therapies.

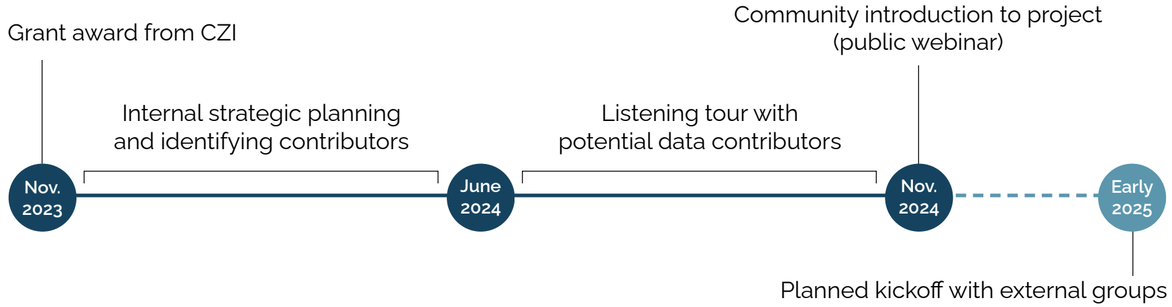
- Dr. Doug Nordli

What

This project is bringing together medical centers, registries, and patient advocacy groups to collect and harmonize clinical, genomic, EEG and other types of data, addressing the urgent need for a unified, high-quality, and highly-annotated data source.

How

D4CG will leverage the University's expertise in the quantitative analysis of EEGs to build a data commons using the same methodologies that were successful in building the PCDC. This effort will integrate raw and analyzed EEG data with genomic information and crucial clinical details to enhance the richness of these data sources and enable more types of research. In future stages, we plan to collect cognitive and outcome measures, as well as treatment data via linkage with electronic health records. This will further increase the value of the data commons and expand the scope of research it can support.



Learn more

Watch our webinar about the epilepsy data commons at **sam.am/EpilepsyVideo**

14 new publications in FY24

51 research projects and consensus papers approved or in progress

Publications from D4CG and collaborators

Automated Matching of Patients to Clinical Trials: A Patient-Centric Natural Language Processing Approach for Pediatric Leukemia in *JCO Clinical Cancer Informatics*

Sociome Data Commons: A scalable and sustainable platform for investigating the full social context and determinants of health in the *Journal of Clinical and Translational Science*

Using A Standardized Nomenclature to Semantically Map Oncology-Related Concepts from Common Data Models to a Pediatric Cancer Data Model in the *AMIA Annual Symposium Proceedings*

Data in pediatric oncology: Something old, something new in *Pediatric Blood & Cancer*

Extracting Electronic Health Record Neuroblastoma Treatment Data With High Fidelity Using the REDCap Clinical Data Interoperability Services Module in *JCO Clinical Cancer Informatics*

Big Data in Pediatric Oncology: Hope, Hype, Reality in *Advances in Oncology*

Targeted Enrollment in Pediatric Oncology Trials: A Vision for Just-in-Time Matching in *JCO Oncology Practice*

GDPR and data sharing: the Pediatric Cancer Data Commons experience in *The Lancet Oncology*

Making sense of artificial intelligence and large language models—including ChatGPT—in pediatric hematology/oncology in *Pediatric Blood & Cancer*

Publications from PCDC consortia

Adjuvant chemotherapy does not improve outcome in children with ovarian immature teratoma: A comparative analysis of clinical trial data from the Malignant Germ Cell International Consortium in *Pediatric Blood & Cancer*

INternational Soft Tissue saRcoma ConsorTium (INSTRuCT) Consensus Statement: Imaging Recommendations for the Management of Rhabdomyosarcoma in the *European Journal of Radiology*

Persistence of Racial and Ethnic Disparities in Risk and Survival for Patients with Neuroblastoma over Two Decades in *EJC Paediatric Oncology*

Consensus and controversy in the management of paediatric and adult patients with ovarian immature teratoma: the Malignant Germ Cell International Consortium perspective in *eClinicalMedicine*

Accelerating pediatric hodgkin lymphoma research: the hodgkin lymphoma data collaboration (NODAL) in the *Journal of the National Cancer Institute*

For a detailed publication list, please see Appendix 2.

Publication highlight

Imaging recommendations for soft-tissue sarcomas

Primary endpoints (the main outcomes measured in a clinical trial to determine if a treatment is effective) for patients with soft-tissue sarcoma are based on imaging-based response assessments. These assessments are subject to confounding factors such as differences in acquisition settings, methods of measurement, and interobserver variability, with little guidance provided in the clinical trial protocols for the acquisition of imaging for response assessment. To address this problem, the INSTRuCT imaging working group published imaging guidelines based on literature summaries, complemented by multidisciplinary expert opinion. They provide general imaging recommendations for both rhabdomyosarcoma and other types of soft-tissue sarcomas, as well as specific recommendations for a selection of histotypes. These guidelines will facilitate more uniform imaging acquisition and more consistent response assessments for patients with soft-tissue sarcomas on clinical trials.

Publication highlight

The Hodgkin Lymphoma Data Collaboration

In the first publication from PCDC consortium Hodgkin Lymphoma Data Collaboration (NODAL), authors from D4CG and NODAL describe the development of the consortium and their goal to harmonize pediatric Hodgkin lymphoma (HL) data to facilitate cross-trial comparisons of patients across all disease stages and subtypes. Given the high survival rates for HL, they discuss the importance of having long-term follow-up data available to study the late effects of cancer treatment, noting that a cross-disease platform such as the PCDC can offer a larger pool of data for such research, since cohorts of patients can be created according to the treatments they received, regardless of their particular cancer diagnosis. Another high priority for NODAL is to use their united data to better study rare subtypes of HL. The paper also addresses the challenges of maintaining such collaborative efforts, particularly the need for sustainable funding and the integration of new data as clinical standards evolve.

Stronger together

National Cancer Institute collaborations

D4CG and the PCDC are an important part of a national data sharing ecosystem through the National Cancer Institute's Childhood Cancer Data Initiative (CCDI). 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.

CCDI Data Federation

With a collaborative group of institutions, we are working to make the multiple data commons involved in CCDI able to interoperate by developing and implementing a common harmonized data model and application programming interface (API) for queries across the PCDC, Kids First Data Resource Center, St. Jude Cloud, and the Treehouse Childhood Cancer Data Initiative.

Data federation allows users to access data from multiple resources as if they were accessing a single virtual database. The data remain at the original source, but the API makes them searchable in one location. In a major milestone for this project, the CCDI Data Federation Resource API became available in 2024.[1]

C3DC

We are participating in developing the Childhood Cancer Clinical Data Commons (C3DC), a data node of the CCDI that acts as the primary source of deidentified individual-level data describing participants' demographic and phenotypic characteristics. C3DC interoperates with other CCDI data type-specific nodes such as genomics, imaging, and proteomics.

Recently, the datasets in the C3DC were updated, and new datasets were added, ensuring the latest information is available for research and exploration. Nodes for treatment, treatment response, and survival were integrated into the C3DC data model, enabling more refined search and analysis, and more new features were added to the user interface.[2]

1. We have received $1,146,292 in funding for this project, 100% of which is financed with federal money.
2. We have received $1,863,270 in funding for this project, 100% of which is financed with federal money.

New UChicago connections

In summer 2024, the University of Chicago announced a new Section of Biomedical Informatics in the Department of Pediatrics, led by D4CG director Sam Volchenboum. Leveraging D4CG's expertise in harnessing data to improve healthcare, the Section creates a resource at the University to foster cutting-edge biomedical research and integrate data-driven solutions into pediatric medicine to ensure the delivery of personalized, precise, and efficient care for every child.

This will create opportunities for D4CG to work closely with UChicago faculty on new projects. The first new faculty member to join the Section was Dr. Lawrence Hunter, an internationally renowned expert in knowledge-based biomedical data science, artificial intelligence for discovery in molecular biomedicine, and computational bioethics who is widely recognized as one of the founders of the field of bioinformatics. We look forward to our collaborations!

> "I'm really happy to be joining. Sam and I already have new projects going, and I look forward to helping build an exciting future for the Section."
>
> - Dr. Lawrence Hunter

The next generation

We were fortunate this year to work with a diverse group of students and interns from the University of Chicago and beyond, connecting us to the next generation of talent in our field. This group included fellows from the Data Science Institute and Harris School of Public Policy, medical students, and undergraduate interns. We also participated as a mentoring organization in Google Summer of Code, a global, online program that pairs students and beginners in open source software development with organizations to work on specific project ideas.

Thank you to our 2023-24 fellows and interns!

Sociome Data Commons
Apoorv Anand
Rishab Bhatt
Juliana Byanyima
Caleb Dimenstein
Evan Hart
Gaga He
Christopher Marasco
Alicia Severiano Perez
Gabrielle Sudilovsky
Utsav Thota

D4CG Technology
Paarth Agarwal
Natalya Cadiche
Zhen Jiang
Joshua Lukin
Oumy Mbaye
Rayat Roy
Siddharth Shastry
Wenbo Qian
Luwei Wang
Joanna Zhang

D4CG Data Standards
Simar Narula

In-person connections

We had a busy, exciting year of connecting with colleagues and sharing our work at conferences, from talks and exhibition booths to productive in-person and hybrid meetings where PCDC consortia advanced their work.

A few highlights of many included:

- We demonstrated PCDC tools and connected with colleagues at our exhibition booths at ASH, ISCAYAHL, SIOP, and SIOP Europe. (A)
- Sam Volchenboum attended the White House Rare Disease Forum, which elevated the experiences and expertise of people and families facing rare diseases in order to share developments, accelerate progress, and identify opportunities to improve research and treatment. (B)

A

B

A

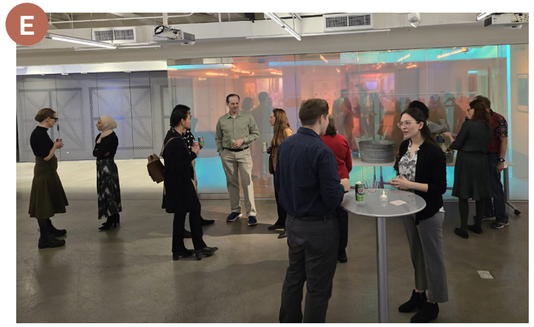
C

D

- At SIOP, Kirk Wyatt gave a talk during the Global Health session and presented a poster about how we implemented and tested safeguards during our pilot release of portal analytics tools. (C)
- Sam Volchenboum gave a keynote, and he and Suzi Birz participated in workshops, at the Paris Conference for an International Childhood Cancer Data Partnership. (D)

- At a meeting of the Leukemia & Lymphoma Society in May, Sam Volchenboum was presented with the inaugural Dr. Louis DeGennaro Innovation Award.
- At ASPHO, Kirk Wyatt gave a workshop on demystifying artificial intelligence.
- At SIOP Europe, Sam Volchenboum gave a talk on bringing together large datasets in AYA cancer.
- We hosted a D4CG Spring Social here in Chicago to connect with our colleagues and friends. (E)

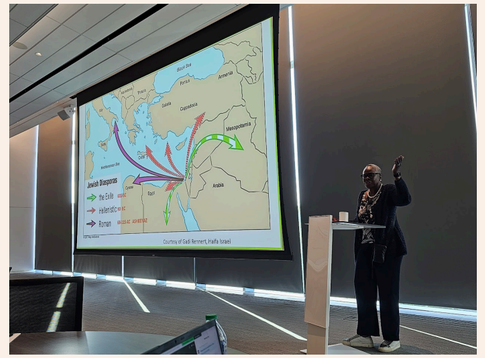
E



ITCR at D4CG

D4CG had the honor of hosting the NCI Informatics Technology for Cancer Research (ITCR) Annual Meeting in September 2023 at the University of Chicago. National Cancer Institute grant recipients and program leaders from across the United States joined us in Chicago and virtually to share their scientific work, hear keynote speakers, and participate in workshops and social events. Thanks to our sponsors Globus, the University of Chicago Department of Pediatrics, and the University of Chicago Comprehensive Cancer Center for helping to make this meeting possible.

Special address from Senator Dick Durbin



Keynote from Dr. Funmi Olopade

Fueling our future

The progress outlined in this report has been made possible thanks to the generous partnership of our funders. Here's a look at what your contributions helped to achieve this year. Thank you for believing in our vision and being an essential part of our team.

Cancer Research Foundation helped to diversify and grow the PCDC by supporting our international outreach.

Children's Cancer Research Fund continued to support the work of our retinoblastoma consortium Global REACH as they made progress that included a completed data dictionary.

Through ***Children's Oncology Group Foundation***, we received support from a coalition of four foundations. ***Friends of T.J. Foundation***, ***Maddie's Promise***, ***SebastianStrong***, and ***Summer's Way Foundation*** joined together to provide funding for a statistician to work with PCDC soft-tissue sarcoma data from INSTRuCT, as well as support for INSTRuCT project management.

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

Comer Development Board helped to sustain the centralized project management, governance, and engagement operations of the PCDC.

Fanconi Cancer Foundation supported the development of FRIENDS and their work to include Fanconi anemia data in the PCDC.

The Fund for Innovation in Cancer Informatics continued their support of our work to improve and expand the GEARBOx clinical trials matching tool.

Gray Foundation continued to support the PREDICT monogenic diabetes consortium as they developed their data dictionary and secured data contributors.

Jeffrey Pride Foundation and ***Aileen S. Andrew Foundation*** supported the Pediatric Cancer Data Commons through Project Every Child.

The Leukemia & Lymphoma Society, as part of its Dare to Dream Project, continued to provide essential support for both our AML consortium INTERACT and the use and further development of GEARBOx.

MIB Agents, through an Outsmarting Osteosarcoma grant possible because of Annaleigh, provided support for our bone tumors consortium HIBiSCus as they harmonize osteosarcoma data.

A contract with the ***National Cancer Institute*** and ***Leidos*** enabled D4CG to participate in developing the Childhood Clinical Data Commons (C3DC), a data node of the CCDI.

Another contract with the ***National Cancer Institute*** allowed us to work on developing and implementing a harmonized data model across multiple CCDI data commons.

A grant from the ***NCI Informatics Technology for Cancer Research (ITCR)*** program supported our continued improvement and expansion of GEARBOx,

The ***Neuroblastoma Children's Cancer Society*** provided funding for INRG, with a focus on efforts to better understand and treat secondary illness and side effects in neuroblastoma survivors.

Rally Foundation for Childhood Cancer Research supported our bone tumors consortium HIBiSCus as they harmonized data, as well as supporting project management for the PCDC as a whole.

The ***Sarah Jane Adicoff Endowment for Research in Rhabdomyosarcoma***, through ***Seattle Children's Foundation***, provided support for INSTRuCT.

St. Baldrick's Foundation supported the PCDC's central operations, including governance, project management, and engagement, as well as enabling our partnership with C3P.

Team Bright Side continued their support of our ALL consortium and the planning and development of our Bright Side Navigator long-term follow-up project.

An anonymous foundation continued to support our development of consensus data dictionaries.

In addition to the funding highlighted here, which was received during the 2023-24 fiscal year, we are grateful and excited to welcome these organizations supporting our work in 2024-25 and beyond. We look forward to reporting more next year on the advances made possible by your support!

- **Chan Zuckerberg Initiative**
- **Hyundai Hope on Wheels**

Stronger together

The Data for the Common Good team brings together a wide range of backgrounds and expertise, united by a shared commitment to harnessing the power of community and data to improve human health.

Suzi Birz, MScMI, LFHIMSS
Regulatory and Data Governance Consultant

Kat Bouzein, MS
Program Director, PCDC

Lauren Chan, PhD, RD
Postdoctoral Researcher

Seong Choi
Senior Software and Data Integration Engineer

Wanda Clark-Jenkins
Administrative Specialist

Spencer Claxton
Front End Developer

Ellen Cohen, MPP
Deputy Director

Bailey Farrow
Program Director, Epilepsy

Nicholas Ferraz, MS
Full Stack Developer

Brian Furner, MS
Senior Director of Data and Technology

Luca Graglia, MS
Director of Software and Infrastructure Services

Zari Henry
Research Coordinator

Hubert Hickman, MS
Senior Data Strategist and Data Operations Lead

Enal Hindi, MS
Senior Project Manager

Steve Krasinsky
Full Stack Developer

Jooho Lee, PhD, MSLIS
Semantic Engineer

Mei Li, MS
Healthcare Data Standards Analyst

Maya Maric
Senior Project Manager

Paul Murdoch
Junior Back End Developer

Kaitlyn Ott, MS
Healthcare Data Standards Analyst

Rolando Palacios, MS
Technical Project Manager

Caitlin Pike
Director of Communications

Thivashnee Sharma, MBA
Senior Project Manager

Sandra Tilmon, MS, MPH
Healthcare Data Scientist

Sam Volchenboum, MD, PhD
Director and Pediatric Oncologist

Michael Watkins, PhD
Manager of Data Standards and Modeling

Kirk Wyatt, MD
Senior Clinical Advisor

Liz Zampino, MS
Program Director, Sociome

Tianyun Zhang, MS
Senior Front End Developer

Stay connected!

Visit our website

commons.uchicago.edu

Explore PCDC tools

pedscommons.org

Sign up for our newsletter and conference previews

sam.am/D4CGnews

Connect with us on social media

linkedin.com/company/d4cg

bsky.app/profile/d4cg.bsky.social

Questions about getting involved or supporting D4CG?

Email Kat Bouzein at kblumhardt@uchicago.edu

Thank you.

Our community is our greatest strength. We are grateful to our consortium members and data contributors, data portal users, advisory group members, funders and partners, University of Chicago and Biological Sciences Division colleagues, and all those working alongside us to improve human health. Most of all, we thank the patients and families without whom our work would not be possible.

Appendix 1: Consortium Names

C3P - Consortium for Childhood Cancer Predisposition

CHIC - Children's Hepatic Tumors International Collaboration

FRIENDS - Fanconi Research Initiative for Education, Networking, and Data Sharing

Global REACH - Global REtinoblastoma Alliance for CHildren

HIBiSCus - Harmonization International Bone Sarcoma Consortium

INRG - International Neuroblastoma Risk Group

INSPiRE - International central Nervous System Pediatric REsearch consortium

INSTRuCT - INternational Soft Tissue saRcoma ConsorTium

INTERACT - INTERnational pediatric Acute myeloid leukemia ConsorTium

MaGIC - Malignant Germ cell International Consortium

NOBLE - NasOpharyngeal carcinoma gloBaL partnErship

NODAL - hodgkiN lymphOma DatA coLlaboration

PREDICT - PREcision DIabetes ConsorTium

Reproductive HOPE - Reproductive Health Outcomes and Preservation Evaluation

Appendix 2: Publications, July 2023-June 2024

Kaskovich S, Wyatt KD, Oliwa T, et al. Automated Matching of Patients to Clinical Trials: A Patient-Centric Natural Language Processing Approach for Pediatric Leukemia. *JCO Clin Cancer Inform.* 2023 Jul:7:e2300009. doi: 10.1200/CCI.23.00009

Vieira AGS, Puthenpura V, Krailo M, et al. Adjuvant chemotherapy does not improve outcome in children with ovarian immature teratoma: A comparative analysis of clinical trial data from the Malignant Germ Cell International Consortium. *Pediatr Blood Cancer.* 2023 Aug;70(8):e30426. doi: 10.1002/pbc.30426

Schoot RA, van Ewijk R, von Witzleben A, et al. INternational Soft Tissue saRcoma ConsorTium (INSTRuCT) Consensus Statement: Imaging Recommendations for the Management of Rhabdomyosarcoma. *Eur J Radiol.* 2023 Sep:166:111012. doi: 10.1016/j.ejrad.2023.111012

Tilmon S, Nyenhuis S, Solomonides A, et al. Sociome Data Commons: A scalable and sustainable platform for investigating the full social context and determinants of health. *J Clin Transl Sci.* 2023 Nov 7;7(1):e255. doi: 10.1017/cts.2023.670

Chennakesavalu M, Pudela C, Applebaum MA, et al. Persistence of Racial and Ethnic Disparities in Risk and Survival for Patients with Neuroblastoma over Two Decades. *EJC Paediatr Oncol.* 2023 Dec:2:100022. doi: 10.1016/j.ejcped.2023.100022

Carlson B, Watkins M, Li M, Furner B, Cohen E, Volchenboum SL. Using A Standardized Nomenclature to Semantically Map Oncology-Related Concepts from Common Data Models to a Pediatric Cancer Data Model. *AMIA Annu Symp Proc.* 2024 Jan 11:2023:874-883. PMID: 38222364

Pashankar F, Murray MJ, Gell J, et al. Consensus and controversy in the management of paediatric and adult patients with ovarian immature teratoma: the Malignant Germ Cell International Consortium perspective. *EClinicalMedicine.* 2024 Feb;69:102453. doi: 10.1016/j.eclinm.2024.102453

Wyatt KD. Data in pediatric oncology: Something old, something new. *Pediatr Blood Cancer.* 2024 Feb;71(2):e30769. doi: 10.1002/pbc.30769

Furner B, Cheng A, Desai AV, et al. Extracting Electronic Health Record Neuroblastoma Treatment Data With High Fidelity Using the REDCap Clinical Data Interoperability Services Module. *JCO Clin Cancer Inform.* 2024 May:8:e2400009. doi: 10.1200/CCI.24.00009

Wyatt KD, Birz S, Castellino SM, et al. Accelerating pediatric hodgkin lymphoma research: the hodgkin lymphoma data collaboration (NODAL). *J Natl Cancer Inst.* 2024 May 8;116(5):642-646. doi: 10.1093/jnci/djae013

Wyatt KD, Volchenboum SL. Big Data in Pediatric Oncology: Hope, Hype, Reality. *Adv Oncol.* 2024 May;4(1):91-99. doi: 10.1016/j.yao.2024.02.005

Wyatt KD, Volchenboum SL. Targeted Enrollment in Pediatric Oncology Trials: A Vision for Just-in-Time Matching. *JCO Oncol Pract.* 2024 May;20(5):603-606. doi: 10.1200/OP.23.00826

Wyatt KD, Minard-Colin V, Schleiermacher G, Willi M, Volchenboum SL. GDPR and data sharing: the Pediatric Cancer Data Commons experience. *Lancet Oncol.* 2024 June;25(6):e227. doi: 10.1016/S1470-2045(24)00250-X

Wyatt KD, Alexander N, Hills GD, et al. Making sense of artificial intelligence and large language models—including ChatGPT—in pediatric hematology/oncology. *Pediatr Blood Cancer.* 2024 Sep;71(9):e31143. doi: 10.1002/pbc.31143.

Appendix 3: PCDC Advisory Group Members

Scientific Advisory Committee

Todd Alonzo
Maja Beck-Popovic
Gianni Bisogno
Susan L. Cohn
Jamie Flerlage (co-chair)
Brenda Gallie
Ajay Gupta
Sumit Gupta
Darren Hargrave
Doug Hawkins
Stefanie Hecker-Nolting
Katie Janeway
Pamela Kearns
Kara Kelly
Ted Laetsch
Sarah Leary
Marie-Cécile Le Deley
Mignon Loh
Akira Nakagawara
Farzana Pashankar
Andy Pearson
Christopher Porter
Praful Ravi
Dirk Reinhardt
Lainie Ross
Erin Rowell
Gudrun Schleiermacher
Furqan Shaikh
Monika Sparber-Sauer
Daisuke Tomizawa (co-chair)
Liliana Vásquez
Anita Villani
Michaela Willi

External Advisory Board

Samuel Blackman (co-chair)
Peggy Bodin
Dana Callow
Hubert Caron
Ellen Clayton
Jacques Demotes-Mainard
Joe Depa
Derek Groothuis
Nur-Ul Haq
Minke Huibers
Parker Moss (co-chair)
Gregory Reaman
Sarah Rostock
Carlos Sandi
Dominik Schneider
Greg Simon
Henry Ting
Gilles Vassal

AYA Research Council

Todd Alonzo
Elysia Alvarez
Carolyn Breinich
Anne-Sophie Darlington
Uta Dirksen
Doug Fair
Andrea Ferrari
Lindsay Frazier
Sam Funt
Hilary Gan
Tara Henderson
Olga Husson
Tyler Ketterl
Suzanne MacFarland
Martin McCabe
Jennifer McNeer (co-chair)
Chaya Moskowitz
Mariana Nana
Megan Othus
Ruzanna Papyan
Max Penzer
Apostolos Pourtsidis
Betty Roggenkamp
Mike Roth (co-chair)
Mary Sammel
Jessica Schulte
Lisa Schwartz
Jennifer Seelisch
Chelsea Self
Dan Stark (co-chair)
Sandra Strauss
Winette van der Graaf
Gabriela Villanueva (co-chair)
Kurt Weiss
AnnaLynn Williams
Aaron Yeo

Lived Experience Advisory Panel

Sarah Bartosz
Kim Buff
Evan Connelly
Sloane Dyer
Mindy Dykes
Mariah Forster-Olson
Parker Moss
Max Penzer
Carlos Sandi
Johnny Schroepfer
Greg Simon
Eshan Vishwakarma
Michaela Willi

THE UNIVERSITY OF CHICAGO

DATA FOR THE COMMON GOOD