N3C Attribution and Publication Principles

Version history

Version	Description of changes
All versions	doi.org/10.5281/zenodo.7126728 This DOI represents all versions, and will always resolve to the latest one.
1.0-2020-08-07	● First draft
1.1-2020-08-25	● First official release ● Link to N3C marker paper updated from preprint to in press DOI ● Manuscript submission form language and link updated
1.2-2020-08-25b	● Citation of marker paper requested ● N3C enclave link added ● Change log added to header
1.3-2020-10-15	● Addition of consortial manuscript criteria ● Peer review processes
1.3-2020-11-05	● Addition of authorship criteria
1.4-2021-05-06	● Refine definition of consortium vs. project-based manuscripts and authorship attribution
1.5-2022-09-09	● Terminology changed from "consortial manuscript" to "manuscript about N3C" ● Clarifying the role of the committee in addressing controversial statements ● Adding acknowledgment of Code-set creators and reviewers

1.6-2023-01-11	● Clarifying language about use of data from AI/AN individuals
	● Abstracts, posters and podium presentations:
	    • Updated condensed acknowledgements
	    • Updated submission requirements
	● Added encouragement to work with domain teams (covid.cd2h.org/teams)
	● Added requirement for inclusion of the approved MSID in the body of the work
	● Data and Code Availability Statement
	● Incorporation by reference of acknowledgements web page
	● Attribution of Codesets/ Concept sets/ Value sets
	● Added encouragement to use the Author Arranger tool from NIH to help with large author lists

1.7-2023-03-30	● Added reference to the NIH policy about deposition in PubMedCentral ● Added clarification about scope to include works about computational models, but exclude derivative works and works by authors who are not registered with N3C. ● Added attribution for Axle Informatics Subcontract Number: NCATS-P00438-B

The N3C is more than simply a data enclave; it is also a collaborative research community committed to the rapid generation and dissemination of scientific knowledge for the public good, and to the advancement of COVID-19 science.
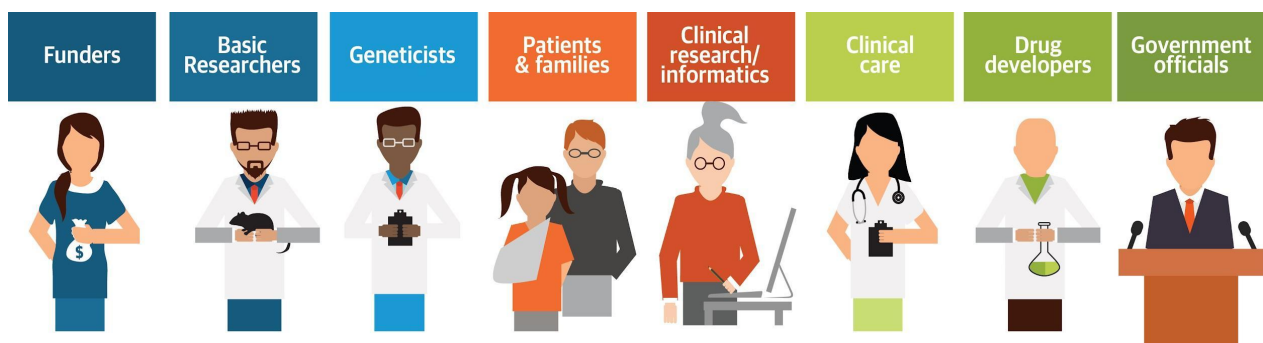
Purpose: This document provides community-driven guidelines and approaches that all users within the N3C research community uphold, and it addresses attribution and publication principles regarding N3C Community dissemination of research. **These publication and attribution principles apply to N3C analysis reports, data, resources, abstracts, presentations,**

preprints, manuscripts, and other publications arising from using content in the N3C Data Enclave. This document will be reviewed and updated regularly as N3C evolves, but at least annually by the N3C Governance Workstream and modified or recertified as deemed necessary.

Overview: These guidelines are intended to enable the following:

- A transparent and collaborative environment within the N3C where all contributions are acknowledged;
- Promote provenance and reproducibility of COVID-19 research;
- Sharing research results with N3C users as soon as they are mature enough;
- Clarify attribution expectations for all N3C artifacts, specifically publications.

Community Attribution Guidelines: The N3C community is dedicated to the goal of collaboration and rapid knowledge generation and dissemination to help combat COVID-19.

To simultaneously support career development and demonstrate contribution and acknowledgment from all types of contributors throughout the research process, the N3C takes a proactive approach to attribute and track all contributions to any result or artifact derived from the use of N3C in alignment with the N3C Community Guiding Principles.

- The N3C Enclave supports robust tracking of provenance and attribution.
  - Masthead authors
  - Block authors (Core contributors)
  - Acknowledgment by attribution
- Every contributing organization (e.g., data contributors, software developers, etc.) will be listed on the N3C website along with their Research Organization Registry identifier where available.
- Each user must have an ORCiD ID to use through N3C registration, supporting attribution within the N3C Enclave, and in subsequent artifact dissemination.
- Users are encouraged to list all contributors to any Data or Resource that is ingested, as well as on any report or outcome, whether they are automatically tracked within the analysis workflow or not.
- All workflows and their resulting reports will indicate any upstream contributors to data and code automatically. Additional upstream or downstream contributors not automatically tracked in the workflow may be added.
- N3C will support authors in appropriately acknowledging contributors.

The N3C Data Enclave uses the Contributor Attribution Model (CAM) established within the NCATS-funded Center for Data to Health (CD2H) program to automatically capture the aforementioned granular contributions as part of provenance tracking. The resulting set of contributors will be included as part of N3C "reports"—finalized reproducible analysis documents

that link to specific versions of all code, data, resources, and contributors. The CAM provides a simple data model for representing information about contributions made to research-related artifacts—for example, a curator contributing to a mapping file.

Working with a domain team

All N3C researchers are strongly encouraged to work with a domain team (see covid.cd2h.org/teams) when doing analyses and preparing manuscripts; doing so avoids both methodological weakness and policy violations. However, even manuscripts that are not prepared with the involvement of a domain team may nevertheless benefit from reuse of code and codeset artifacts produced by the domain team, and in this case, consortial authorship for the contributors to those artifacts is appropriate.

Scholarly works in scope for review

- **Individual DUR project-based works**. Scholarly works put forth by members of the N3C community describing the research, software development, etc., but which are not about the N3C infrastructure or methodology and do not represent the N3C community as a whole. Project-based manuscripts are largely limited to discoveries made with the available data infrastructure as a result of one or more Data Use Requests (DURs) submitted and approved for access to the N3C Data Enclave. Examples:
  - Description of specific findings in the data
  - Description or validation a novel computational model that uses the data

- **Works about N3C.** These manuscripts describe the N3C infrastructure, methods, or community and represent the program as a whole. Such manuscripts and associated works are central to the N3C initiative. Examples of consortium manuscripts are:
  - Commentaries on the vision of the N3C program
  - Methodological papers describing the N3C data, quality assurance processes, or other endeavors requiring broad community contribution and endorsement
  - Focus on N3C processes or community development
  - Focus on core infrastructure, data harmonization or quality assurance, or its governance and administration

- **All formats.** In addition to manuscripts, all abstracts, podium presentations, and posters must be submitted to the Publication Committee prior to submission to the conference. This is to ensure compliance with download policy as well as to track outputs. Authors are actively encouraged to consider inclusive authorship in their abstracts and podium presentations. Consortial authors will not be collected for posters and presentations, but will be collected for all full length manuscripts, whether conference proceedings or traditional peer-reviewed venues. The Publication Committee will review abstracts within 2 business days.

- **Out of scope for review:**
  - Works that merely reference public information about N3C but that do not use the secure data enclave or infrastructure at any point
  - Works whose authors include no registered N3C researchers
  - Works derived directly from content approved by the Publication Committee (For

example an approved manuscript that was reformatted as a presentation). To be exempt from review, all such derivatives must be devoid of new or updated downloads, and conclusions must not be different.

Authorship Designations & Acknowledgment

N3C fosters team science and insists that authors and collaborators be identified in publications to ensure appropriate credit and recognition.

Authorship criteria

One of the important determinations to make for any contributor is whether their contribution constitutes masthead authorship. We recommend discussion between the lead author and the contributor(s) to the research described in the manuscript to determine the masthead authorship. Contribution to the N3C Data Enclave does not provide automatic authorship on a data user's manuscript. Co-authorships should reflect a common-sense assessment but in general, should reflect substantial work or intellectual contributions to the topic of the publication. The N3C community recognizes that the science we undertake within the N3C is necessarily dependent on the work of many people. N3C supports adherence to the ICMJE standards. A contributor must declare a significant or critical contribution to be included as an author. Authors are expected to do at least one of the following: edit or write a manuscript; provide data, figures, or tables; provide critical ideas; and/or write code. For example, a codeset is used for critical success of a manuscript. There should be communication with codeset authors/reviewers as to who the critical contributors were, and consideration should be given as to the inclusion of those contributors as authors on the manuscript. For manuscripts intended for journal publication, include a statement that authorship was determined using ICMJE recommendations.

	Masthead Authorship	Consortial Collaborators/ Consortial Authors	In Text Acknowledgements
Indexed in Medline	Yes	Yes	No
Nature of contribution	Substantial intellectually important contributions made directly to the manuscript	Less substantial but still intellectually important contributions made directly to the study/manuscript or to knowledge artifacts used directly by the study** (e.g. phenotype definition, knowledge store products, variable codeset)	Important contributions directly to the N3C infrastructure. Note that in text acknowledgement may occasionally overlap with masthead and / or consortial authors.
Who is invited	Exclusively at the discretion of contact & senior authors	by default: Core Authorship blocks as relevant: codeset and other artifacts creators/reviewers and selected domain teams***	Core Authorship blocks
Publication Committee responsibilities	*	Pub Committee notifies core contributors, Codeset or other core artifacts creators/reviewers & relevant domain teams of manuscripts. Pub Committee alerts contact author of any responses to consortial-level contributions	Pub Committee maintains the core contributors table and the default pastable acknowledgements text. Pub Committee notifies core contributors of manuscripts. Pub Committee alerts contact author of any opt-outs.
Contributor responsibilities	*	Contributors respond within 3 business days via contributor's form.	Contributors do not need to take action to be acknowledged; they may optionally opt-out within 3 business days via contributor's form.
Contact Author responsibilities	*	Contact author is provided with a spreadsheet of those that have responded to the contributions form. The spreadsheet will include Names, Emails, ORCIDs, and Contributions. Please upload this file with your submission. Do NOT publish email addresses.	If anyone opts out, contact author will be provided with the names of individuals to omit from acknowledgement, else use the default acknowledgements text. Contact author will paste the acknowledgements text in the appropriate section in the body of their manuscript.

At the contact author's discretion the contributor's form CAN be used to collect info from masthead authors; however the policy does not require this. For masthead authors, the form is not usually necessary unless there are dozens of masthead authors or when the journal requires that each author specifies their contributions. Authors are encouraged to use the NIH https://authorarranger.nci.nih.gov/#/web-tool author arranger tool to manage large author lists.

*** In cases where the N3C is the subject of the manuscript rather than just the platform on which the study was done, consortial authors list will be more comprehensive.*

**** To minimize overlaps and maximize efficiency, all investigators are strongly encouraged to collaborate with the relevant domain team(s) from the start of the project. covid.cd2h.org/teams*

Disclaimer and Acknowledgements text

Disclaimer and attribution to include in all works including posters and podium presentations

The N3C Publication committee confirmed that this *[manuscript/abstract/poster/presentation]* <MSID:XXXX.XX> is in accordance with N3C data use and attribution policies; however, this content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the N3C program.

The analyses herein were conducted using the NCATS N3C Data Enclave supported by NCATS U24 TR002306, Axel Informatics Subcontract Number: NCATS-P00438-B and made possible because of the patients whose data was contributed by partner organizations (covid.cd2h.org/dtas). We gratefully acknowledge the scientists who have contributed to the on-going development of this community resource (covid.cd2h.org/acknowledgements) [1]

1. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, Payne PRO, Pfaff ER, Robinson PN, Saltz JH, Spratt H, Suver C, Wilbanks J, Wilcox AB, Williams AE, Wu C, Blacketer C, Bradford RL, Cimino JJ, Clark M, Colmenares EW, Francis PA, Gabriel D, Graves A, Hemadri R, Hong SS, Hripscak G, Jiao D, Klann JG, Kostka K, Lee AM, Lehmann HP, Lingrey L, Miller RT, Morris M, Murphy SN, Natarajan K, Palchuk MB, Sheikh U, Solbrig H, Visweswaran S, Walden A, Walters KM, Weber GM, Zhang XT, Zhu RL, Amor B, Girvin AT, Manna A, Qureshi N, Kurilla MG, Michael SG, Portilla LM, Rutter JL, Austin CP, Gersing KR, N3C Consortium. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. J Am Med Inform Assoc. 2021 Mar 1;28(3):427–443. http://dx.doi.org/10.1093/jamia/ocaa196 PMCID: PMC7454687

Additional acknowledgement to include in published manuscripts:

The specific acknowledgements for manuscripts vary over time. Please see covid.cd2h.org/acknowledgements for the latest and copy each of the sections into the manuscript.

Declaration, Attribution of Codesets/Concept sets/Value sets

Codesets/Concept sets/Value sets and harmonized variables will be placed into a shared folder in the N3C Data Enclave and indexed in the Data Resource Library (in development), which is an expanded data dictionary. Codesets/Concept sets can be requested to be exported to GitHub (see Results Download Policy for details). Provenance and attribution metadata will be incorporated into the library. Codeset and variable artifacts will be licensed for reuse along a class of Creative Commons licenses, designated by the contributor. The CC BY 4.0 license is the preferred license class, and is the default license if not otherwise indicated. Please consult with your Institution's local Technology Transfer office for guidance and feedback on the use of the CC BY 4.0 license.

Authors are strongly encouraged to include the following statement about data and code availability.

The N3C Enclave is available for public research use. To access data including that used in this manuscript, institutions must have a signed Data Use Agreement executed with the U.S. National Center for Advancing Translational Sciences (NCATS) and their investigators must complete mandatory training

and must submit a Data Use Request (DUR) to N3C. To request N3C data access, researchers must follow instructions at https://covid.cd2h.org/onboarding.

This project utilized <specify Level 1, Level 2, or Level 3> data from N3C release <provide release number>, which can be replicated within the N3C Enclave by fully onboarded, trained users with an approved DUR. All concepts and definitions are available at <provide link or supplementary data number>. More than 4,000 researchers currently have access to data in N3C; together they represent more than 300 US research institutions. Code used in this analysis and all concept sets are available at <enclave URL or, if appropriate and approved, the Zenodo-provisioned DOI corresponding to the GitHub URL>. Reviewers can confidentially request access by these same means.

N3C Publication Committee

Membership consists of volunteer members of the N3C user community. Anyone is welcome to inquire; members are selected for inclusive and varied participation in N3C, diversity of expertise and career level, geographic and institutional diversity, and racial and gender balance.

Role: The Publication Committee's role is to set community standards and expectations, keep track of productivity and outputs of the N3C member community, and facilitate inclusive and appropriate attribution, as well as promote collaborations and troubleshoot authorship questions with regard to publication of specific projects or efforts within the N3C community.

Review Scope: All work arising from the use of content in N3C Data Enclave will be submitted to the N3C Publication Committee. Instructions are posted on the N3C Publication Review webpage.

- **All full manuscripts** intended for journal publication or preprint must be submitted to the Publication Committee prior to submission to a journal/preprint server. The Publication Committee does not provide a full scientific peer review of the manuscript though they may offer suggestions regarding enhancing reproducibility of the research. The Publication Committee will also, within 10 business days of submission:
    - Support the authors in confirming authorship and attributions.
        - Determine whether the paper is about N3C or is project-based
        - Invite core contributors to declare contributions if they meet consortial authorship requirements
        - Suggest domain teams that may have contributed knowledge artifacts (eg. code sets)
        - Collate responses and return them to corresponding authors
    - Verify alignment with the Data Use Request (DUR) including COVID-relatedness.
    - Verify that no individual or line level data are included per the Data Use Agreement (DUA).
    - Verify that the aggregate data included in the manuscript have been approved by the Data Download Committee per the Results Download Policy.
    - Verify that the reported data pertains to 20 or more persons. Exceptions can be granted by NCATS to reduce the reportable cell size threshold to 10 persons based on scientific interpretation. Documented Appeal decisions must be provided to the Publication Committee. Authors must include an asterisk or footnote in their manuscript for affected values and state "***The N3C Download Committee***

> ***reviewed and granted an appeal for this manuscript to report cell patient numbers as low as 10 (instead of the normal threshold of 20)***".

- ○ Verify respect for sovereignty of American Indian/Alaska Native (AI/AN) individuals. N3C contains no Tribal affiliation and the use of AI/AN data and ZIP code information to make assumptions about Tribal affiliation is not valid or appropriate.

- ○ Offer feedback to authors discussing particularly sensitive or controversial topics.

- ○ Encourage authors to reference a link to their analytics pipeline in the N3C Data Enclave as a meaningful demonstration of N3C's commitment to reproducibility.

- ○ Encourage authors to publish in open-access journals and make committee-approved works immediately available. Immediate availability can be accomplished via preprint submission to Medrxiv or other established preprint registry such as Zenodo. Published manuscripts must be submitted to PubMed Central and should be made publicly available immediately upon acceptance to publication without an embargo period.

If these processes uncover issues specific to the [N3C Data User Code of Conduct](#), or if the manuscript addresses controversial topics, the Publication Committee may contact the user and NCATS staff for follow-up action.

Confidentiality and endorsement

Draft manuscripts, abstracts, and podium presentations submitted to the N3C Publication Committee prior to publication will be treated as confidential by the N3C Publication Committee and N3C leadership. This includes, but is not limited to, information shared via the Publication Intent Form. Access to N3C Data Enclave resources does not imply endorsement of the research project and/or results by NIH or NCATS.

Press releases and announcements

Press releases or major announcements about the N3C, including project-based results, must be coordinated with and reviewed by the NCATS Communications Team prior to distribution.

Brief announcements are allowed but must include disclaimer language such as "**The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the N3C Program**."