N3C Attribution and Publication Principles

Version history

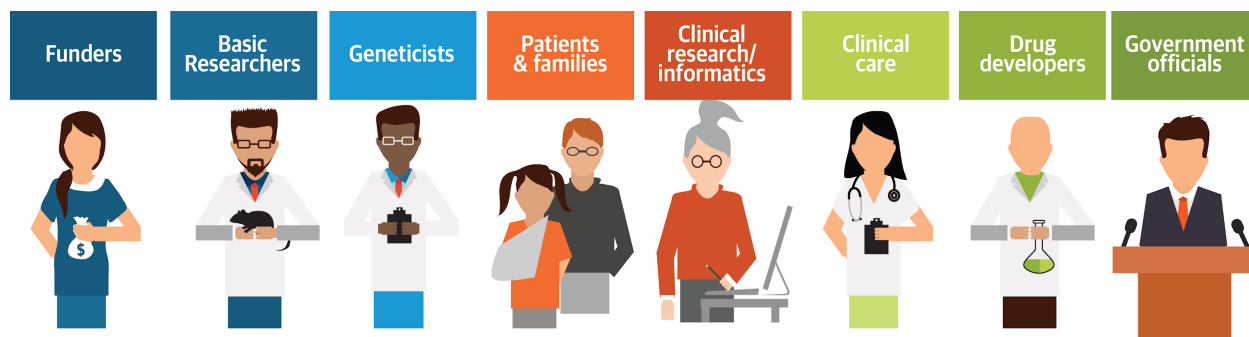| Version	Description of changes
All versions	https://doi.org/10.5281/zenodo.3992394 This DOI represents all versions, and will always resolve to the latest one.
1.0-2020-08-07	● First draft
1.1-2020-08-25	● First official release ● Link to N3C marker paper updated from preprint to in press DOI ● Manuscript 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

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 the use of content in 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 attributing and tracking 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
  - 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.
- All users share ORCiD IDs 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 Center for Data to Health (CD2H) program to capture the aforementioned granular contributions automatically 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.

Manuscript Designations

- **N3C Consortium manuscripts.** These are manuscripts describing research, software development, etc., that involve substantial contributions from the N3C community. Such manuscripts and associated works are central to the N3C initiative. Examples of consortium manuscripts are:
  - o Commentaries on the vision of the N3C program
  - o Methodological papers describing the N3C data, quality assurance processes, or other endeavors requiring broad community contribution and endorsement
  - o Focus on N3C processes or community development
  - o Focus on core infrastructure, data harmonization or quality assurance, or its governance and administration
  - o Discoveries made by Domain Teams working together
- **Individual DUR project-based manuscripts**. Manuscripts put forth by members of the N3C community describing research, software development, etc., but do not involve substantial contributions from the N3C community workstream(s). Such manuscripts and associated works do not represent the N3C community as a whole. Project-based manuscripts are largely limited to discoveries made with the available infrastructure as a result of one or more Data Use Requests (DURs) submitted and approved for access to the N3C Data Enclave.

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: 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 manuscript; provide data, figure, or table; 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 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: 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 to determine the masthead authorship. If a manuscript's critical findings depend on earlier work (e.g. phenotype definition, knowledge store products,

variable codeset), strong consideration should be given to masthead authorship for the key contributors to that earlier work. Masthead authors are named in the manuscript byline.

Block authorship: Block authorship is used to recognize the contribution of multiple investigators (e.g. Domain Teams or working groups). Block authors in any specific manuscript must be enumerated in a supplemental file to be MEDLINE listed. Mention of block authorship without enumeration is prohibited. Block authorship appears as "N3C Consortium" in the masthead. Note: Only Masthead and Block Authors are indexed in MEDLINE.

Acknowledgments: It is the expectation of the N3C community that every person who contributed to the work being published is recognized.

- General attribution: Will be given to those who provided the infrastructure necessary to create the publication, as:
  - "This research was possible because of the patients whose information is included within the data and the organizations and scientists who have contributed to the on-going development of this community resource https://doi.org/10.1093/jamia/ocaa196."
- Additional attribution: Should be given to individuals or groups who generally supported the publication but whose efforts are not consistent with ICMJE guidelines to warrant authorship.
- Grant acknowledgment statements: Funding sources must be acknowledged.
  - For publications involving use of data from the N3C Data Enclave, the following acknowledgment statement (using relevant grant numbers) must be included:
    - "The analyses described in this [*publication/report/presentation*] were conducted with data or tools accessed through the NCATS N3C Data Enclave https://covid.cd2h.org and N3C Attribution & Publication Policy v 1.2-2020-08-25b, and supported by NCATS U24 TR002306 and [*additional funding agencies or sources & reference numbers*]."
  - For all publications involving an academic institution that resides at a Clinical and Translational Research (CTR) site under the Institutional Development Award (IDeA) Program grant, the following acknowledgment statement must also be included:
    - "The project described was supported by the National Institute of General Medical Sciences, 5U54GM104942-04 and [*additional funding agencies or sources & reference numbers*]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH."

N3C Publication Committee

N3C Publication Committee: Consists of volunteer members of the N3C user community. 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.

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

- Abstracts, podium presentations and press releases should be submitted once they have been accepted by the conference to allow for promotion and to track output; however, these will not be reviewed by the Publication Committee. Authors are actively encouraged to consider inclusive authorship in their abstracts and podium presentations.

- All full manuscripts intended for journal publication will be submitted to the Publication Committee. The Publication Committee will:
    - Support the authors in confirming authorship and attributions.
    - Verify alignment with the Data Use Request (DUR).
    - Offer editing support on a case-by-case basis at the committee's discretion.

- Draft manuscripts will be examined by the Publication Committee within 10 business days of submission.

- Analytics pipeline: Authors are encouraged to reference a link to their analytics pipeline in the N3C Data Enclave as a meaningful demonstration of N3C's commitment to reproducibility.

- N3C Users will strive to publish in open-access journals and will strive to make their results immediately available. Immediate availability can be accomplished via preprint submission to the N3C Zenodo community or other established preprint registry. Published manuscripts should be submitted to PubMed Central to be made publicly available immediately upon acceptance to publication without an embargo period.

Confidentiality and endorsement

Draft manuscripts, abstracts, and podium presentations shared within N3C prior to publication will be treated as confidential within 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 NCATS staff prior to distribution.

Brief announcements are allowed but must include disclaimer language such as "these results represent the authors only".