

Tribal Consultation Report and Response

National COVID Cohort Collaborative (N3C)

National Center for Advancing Translational Sciences

National Institutes of Health

July 29, 2022

Table of Contents

Executive Summary.....	1
Description of the Critical Event	2
N3C Program Background.....	2
N3C Architecture.....	3
N3C Data Policy.....	3
Data Protections	4
Overview of the N3C Data Set	5
AI/AN Data	5
Steps Taken to Initiate a Consultation	6
Perspectives and Recommendations Provided During the Consultation Process and NCATS Response.....	6
Next Steps	9
Closing.....	10
Addendum	10

Executive Summary

In fall 2020, the National Center for Advancing Translational Sciences (NCATS) formally launched the [National COVID Cohort Collaborative \(N3C\)](#) in response to the urgent need for access to clinical information to help researchers study the COVID-19 pandemic. As part of NCATS' efforts to understand whether and how to make American Indian and Alaska Native (AI/AN) data held within the N3C accessible to researchers, NIH initiated formal [Tribal Consultation](#) to solicit input and feedback from Tribal Nations. Specific feedback received included the following:

1. Dissemination of significant findings is an important consideration so that if AI/AN researchers access the data and identify significant findings within the data, other communities have access to that information.
2. The N3C and access to AI/AN data provide a mechanism for improving understanding of how COVID-19 is impacting regional and local communities, as well as how to prepare for future pandemics.
3. The Code of Conduct could address specific criteria and expectations for accessing AI/AN data.
4. Although there are THRO representatives on the NCATS N3C Data Access Committee (DAC) THRO does not represent Tribal Nations.
5. It was noted that Tribal Nation representation in the data will be largely absent from the N3C without an agreement/collaboration with the Indian Health Service (IHS). AI/AN data will likely be represented by urban dwelling populations. Limitations of the data and potential to harmonize with IHS data needs to be considered.
6. Some level of data related to Native American COVID-19 data should be unobscured to more accurately examine why Native Americans were hit particularly hard by COVID-19 and how to mitigate these issues in the future. Going forward, the N3C should create a stand-alone category for Native Americans rather than including Native Americans in the "other" category. However, data regarding ZIP codes of COVID-19 diagnosis should continue to be obscured.
7. Currently no Tribal enrollment or affiliation data are collected regarding COVID-19. This should remain intact. Under no circumstances should Tribal enrollment information be collected.
8. The N3C must apply safeguards to ensure the proper balance is struck so that enough data are available to make adequate research possible, but not enough to violate general privacy rights of AI/AN people.
9. The N3C should continue to hold ongoing consultations with Tribal Nations to foster open communication. Furthermore, the N3C should work and collaborate with IHS, IHS facilities, and health care facilities operated by Tribes or Tribal-serving organizations to provide assistance and knowledge to AI/AN researchers.

This document outlines and describes NCATS' response to the feedback received through Tribal Consultation. NCATS intends to remove the obscuring practices currently in place on AI/AN data in the

N3C so that the benefits of the data for individuals identified as AI/AN can be realized. NCATS will continue to engage Tribal Nations on the use of AI/AN data in the N3C.

Description of the Critical Event

Established in response to the COVID-19 pandemic, the N3C formed in May 2020 and NCATS opened the N3C Data Enclave¹ to researchers in September 2020. The N3C Data Enclave contains real-world data from patients who have had a COVID-19 test or were suspected of having COVID-19, whether the test was positive or negative. These data come from partner academic medical centers and their affiliated clinical partners from across the United States. Within these data are patient records where race and ethnicity data are self-identified or assigned by the health care organization as AI or AN in the electronic health record (EHR).

N3C Program Background

Since the beginning of the pandemic, health care providers and researchers have worked urgently to understand the novel SARS-CoV-2 virus and the disease it causes, COVID-19. Our early understanding of COVID-19—its signs, symptoms, and effective courses of treatment—was very limited. While this disease rapidly spread through communities, cities, and countries and crossed borders and oceans, important patient health information was being collected in many clinics and hospitals. Data collected in this manner were difficult to put together for researchers and physicians to understand the characteristics of this new disease and to be able to address critical questions, such as how it spread, who was most at risk, which treatments helped, and what the effects of the disease were, including long-term effects.

In response to this urgent need, NCATS and its partners developed the N3C to collect existing EHR data from hospitals and clinics and to harmonize these data and make them accessible to researchers seeking to understand COVID-19. The N3C effort is centered on the following:

1. Establishing a secure data repository for studying COVID-19-related data.
2. Receiving existing patient data derived from EHRs provided by participating U.S. health care sites.
3. Harmonizing EHR data from across different health systems so that they can be combined for research.
4. Providing support for researchers using the N3C Data Enclave to collaborate on COVID-19 research.
5. Ensuring research using the N3C Data Enclave follows the applicable rules, policies, and expectations of NIH and NCATS and its partners to keep the data secure and protect patient privacy.

¹ The N3C Data Enclave is a secure platform through which the harmonized clinical data provided by our contributing sites is stored. The data itself can only be accessed through a secure cloud portal hosted by NCATS and cannot be downloaded or removed.

The N3C is a collaborative, community-driven initiative that allows research teams nationwide to study clinically important questions about COVID-19, including risk and protective factors in particular populations, medications that may mitigate or promote severe infection, and long-term effects of infection. The N3C Data Enclave is the largest collection of patient EHRs and associated clinical information available for COVID-19 research. It is a partnership among the NCATS-supported [Clinical and Translational Science Awards \(CTSA\) Program](#) institutions, the [National Center for Data to Health](#), and [National Institute of General Medical Sciences–funded Institutional Development Award Networks for Clinical and Translational Research \(IDeA-CTR\)](#) institutions, with overall stewardship by NCATS.

N3C Architecture

The N3C Data Enclave maintains real-world clinical data from patients who have had a COVID-19 test or were suspected of having COVID-19, whether the test was positive or negative. The data, which are refreshed weekly, include demographics, symptoms, laboratory test results, procedures, medications, medical conditions, physical measurements, and other information that captures two years of clinical history data prior to COVID-19 testing.

Partner academic medical centers and their affiliated clinical partners from across the United States transfer data to NCATS, which harmonizes the information and oversees the centralized, secure data enclave. Researchers may apply to access the data only for COVID-19 research purposes, and their [project descriptions](#) are publicly posted.

The N3C encourages researchers to participate in the broader N3C community by joining Domain Teams to engage with others who have similar COVID-19 research interests. The Domain Teams address diverse topics, such as machine learning, rural health, nursing, and cardiovascular disease. Training to understand how to navigate and work within the N3C is also an important effort. There are “office hours” for N3C users, as well as training videos and tutorials to assist individuals working in the N3C data platform.

N3C Data Policy

EHRs are digital, machine-readable versions of patients’ paper medical charts. They contain clinical information, such as medical history, diagnoses, demographics, immunization records, laboratory results, medications, and more. EHRs also contain data that may identify a person, also known as protected health information. Patients are not asked to provide data to the N3C. Rather, contributing sites de-identify existing data from individuals who are tested for COVID-19.

Under the [1996 Health Insurance Portability and Accountability Act \(HIPAA\)](#), covered entities—such as health care providers—may release data for research without obtaining an individual’s authorization if direct identifying information is removed and appropriate oversight and agreements are in place. Accordingly, NCATS requested that participating health care institutions transfer what is known as a “Limited Data Set”² to the N3C so that these needed clinical data could be used for COVID-19 research. A [Limited Data Set](#) (LDS) is protected health information that excludes 16 types of direct identifying information for the purposes of research, public health or health care operations without obtaining an individual’s authorization. A data use agreement must describe the terms and conditions of the data use. NCATS requested that the LDSs include the five-digit ZIP codes (HIPAA allows ZIP codes to be

² [HIPAA Privacy Rule and Its Impacts on Research \(nih.gov\)](#)

included in the LDS, but not the complete residential address) and dates of service. These data would be critical for tracking the progress of the pandemic over time and providing some specificity to regional outbreaks. The NIH Institutional Review Board (IRB) has determined there is minimal risk to individuals, and NCATS received a waiver of informed consent from the NIH IRB, conforming to the Federal Policy for the Protection of Human Subjects (or “Common Rule”).

As the steward of the data, NCATS is taking multiple precautions to protect the confidentiality, security, and integrity of the data. For example, NCATS oversees a robust governance and oversight process that includes [Data Transfer Agreements](#) (DTAs) and [Data Use Agreements](#) (DUAs) with institutions, [Data Use Requests](#) (DURs) with users, and a DAC that grants N3C Data Enclave access to qualified researchers. Both the [DUA](#) and [N3C Data User Code of Conduct](#) prohibit the re-identification of individuals, individual providers, and sites of care. Furthermore, all work must be done within the enclave, and no data from the EHRs may be downloaded. Users are allowed to download research results after they are reviewed by a data download committee.

Data Protections

NCATS is carefully balancing the need for researchers’ access to patient data to address the pandemic with the need to protect patient information and privacy. At the same time, NCATS and the N3C encourage research collaborations and rapid dissemination of research results. NCATS closely works with its partners to discuss and develop procedures that align with **regulations and policy**, to develop **privacy measures** to keep the data protected, to test and monitor the N3C Data Enclave and data cloud **security**, and to review **researcher responsibility** expectations to ensure that users do their part to keep the data secure and protect the privacy of the patient health information. The following table shows the efforts prior to Tribal Consultation that NCATS has undertaken to ensure the security, confidentiality, and privacy of N3C data.

Regulatory and Policy	Privacy Measures	Security Testing and Monitoring	Researcher Responsibilities
<ul style="list-style-type: none"> • Data-contributing sites abide by the HIPAA Privacy Rule • N3C research is subject to the Federal Policy for the Protection of Human Subjects in research ("Common Rule") • Data are provided as a HIPAA-defined limited data set • NIH IRB oversight and waiver of consent • For COVID-19–related research only • No genomic data • No emergency public health authorities were used to obtain the data under these conditions 	<ul style="list-style-type: none"> • Certificate of Confidentiality • Data stay within the N3C Data Enclave: No download or capture of raw data • Privacy Impact Assessment • Review of project requests by the Data Access Committee • Additional Tribal data privacy measures (while seeking a consultation with Tribal Nations) 	<ul style="list-style-type: none"> • Federal government–compliant enclave managed by NCATS • Meets government security controls for cloud security and privacy • Data encryption in transit and at rest, without exception • Scheduled penetration testing • Active monitoring and logging by NIH and HHS • Auditing of activities in the N3C Data Enclave 	<ul style="list-style-type: none"> • A user's organization signs a Data Use Agreement with NCATS for terms and conditions of use • Users adhere to the N3C Data User Code of Conduct • Required NIH IT security training • Required Human Subjects Research Protection training • Follow N3C's Community Guiding Principles

Overview of the N3C Data Set

The [N3C Cohort Exploration dashboard](#) provides an overview of key metrics and distributions by age and sex, race and ethnicity, and comorbidity within the N3C Data Enclave. As of June 27, 2022, the N3C Data Enclave included patient records from 14.3 million individuals, including about 5.6 million patients with COVID-19. These data were contributed by 93 organizations, with 306 DUAs executed.

As of June 2022, the N3C has made data accessible to more than 3,000 researchers and clinicians to study the progression of COVID-19, identify risk and protective factors, search for effective treatments, understand the long-term disease effects, and determine how best to care for those with the disease. The current plan is for the N3C to operate for five years, beginning in 2020, and to continue to update the data during this time period. During the fifth year, NCATS will evaluate the status of the enclave and researcher needs to determine whether to extend the agreements another five years or stop the program.

AI/AN Data

Since the establishment of the N3C, NCATS has engaged in extensive discussions with the NIH Tribal Health Research Office (THRO) and met with the NIH Tribal Advisory Committee to identify potential issues of concern to Tribal Nations. NCATS developed a temporary strategy until formal Tribal Consultation could be held, where ZIP codes were not accessible for research when they represent regions where the majority of residents are AI or AN. In addition, EHRs where race and ethnicity data are

self-identified as AI/AN had their race information aggregated into an “Other” category to temporarily obscure the data until Tribal Consultation could occur. It is important to note that the N3C *does not* have EHR data from IHS, nor does it contain information about Tribal affiliation.

A summary review by NCATS to characterize obscured data indicates that as of March 24, 2022, more than 50,000 patient records self-identify as AI or AN, representing all gender, age, and COVID-19 status demographics. There has been no review of obscured ZIP codes or scientific analysis of unobscured AI/AN data.

Steps Taken to Initiate a Consultation

Throughout the development and implementation of the N3C, NCATS has maintained an ongoing engagement with THRO and met with several external partners, such as the Johns Hopkins Center for American Indian Health (August 2020), the Southcentral Foundation (September 2020), and Tribal Epidemiology Center Directors (September and November 2020), as well as with other NIH staff members who have recently completed a Tribal Consultation (e.g., RADx), to prepare for a Tribal Consultation.

NIH convened a one-hour virtual NIH Tribal Consultation and Urban Confer on the National COVID Cohort Collaborative on February 11, 2022, from 1:30 to 2:30 p.m. EST. A [Dear Tribal Leader Letter](#) announcing the Tribal Consultation was sent to Tribal leaders, and a recorded [informational webinar](#) about the N3C Program was made available prior to this engagement on the [NCATS Tribal Consultation webpage](#).

Topics suggested for consideration at the Tribal Consultation included the following:

- Should AI/AN data be unobscured and made available for COVID-19 research?
 - How would Tribal communities want to be involved in governance, access, use, etc.?
- What steps would the N3C need to take to consider making the data available?
 - Which aspects of governance would be particularly critical for this resource?
 - How can ongoing partnerships with Tribal communities and researchers be instituted and maintained?
- If Tribal Nations identify benefits to using the N3C resource, what outreach is needed to better engage AI/AN researchers?

After the Tribal Consultation, NCATS engaged with the NIH Tribal Advisory Committee at its monthly meeting to gain additional feedback.

Perspectives and Recommendations Provided During the Consultation Process and NCATS Response

Feedback item 1: Dissemination of significant findings is an important consideration so that if AI/AN researchers access the data and identify significant findings within the data, other communities have access to that information.

NCATS Response

NCATS is committed to dissemination. Through the Code of Conduct, to which researchers must attest, and DUAs, users and their institutions are encouraged to publicly disseminate findings or new insights into COVID-19 or related disease etiology, progression, or pathology based on research conducted using the N3C Data Enclave, ideally through open-access formats. The main way to disclose findings is through the N3C Dashboard, which is open to the public. The N3C Dashboard provides an overview of the data that can be found within the resource, along with publications and presentations resulting from researchers using the N3C. NIH has additional platforms for disseminating findings and publications, as well.

Feedback item 2: The N3C and access to AI/AN data provide a mechanism for improving understanding of how COVID-19 is impacting regional and local communities, as well as how to prepare for future pandemics.

NCATS Response

NCATS recognizes that the AI/AN data that are not currently accessible could be useful for understanding the impact of COVID-19, identifying effective interventions to improve public health, and generating knowledge directly useful to Tribal communities and public health responses to COVID-19. Research conducted by accessing the N3C is already beginning to contribute to our understanding of COVID-19 and is revealing new knowledge, such as populations at greater risk for severe COVID-19 outcomes, COVID-19 severity when considering social determinants of health, and identification of effective drugs repurposed as a treatment for COVID-19. The understanding of how these clinical data can be used as real-world evidence is growing, allowing examination of pandemics at regional and local levels.

We would like the research utilizing the data in the N3C to inform the health of Tribal Nations, as well. For these reasons, the N3C has decided to move AI/AN data into a stand-alone category so that these benefits can be realized.

Feedback item 3: The Code of Conduct could address specific criteria and expectations for accessing AI/AN data.

NCATS Response

Changes to the Code of Conduct will be incorporated to emphasize AI/AN data access expectations.

Feedback item 4: Although there are THRO representatives on the NCATS N3C Data Access Committee (DAC), THRO does not represent Tribal Nations.

NCATS Response

The NCATS N3C DAC must be composed of federal officials because HHS/NIH/NCATS is assuming fiduciary responsibility for data security and access control. To ensure Tribal representation, NCATS proposed options for consideration during the Tribal Consultation process, including the formation of a Tribal DAC that reviews all DURs involving AI/AN-specific research or the

submission of a DUR that would be specific for AI/AN-focused research. Should it be determined that one of these approaches is needed, this feedback on representation will be incorporated.

Feedback item 5: It was noted that Tribal Nation representation will be largely absent from the N3C without an agreement/collaboration with IHS. AI/AN data likely will be represented by urban-dwelling populations. The limitations of the data and potential to harmonize with IHS data need to be considered.

NCATS Response

The lack of harmonization of EHR data is a long-standing and significant issue. NCATS welcomes a collaboration with IHS, but a breadth of data is available in the N3C that could be generalizable for some communities.

An internal analysis by NCATS to characterize the obscured data indicates that as of March 24, 2022, more than 50,000 patient records are indicated as AI or AN according to self-identified or clinic-assigned demographic data from EHRs and are inclusive of all gender, age, and COVID-19 status demographics. There has been no review of obscured ZIP codes or scientific analysis of unobscured AI/AN data. Because Tribal Nation affiliation is not collected, it is difficult to fully understand the population of data coming from Tribal Nations versus urban-dwelling populations, without removing the obscured ZIP code protections.

Given the limited identifying information in the AI/AN data and the potential for broader understanding of risk factors and treatment of COVID-19, NCATS views the potential risks of unobscured AI/AN data as low.

Feedback item 6: Some level of data related to Native American COVID-19 data should be unobscured in order to more accurately examine why Native Americans were hit particularly hard by COVID-19 and how to mitigate these issues in the future. Going forward, the N3C should create a stand-alone category for Native Americans rather than including Native Americans in the “other” category. However, data regarding ZIP codes of COVID-19 diagnosis should continue to be obscured.

NCATS Response

The N3C will move AI/AN data back to a stand-alone category. ZIP codes contained within AI/AN data will be truncated to the first three digits or omitted entirely if the region represented by those first three digits contains fewer than 20,000 people.

Feedback Item 7: Currently no Tribal enrollment or affiliation data are collected regarding COVID-19. This should remain intact. Under no circumstances should Tribal enrollment information be collected.

NCATS Response

Tribal enrollment and affiliation data will not be collected within the N3C.

Feedback Item 8: The N3C must apply safeguards to ensure the proper balance is struck so that enough data are available to make adequate research possible, but not enough to violate general privacy rights of AI/AN people.

NCATS Response

NCATS is carefully balancing the need for researchers' access to patient data to address the pandemic with the need to protect patient information and privacy. At the same time, NCATS and the N3C encourage research collaborations and rapid dissemination of research results. NCATS works closely with its partners to discuss and develop regulatory and policy issues that affect N3C operations; to develop privacy measures to keep the data protected; to test and monitor the N3C Data Enclave and data cloud security; and to review researcher responsibility policies to ensure that users do their part to keep the data secure and to protect the privacy of the patient health information. The safeguards implemented by the N3C are described in the four pillars shown above and described [online](#). Per the next steps, we are taking additional measures to make the AI/AN data available, with safeguards around the data availability.

Feedback Item 9: The N3C should continue to hold ongoing consultations with Tribal communities to foster open communication. Furthermore, the N3C should work and collaborate with IHS, IHS facilities, and Tribally run health care facilities to provide assistance and knowledge to AI/AN researchers.

NCATS Response

NCATS, which leads the N3C Program, looks forward to engagement with Tribal communities as issues for discussion arise involving AI/AN communities and research. NCATS will work with THRO to maintain communication. NCATS agrees that additional collaboration with IHS could be beneficial.

Next Steps

Based on feedback from Tribal Consultation, NCATS will take the following steps to make AI/AN data available for research through its standard Data Use Request process:

1. AI/AN data will be moved back to a stand-alone category. With this change, AI/AN data will be available in any N3C analysis that provides race and ethnicity distribution.
2. ZIP codes that overlap with Tribal communities will be available for research in the following manner:
 - a. ZIP codes for all geographic units containing 20,000 or fewer people will be removed entirely. This is standard practice for all geographic units and will be applied the same way when AI/AN data are restored to a separate category.
 - b. Currently, specific ZIP codes representing rural populations predominantly with AI/AN-identifying individuals are hidden. These will now be visible in both the LDS and de-identified data. This means that AI/AN data will be managed as others are managed with the exception that the full five-digit ZIP codes will never be shown.
3. The N3C Data User Code of Conduct will be modified so that data users will be asked to attest that they understand the N3C contains no Tribal affiliation data and that use of AI/AN data and ZIP code information to make assumptions about Tribal affiliation is not valid or appropriate. This statement will be included as a reminder when a DUR is received, a DUA is executed, when data are accessed in the N3C platform, and during the publication processes.
4. The N3C will continue to engage THRO and Tribal Nations as issues for discussion arise.

Closing

NCATS appreciates the opportunity to participate in the Tribal Consultation process and has found the feedback received from it to be extremely informative. The Center is grateful to Tribal Nations and leaders for their insights and looks forward to continuing these dialogues in the future.

Addendum

In response to a Tribal Nation request for an extension on the deadline to provide written testimony, NCATS delayed release of this report past the intended 90-day timeline indicated in the [NIH Tribal Consultation Policy](#). NCATS felt it was critical that all feedback was heard and wanted to accommodate Tribal leaders.