

N3C Results Download Policy

Title: N3C Results Download Policy	
VersionNo: 3.0	Effective Date: 2023-01-11
Authors:	Chris Chute, Christine Suver, Johanna Loomba, Melissa Haendel, Shawn O’Neal, and Anita Walden

Version history

Version	Description of changes
All versions	10.5281/zenodo.7942011 This DOI represents all versions and will always resolve to the latest one.
1.0-2021-05-07	<ul style="list-style-type: none"> First draft (doi:10.5281/zenodo.4743235)
2.0-2022-03-01	Updates Section: Knowledge objects - modified Licensing Information updated Cell size threshold appeal process (doi:10.5281/zenodo.4743235)
3.0-2023-01-11	<ul style="list-style-type: none"> Added definitions and clarified the conditions for downloading operational results, knowledge artifacts, and images.

Goals

Investigators using the rich data compiled in the N3C Data Enclave are expected to generate quantitative results in the forms of tables, figures, parameter estimates, and aggregated statistics. These results may have broad impact and will be shared, typically in the form of manuscripts, reports, and visualizations for websites and seminars. Additionally, many knowledge artifacts may be developed in the Enclave and made useful to others by being exported for use in open-source communities such as in GitHub, or provided as supplementary materials to publications. The purpose of this policy is to clarify the conditions for sharing and downloading quantitative artifacts used in analyses, while at the same time ensuring protection and security of the data in the Enclave and compliance with the NCATS [Data Transfer](#) and [Data Use](#) Agreements.

Definitions

The following terms are used as defined for the purposes of this policy:

Results: Any artifact derived from data protected by a Data Use Request (DUR), including summary tables, graphical images or visualizations, or other quantitative information.

Download Request: Formal request for a table, image, or report to be reviewed by Results Download Review Subcommittee. This request is submitted via a form in the N3C Enclave.

Concept Sets (Value Sets): Sets of OMOP concept identifiers, or identifiers used by other medical or non-medical vocabularies not derived from N3C patient data.

Site IDs: Identifiers used to group sources of N3C patient data by contributing site/DTA, as represented in protected N3C data.

Protected Health Information (PHI): Individual's medical record and other personal health information protected under the HIPAA privacy rule

Program-Only Code: Software-defined processes written in Python, R, SQL, or another programming language for data analysis or visualization, that does not include any results (as defined above) or non-obfuscated site IDs (as defined above).

Knowledge artifact: A user-defined block of logic that enables user to infer meaning from data. Knowledge artifacts include workflows, concept sets, mappings or terminological artifacts, and phenotyping algorithms, as examples.

Research Results

Before researchers can request access to N3C data, their home institutions must have an [Institutional Data Use Agreement \(DUA\)](#) in place with NCATS. Access to restricted N3C data for research purposes is granted after approval of Data Use Requests (DURs), which outline the goals of the research to address COVID-19-related questions. Completing a DUR involves attesting to the [N3C Data User Code of Conduct](#) and this Results Download Policy, which requires approval for the export of all research findings in the form of quantitative, graphical, or tabular results. Such results may be exported from the N3C Data Enclave environment with the following provisions:

- No observations or patient data at an individual level may be downloaded.
- Aggregate data may be downloaded if the aggregation pertains to 20 or more persons in the N3C dataset. No table or figure may report cell results as an actual value deriving from fewer than 20 persons unless the value is zero. Cells with counts less than 20 will be identified with a symbol (often seen as "<20" or *). Users may appeal to reduce the

reportable cell size threshold to 10 based on scientific interpretation. Appeal requests must include the scientific rationale, indicate the context of how the data will be presented, and discuss the potential risks of re-identification and benefits for individuals and populations. Appeals may be approved on a case-by-case basis. Successful appeal would permit using a reporting moniker such as “<10” rather than <20 and reporting actual values between 10 and 19; under no circumstance would a successful appeal imply that actual values <10 could be reported. 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)”***

- The obfuscated site IDs used in the Enclave must be re-mapped to a protocol/paper-specific remapping, e.g. sites 1-30. Actual Enclave site IDs may not be used in any public documents.
- Ages may be included in the data unless the patient is over 89 years old. Patients over age 89 may be included in their combined category of ages 90 and older (often seen as “>89”). All quantitative data to be downloaded will be copied to a staging folder on the N3C Data Enclave. It will also be archived on an NCATS hosted permanent location for monitoring purposes.

Download Process

- Investigators must submit requests for data and image download to the Results Download Review Subcommittee. The Results Download Review Subcommittee of the N3C Publications Committee includes members with clinical, data science, and privacy backgrounds. NCATS and the Results Download Review Subcommittee will review requests and may refer them to other N3C committees for additional input. Additional details on how the operations of this committee will be implemented will be available in the Results Download Review process document.
- Metadata about the downloading user, including their ORCID, will be attached to the staging and archive area data.
- PHI and date detector software will be run over newly staged materials.
 - Dates may suggest row-level data, the download of which is prohibited
 - Five-digit zip codes as fields may also suggest row-level data
- A manual audit of all proposed data will be conducted to ensure it does not represent row-level data.
- The download requestor will be informed by email as to how to retrieve the manually-approved results.
- Datasets that are determined to violate these policies after manual review will be denied for downloading. Violations will be addressed and may lead to sanctions as outlined in the Code of Conduct and the NCATS N3C Data Use Agreement.
- Users’ access to the NCATS N3C Data Enclave must be renewed annually. Users’ continued access to their analyses and findings is contingent on renewing the approval of their Data Use Request.

Images/visualizations

Analytic workflows often generate graphical images or figures such as maps, histograms, violin plots, feature importance plots, and model performance plots. Although it is technically possible to directly download images, users who do so without an approved download request are in violation of policy. However, in order to export images in svg/png/jpg formats, users MAY download or screenshot any image that appears in a report that has been EXPLICITLY approved for download by the Results Download Review Subcommittee. Note that the Results Download Review Subcommittee will periodically inspect a random sample of logged image downloads to assess compliance with the Results Download Policy and whether any data on a patient instance has been exported. Violations will follow the process outlined in the User Code of Conduct and the NCATS N3C Data Use Agreement.

From Section F of the NCATS [N3C Data Use Agreement](#), users agree not to photograph, create screenshots, or download row-level data (or images representing non-aggregate data) viewed on the NCATS N3C Data Enclave.

Knowledge artifacts (logic, concept sets, or program code only; no data or protected health information)

Investigators and research teams will generate derived variables, harmonized data elements, concept sets/codesets, machine learning models, code workbooks, logical workflows, and other knowledge artifacts. None of these resources will include row-level data or PHI from clinical sources to enable voluntary and fair sharing. Most of these knowledge artifacts will be beneficial to other users within the N3C community. The goal is to enable the fair sharing of such resources with appropriate attribution and documentation (particularly through Zenodo and/or GitHub). Sharing knowledge artifacts, without row-level data or PHI, will not require prior approval from the Result Download Review Subcommittee but must be documented.

N3C researchers who export knowledge artifacts are responsible for ensuring that no data, no person counts, and no N3C data partner IDs are included in the artifacts code or files.

Special Circumstances

Operational Data Use Requests are designated for, operational, non-research uses of protected N3C data. Summary statistics of data findings approved by NCATS for sharing and made available on operational dashboards of the N3C Enclave website may be downloaded by

authorized personnel listed in the operational DURs that correspond to those dashboards via NCATS approved processes.

Note that screen sharing views of the N3C enclave does not constitute a download when no screenshots or recordings are made. However, to align with access policies, you can only screen share with other N3C users who have the same level of data access through an approved Data Use Request.

Release for archival purposes

Once any artifact is ready for public release - e.g. it is the final code or knowledge artifact used in a published manuscript or is otherwise ready for fully provenanced community reuse, an archival record with a DOI for the artifact may be created in the N3C Zenodo community. It is straightforward to create a versioned release archive in Zenodo directly from Github (see <https://guides.github.com/activities/citable-code/>). Please request the inclusion of your artifact in the [N3C Zenodo collection](#) so that the N3C community can more readily find them.