Agency Priority Goal | Action Plan | FY 2026 – FY 2027

Embracing AI to Combat Childhood Cancer

Goal Leader: Warren A. Kibbe, Deputy Director for Data Science and Strategy, The National Cancer Institute (NCI)

Deputy Goal Leader: Joseph D. Rogers, Lead Health Scientist (Informatics), Division of Cancer Prevention and Control (DCPC), Cancer Surveillance Branch (CSB), Centers for Disease Control and Prevention (CDC)

Goal Overview

Goal statement

Embracing AI to combat Childhood Cancer: NCI and CDC aim to advance research on pediatric cancers by ensuring that data in the Childhood Cancer Data Initiative (CCDI) and National Program of Cancer Registries (NPCR) data modernization priorities are AI-ready through enhanced integration and interoperability, enabling investigators to leverage AI with these data for advanced and innovative visualization and analysis.

- Improve data integration by increasing the number of cancer cases shared and the number of organizations sharing. Aiming for 30,000 number of cancer cases shared and at least 6 organizations participating by October 2027.
- Improve data interoperability of genomic data and incorporate additional data (diagnosis, germline predisposition, treatment, Adverse Events (AEs), clinical outcomes) as available. By October 2027 there will be genomic data from 8,500 children, with at least 1,200 new cases added per year, and 3,000 whole slide pathology images from tumors of pediatric cancer patients. These data are important for training AI models and discovering new opportunities to improve diagnosis, prognosis, treatment, and long-term outcomes for children with cancer.
- Connect 2 additional data discovery and analysis tools to the CCDI Data Hub by October 2027.
- Support or perform 4 integrated AI data analysis projects by October 2027 and ensure the data and results generated are deposited into the CCDI Ecosystem.
- In collaboration with CDC, improve the quality and scalability of extracting clinical data for integration into the CCDI Ecosystem, incorporating new childhood cancer-relevant data domains, like the College of American Pathologists (CAP) Cancer Protocol and United States Core Data for Interoperability+ Cancer (USCDI+) by October 2027.

Problem to Be Solved

While childhood cancers are the leading cause of death by disease in children over the age of 1, they are collectively rare. Childhood cancers make up about 1%–3% of cancers diagnosed annually in the United States. Because individual pediatric cancer types are uncommon and patient data are often fragmented across hospitals, research institutions, registries, and clinical systems, it remains difficult to answer critical scientific and clinical questions at a national scale. Fragmented and inconsistently-structured data limit the ability to rapidly identify pediatric cancer cases, integrate molecular and clinical findings, longitudinally follow survivors, and fully leverage AI and advanced analytics to improve diagnosis, treatment, and outcomes. Differences in data standards and interoperability across systems further slow the generation of actionable insights.

CCDI and CDC's pediatric cancer data modernization activities aim to address these challenges by creating interoperable, longitudinal, and AI-ready data ecosystems that support national-scale learning from every child, adolescent, and young adult (AYA) with cancer. These efforts are intended to accelerate discovery, improve diagnosis and treatment decision-making, enhance survivorship understanding, and support broader collaboration across the pediatric cancer community.

What Success Looks Like

Comprehensive and meaningful data sets are necessary to learn from and develop better treatments and outcomes for all pediatric and AYA patients experiencing cancer. CCDI funding is supporting research studies and tools that drive our ability to use data effectively, including establishing cohorts that can help answer key scientific questions in childhood cancer. The initiative continues to evolve its programs, resources, and tools to make them as effective and helpful as possible, with community input and priorities in mind. The CDC, in support of the Childhood Cancer Survivorship, Treatment, Access, and Research (STAR) Act, developed a secure cloud-based, AI-driven infrastructure known as the National Program of Cancer Registries (NPCR) National Oncology Rapid Ascertainment Hub (NOAH). This system enables rapid case reporting for childhood and young adult cancers and has been piloted in six NPCR-funded cancer registries. NPCR NOAH supports electronic pathology reporting from laboratories to registries, using AI-enabled processes to enhance data extraction and transmission, improve the timeliness and efficiency of data collection, and allow registries to identify and report childhood cancer cases more quickly.

AI-enabled analyses supported through CCDI and NPCR are intended to augment, and not replace, clinical expertise by helping clinicians and researchers identify patterns, refine diagnoses, and generate evidence-informed insights that can support decision-making and future research.

Success is demonstrated by measurable progress in data sharing, integration, and standardization, along with faster and more accurate pediatric cancer case identification and increased use of these data and AI tools to generate new insights and improve pediatric health outcomes.

Strategic Alignment (HHS, NIH/NCI & CDC)

This collaboration directly aligns with HHS, the National Institutes of Health (NIH)/NCI, and CDC priorities to modernize pediatric cancer data infrastructure, accelerate responsible use of AI in biomedical research, and improve outcomes for children and AYAs with cancer.

CCDI was launched following President Trump's 2019 State of the Union call to "learn from every child" with cancer and continues to advance that vision through nationwide pediatric cancer data sharing, molecular characterization, survivorship research, and ecosystem development. The strategic direction of CCDI is also informed by and directly

aligned with the 2025 Executive Order “Unlocking Cures for Pediatric Cancer with Artificial Intelligence,” which calls for accelerating pediatric cancer research and treatment through AI-enabled analysis, interoperable data infrastructure, and advanced computational approaches.

In coordination with CDC and activities supported under the Childhood Cancer STAR Act, CCDI is helping strengthen national pediatric cancer data resources through efforts such as molecular profiling, biospecimen collection, and survivorship data integration. Together, these efforts support development of a longitudinal pediatric cancer learning health system capable of enabling AI-supported discovery, improving diagnosis and treatment decision-making, and advancing survivorship research.

Tracking the goal

Goal Target(s)

The column in **blue** will reflect the quarterly update except for the initial action plan.

	We will…	**Name of indicator** (units in parentheses)	Start value (Date)	**Current value**	**Target value**
1	Improve data integration by increasing the number of cancer cases shared through the CCDI Hub and the number of federated organizations/ecosystems sharing. [NCI]	(Number of cancer cases shared and number of organizations participating)	23,802 cases; 4 orgs. (October, 2025)	24,621; 4 orgs.	30,000 cancer cases shared; 6 organizations
2	Improve data interoperability of genomic data and incorporate additional data (diagnosis, germline predisposition, treatment, AEs, outcomes clinical) as available. [NCI]	(Genomic data cases and whole slide pathology images from tumors of pediatric cancer patients)	7,023 genomic cases; 3,908 WSI (October, 2025)	7,842 cases; 4,407 WSI	8,500 cases 5,000 whole slide pathology images
3	Connect additional data discovery and analysis tools to the CCDI Data Hub. [NCI]	(number of additional tools)	N/A	1 suite of tools (MOSSAIC): data structuring, pathology reporting, and imaging resources	2
4	Support or perform integrated AI data analysis projects and ensure the data and results generated are deposited into the CCDI Ecosystem. [NCI]	(number of projects)	N/A	1	4
5	Improve the quality and scalability of extracting clinical data for integration into the CCDI Ecosystem, incorporating new childhood cancer-relevant data domains through efforts such as the CAP Cancer Protocol and USCDI+ Cancer. [NCI + CDC]	(qualitative progress or turn this into milestones?)	N/A	TBD (baseline needed)	CDC/NPCR will consult with the College of American Pathologists (CAP) to identify the number of pathology laboratories currently utilizing the CAP Cancer Protocol.

6	Increase **adoption** of Association of Public Health Laboratories (APHL) Informatics Messaging Services Platform (AIMS) based pathology reporting by labs to support timely identification of cancer cases [NCI SEER + CDC]	(number of labs utilizing APHL AIMS for reporting pathology reports)	N/A	TBD (baseline needed)	CDC/NPCR will secure access to the APHL AIMS dashboard to obtain necessary data.
7	Increase electronic pathology reporting through APHL AIMS by labs to improve the **timeliness (≤ 30days)** of cancer case identification. [NCI SEER + CDC]	(number of e-reports sent to APHL AIMS by labs ≤30 days)	N/A	TBD (baseline needed)	CDC/NPCR will secure access to the APHL AIMS dashboard to obtain the necessary data
8	**Quantify** the number of pediatric cancer cases identified from the e-reports sent to APHL AIMS by NPCR NOAH at the central cancer registries (CCRs) and confirmed through validation. (*NPCR NOAH is a secure cloud-based, AI-driven infrastructure.)* [CDC]	(number of true pediatric cancer cases identified by NPCR NOAH at the central cancer registries (CCRs) after validation)	N/A	TBD (baseline needed)	CDC/NPCR will work with recipients to identify central cancer registries (CCRs) using NPCR NOAH and to determine pediatric cancer cases based on APHL AIMS reports.

* Qualitative progress may be noted as “Yes, we achieved it”, “in-progress”, or “No, not yet.”

Summary of Progress– FY 2026 Quarter 1 and 2

Goal 1 details: Improve data integration by increasing the number of cancer cases shared through the CCDI Hub and the number of organizations/ecosystems sharing.

TARGET

- 30,000 cancer cases shared through CCDI Hub
- 6 organizations/ ecosystems

UPDATE

- 24,621 cases (CCDI Hub)

Add'l cases (collaborations/federation):

- 30,717 cases (CCSS; St. Jude)
- 5,284 cases (KidsFirst; CRDC; CHOP DRC)
- 304 models (PIVOT; Jackson Labs)

The CCDI Hub shares pediatric and AYA cancer data and resources at multiple levels, including detailed row-level data, aggregate data, and federated datasets from partner ecosystems. Resources include clinical, survivorship, and molecular data, as well as models, and tools generated through both CCDI-led and broader NCI-supported initiatives.

- **Data from CCDI-led initiatives**: Data have been generated through efforts like the Molecular Characterization Initiative (see goal 2) and collected from CCDI-led studies to provide a foundation for learning from every child with cancer. We have directly funded and led research to generate and gather comprehensive, row-level case data (24,621). These efforts include data gathered from more than 25 Cancer Centers.
- **Data from CCDI-supported/NCI-led initiatives**: Resources and data benefit from the infrastructure and framework established in the Hub. (No specific APG targets for these; numbers displayed on the CCDI Hub include these participants.)
- **Data searchable through federated partners**: The Hub also makes available data through federated repositories/ecosystem partners, some of which may also reside within the CCDI Hub while some data are only available at the federated partner. (No specific APG targets for these; numbers displayed on the CCDI Hub include these participants.) CCDI has piloted federation with four platform partners to-date to develop an initial API.

The extensive data resources available to researchers support a comprehensive national Learning Health System for pediatric cancer research. We are on-track to reach our targets for this goal.

Goal 2 details: Improve data interoperability of genomic data and incorporate additional data (e.g., diagnosis, germline predisposition, treatment, Adverse Events, clinical outcomes) as available.

Molecular characterization (clinical sequencing) data are available to the research community from an additional 819 patients since October 2025. Of the total 7,842 cases, research sequencing for these patients is currently being processed, which will create a comprehensive molecular profile that can be used to inform diagnostics and identify therapeutic targets.

An additional 499 whole slide images are available from the same timeframe. We are on-track to reach our targets for this goal.

Expansion of whole slide pathology imaging within the CCDI ecosystem creates new opportunities for AI-enabled image analysis and multimodal learning approaches that integrate pathology, molecular, and clinical data. These efforts may support improved diagnostic refinement, identification of novel disease patterns, improved prognostic modeling, and enhanced understanding of treatment response across pediatric cancers.

Goal 3 details: Connect additional data discovery and analysis tools to the CCDI Data Hub.

One new suite of tools has been added to the CCDI Hub (MOSSAIC): data structuring, pathology reporting, and imaging resources. These are in addition to:

- Exploration tools for CCDI platforms (NCCR*Explorer, cBioPortal, C3DC)
- Knowledgebases (Molecular Targets Platform, CiVIC)
- Toolsets leveraged from NIH/NCI platforms (dbGaP/SRA, caDSR, Cancer Genomics Cloud, SEER tools)

We are on-track to reach our targets for this goal.

Goal 4 details: Support or perform integrated AI data analysis projects and ensure the data and results generated are deposited into the CCDI Ecosystem.

On September 30, 2025, the Executive Order “Unlocking Cures for Pediatric Cancer with Artificial Intelligence” established a national priority to accelerate pediatric cancer research and treatment through responsible application of AI technologies. CCDI directly supports this effort by developing interoperable, longitudinal, and AI-ready pediatric cancer datasets that can support advanced analytics across genomic, imaging, pathology, clinical, treatment, and survivorship domains.

CCDI is working to enable AI-supported approaches that improve data structuring, diagnostic refinement, predictive modeling, biomarker discovery, treatment response analysis, and survivorship research. Current and planned activities include AI-enabled pathology and imaging analysis, machine learning approaches for harmonization and data extraction, integration of

multimodal datasets, and support for external collaborators developing AI tools using CCDI resources. CCDI-supported AI efforts are intended to accelerate discovery while supporting transparency, reproducibility, responsible data use, and clinician-informed interpretation.

This component of CCDI is still in early stages. CCDI is working with the Department of Energy on the RADIANCE project, specifically on machine learning and LLM for data structuring and pathology imaging and analysis pipelines.

Goal 5 details: Improve the quality and scalability of extracting clinical data for integration into the CCDI Ecosystem, incorporating new childhood cancer-relevant data domains through efforts such as the CAP Cancer Protocol and USCDI+ Cancer. [NCI + CDC]

Beginning with cancer cases diagnosed January 1, 2024 and forward, new data items for Pediatric and Adolescent/Young Adult populations are available for use in hospital or central cancer registries. The staging elements collected are based on Toronto Childhood Cancer Staging Guidelines v2, along with additional data for cancer surveillance and include pediatric primary tumor; pediatric regional nodes; and pediatric metastasis. SEER central cancer registries are required to submit these data items when available. Availability of these data items in vendor systems will increase consistent and structured capture of pediatric staging.

USCDI+ Cancer, a collaboration between ONC, NCI, and CDC, has defined a core set of cancer research-specific data concepts to enhance data exchange and harmonization with the goal of accelerating cancer research, prevention, diagnosis, treatment, and care. USCDI+ Cancer published the Cancer Registry draft data element list, which is the minimum necessary data set to identify and extract required data and support the current data sharing and linkage approaches for SEER and NPCR registries. NCI and CDC are currently working on finalizing and testing a FHIR implementation guide for these data elements. Adoption of the USCI+ Cancer Registry data element list and FHIR implementation guide will also benefit more timely and structured reporting of pediatric cancer data from electronic health records.

Goal 6 details: Increase adoption of APHL AIMS based pathology reporting by labs to support timely identification of cancer cases [NCI SEER + CDC]

Goal 7 details: Increase electronic pathology reporting through APHL AIMS by labs to improve the **timeliness (≤ 30days)** of cancer case identification. [NCI SEER + CDC]

Goal 8 details: Quantify the number of pediatric cancer cases identified from the e-reports sent to APHL AIMS by NPCR NOAH at the central cancer registries (CCRs) and confirmed through validation. *NPCR NOAH is a secure cloud-based, AI-driven infrastructure.* [CDC]

Goal Strategies

NCI's CCDI and CDC's NPCR data modernization priorities including NPCR NOAH implementation, represent a national commitment to harnessing and sharing data in ways to make faster progress in childhood and young adult cancers. These initiatives have made significant progress on all foundational goals and the work continues. The goals are:

- gathering data from every child and AYA diagnosed with a childhood cancer, regardless of where they receive their care
- creating a national strategy of appropriate clinical and molecular characterization to speed diagnosis and inform treatment for all types of childhood cancers
- developing a platform and tools to bring together clinical care and research data that will improve preventive measures, treatment, quality of life, and survivorship for childhood cancers

Goal 1 - Improve data integration by increasing the number of cancer cases shared through CCDI Hub and the number of organizations/ecosystems sharing: The CCDI Hub serves as a centralized entry point for researchers to discover, explore, and access pediatric cancer data, tools, and resources across the CCDI ecosystem. Through integrated applications such as the Explore Dashboard and the CCDI Data Federation Resource, users can search deidentified participant-level genomic, clinical, imaging, and biospecimen data across multiple studies and institutions, create virtual cohorts, and identify relevant datasets using standardized metadata and APIs. Rather than moving data into a single repository, CCDI uses a federated data model in which data remain at their original source but are made searchable and interoperable through agreed-upon, shared CCDI standards and APIs. The federation APIs provide open-access metadata and dataset locations, while access to underlying data files is governed by each contributing resource's data access policies.

Goal 2 - Improve data interoperability of genomic data and incorporate additional data (diagnosis, germline predisposition, treatment, Adverse Events, clinical outcomes) as available: The CCDI Molecular Characterization Initiative (MCI) provides comprehensive molecular profiling at no cost for children, adolescents, and young adults with select pediatric cancers treated at Children's Oncology Group (COG) affiliated institutions. Cancers studied include central nervous system (CNS) tumors, soft tissue sarcomas (STS), rare childhood cancers, high-risk neuroblastoma (NBL), and metastatic Ewing sarcoma (EWS). Eligible patients are enrolled through Project: EveryChild (APEC14B1), which supports patient consent, biospecimen collection, and clinical data annotation. Submitted inputs include tumor tissue, matched normal specimens, and associated clinical and phenotypic data. These samples undergo CLIA-certified DNA and RNA sequencing, along with methylation array profiling, to identify clinically relevant molecular alterations. Clinical reports are returned to treating physicians within two weeks to support diagnosis, treatment planning, and clinical trial eligibility. Early important findings from this project include discovery of 14% of the children analyzed have germline mutations, children with ultra rare pediatric cancers are nearly twice as likely to have germline mutations, highlighting the value of

genetic counseling and cascade testing for families affected by these cancers. In addition, 11% of the children profiled were enrolled on clinical trials based on their profile results. A similar percentage, just over 10% of the children profiled, were treated with a targeted therapy based on their profile results. To support the goal of learning from every child with cancer, molecular profiling data are submitted to the CCDI Ecosystem monthly, while associated clinical and phenotypic data are contributed every six months. This will support ongoing pediatric cancer research and improve understanding of disease biology. For an overview of this initiative**,** access the MCI web page on cancer.gov.

Goal 3 - Connect additional data discovery and analysis tools to the CCDI Data Hub.

Through the CCDI Hub and its Explore Dashboard, researchers can discover, filter, visualize, and access CCDI-managed datasets alongside linked platforms, analytical tools, and complementary resources that support childhood and adolescent and young adult (AYA) cancer research. The CCDI ecosystem is designed to support a wide range of use cases, including data discovery, cohort identification, longitudinal and survivorship analyses, harmonization and structuring of heterogeneous datasets, complex and federated search functions, cross-study comparison, and integration with analytical workflows and AI-enabled methods. Recognizing that users have varying technical expertise and research needs, CCDI aims to provide data access and functionality at multiple levels of complexity, from intuitive visualization and cohort exploration tools to advanced computational and analytical environments.

To advance a more integrated, interoperable, and accessible research ecosystem and support the development of a learning health system, CCDI will expand and incorporate tools and capabilities that enable a broad range of stakeholders to effectively understand, access, and use the data. These stakeholders include clinicians, data scientists, computational biologists, basic and translational researchers, public health researchers, patients, families, and advocacy communities. CCDI is actively working with partners who have developed relevant tools to incorporate into the ecosystem.

Goal 4 - Support or perform integrated AI data analysis projects and ensure the data and results generated are deposited into the CCDI Ecosystem.

On September 30, 2025, an EO titled "Unlocking Cures for Pediatric Cancer with Artificial Intelligence" was signed to significantly accelerate pediatric cancer research and treatment by leveraging AI technologies. Informed by and in alignment with the EO, CCDI is working to apply AI to analyze complex biological datasets, improve predictive modeling for patient responses, and identify new biomarkers for better diagnosis and treatment across diverse data sets (e.g., genomic, imaging, clinical records). Key priorities include optimizing clinical trials, improving diagnostics, increasing data privacy protections, facilitating equitable access to advancements, and transforming both the CCDI data ecosystem and the diverse data held in that ecosystem to be AI-ready. CCDI has already begun supporting existing NCI AI-related activities and will continue to develop new initiatives in support of the EO.

Future CCDI activities will include workforce development efforts focused on pediatric cancer data science and AI literacy, use of AI for novel research, and supporting responsible and effective incorporation of AI tools and techniques into the CCDI Ecosystem for use across the cancer community.

Goal 5 - Improve the quality and scalability of extracting clinical data for integration into the CCDI Ecosystem, incorporating new childhood cancer-relevant data domains, through efforts such as the CAP Cancer Protocol and USCDI+ Cancer. [NCI + CDC]

Goal 6 - Increase adoption of APHL AIMS based pathology reporting by labs to support timely identification of cancer cases [NCI SEER + CDC]

Goal 7 - Increase electronic pathology reporting through APHL AIMS by labs to improve the **timeliness (≤ 30days)** of cancer case identification.
[NCI SEER + CDC]

Goal 8 - Quantify the number of pediatric cancer cases identified from the e-reports sent to APHL AIMS by NPCR NOAH at the central cancer registries (CCRs) and confirmed through validation. [CDC]

Key Indicators

To be added in future reports, if required; our metrics are our main key indicators.

Key Milestones

Key Milestone	Milestone Due Date	Milestone Status	Owner	Comments [Optional]
Incorporate tools developed through ARPA-H/NIH collaboration for Biomedical Data Framework	Q1, FY 27	On-track	NCI	Establish contract with DOI to procure BIOME AI software
Community-focused engagement on AI needs	Q4, FY26	On-track	NCI	Hold workshop/symposium to determine potential AI-focused projects
Deliver next batches of MCI samples for Azenta sequencings	Q4, FY26	On-track	NCI	Working to align research characterization with clinical sequencing for MCI
Determine target values for items 5 through 8			CDC	Mentioned as a milestone because these are important to determine to evaluate progress.
Data extraction process determined (Target Goal 5)			CDC	Determining how and by whom this data is extracted.
Data integration process for CCDI Ecosystem (Target Goal 5)			CDC	Determining how and by whom this data is integrated
Goal Target 7 – Need to assess percent of pediatric cancer cases coming from APHL AIMS Dashboard			CDC	After access has been granted (in process) we can start to determine if this is a viable method to determine pediatric cancers.

Data Accuracy & Reliability

Please include a brief description of how the team will ensure the accuracy and reliability of the data and will compensate for any data limitations.

Data Source	Accuracy	Reliability	Data Limitations?	Comments
Molecular Characterization Initiative	High accuracy	High reliability	Data are limited to what is agreed-upon in our MCI protocol (contracts with IGM and Azenta)	CCDI data pipeline validates and aligns incoming data to standards and metadata requirements, assuring data quality
NCCR	High accuracy	High reliability	Agreements with individual state registries and SRP contracts for cancer surveillance reporting	Public health reporting for state registries is highly regulated and standardized
Federated data system partners	High accuracy queries	High reliability queries; variable access requirements	Search limited by API; access limited by rules of individual ecosystems	Federation allows users to find datasets across systems but access is not currently seamless
APHL AIMS	High accuracy	High reliability	Not all pathology labs report through APHL AIMS Platform	Every state cancer registry acquires some path lab data from APHL AIMS platform.

Additional Information

Related Policy and Partners

- President's Management Agenda:
  - Leverage Technology to Deliver Faster, More Secure Services

- Policies:
  - Executive Order – Unlocking Cures for Pediatric Cancer with Artificial Intelligence
  - HHS Strategic Plan - *forthcoming*

Goal Team

Questions regarding this APG can be directed to performance@hhs.gov

National Institutes of Health

Goal Lead:

- Warren Kibbe, Director for Data Science and Strategy, The National Cancer Institute

Implementation Team:

- Jaime Guidry Auvil
- Eve Shalley
- Subhashini Jagu
- Betsy Hsu

Centers for Disease Control and Prevention

Goal Lead:

- Joseph D. Rogers, Lead Health Scientist (Informatics, Division of Cancer Prevention and Control, Centers for Disease Control and Prevention

Implementation Team:

- Sean Porter
- Pradeep Podila
- Paran Pordell
- Brandy Peaker