

Overview:

This is a reference for our ARIC CSCC statisticians, programmers, and ultimately our external end-users of ARIC datasets files, on how these files are named for distribution.

Summary of Website Datasets Policies and Procedures:

When they are finalized, dataset files are posted to the Secure Download portion of the ARIC website. This page and its files are only visible to those with Data Recipient access.

Details about upcoming dataset distributions are given in the Anticipated Data Delivery Plan document on the [Obtain/Submit Data page](#). ARIC data sharing policies are also housed publicly on this page.

Each listing (table row) on the Secure Download page has a descriptive Title, the Date, the Zip files, and the datasets files within.

Secure Download - Datasets

Category: Visits Items per page: 50

Visits			
Title	Visit	Date	Files
V7 Form/Reading Center & Derived Data	V7	02/28/2024	V7 Form V7 Derived V7 NCSDERV
V9 Form/Reading Center & Derived Data	V9	02/28/2024	V9 Form V9 Derived V9 NCSDERV
V6 Form/Reading Center & Derived Data	V6	02/28/2024	V6 Form V6 Derived V6 NCSDERV
V8T Form/Reading Center & Derived Data	V8T	02/28/2024	V8T Form V8T Derived V8T NCSTDERV
V5 Forms/Reading Center & Derived Data	V5	02/28/2024	V5 Core Forms Group 1 V5 Form Reading Center Group 2 V5 Derived V5 NCSDERV

General guidelines for datasets files:

1. **Only datasets containing no PHI (protected health information) are posted to the website.** These files typically contain “_np” in the filename.
2. Slashes are not to be included in filenames, as this can hinder import of the files by external users into other software systems.
3. When datasets are updated the files are replaced on the website, and version history is captured in the corresponding data dictionary.
4. Ancillary study codebooks are not posted separately on the website, but are packaged in with datasets files.
5. Interim files (form data and derived data) have a retrieval date on the dataset name. Final files do not have a retrieval date suffix on the dataset name.
6. Datasets with data across multiple visits have this information in the filename, such as V1_V5_ANALYTES, or V2_V#_CNFA, with # used to denote all visits onward.

Dataset Type	Subtype	Release Frequency	Codebook location on website	Zip File Naming Conventions	Dataset Naming Conventions
Surveillance	Form and Derived	Annual	Event Surveillance → Surveillance Data Collection	The annual Surveillance datasets include the event type and event year. Example: 2021 HF Surveillance Datasets Without PHI.zip	C - Cohort XX - Year of event freeze Formcode # - version number _np Ex: C21CELB1_NP
Retention and Follow-Up	Form and Form composite Final	Annual	Cohort Studies → Annual and Semi-Annual Follow-up	The annual Follow-Up Composite file includes the most recent contact year (CY), such as CY2-32. Ex: CY2-34.zip	SC request #_formcode_np
Visits	Form Raw Interim	Varies	Cohort Studies → Current and Archived Visit Documents	The word Interim is included in the zip file. More specific information may be listed (e.g., no lab data). Ex: Visit 10 Interim Form Data NP.zip	formcode_mmddyy_np
	Form Raw Final	Varies, based on Visit close	Cohort Studies → Current and Archived Visit Documents	The word Interim is removed.	formcode_np
	Derived Form Datasets	Varies, based on Visit close	Cohort Studies → Current and Archived Visit Documents → Derived/Lab/RC Codebooks	Some form datasets have derived computational data (but are not to be confused with the DERIVE datasets). These have derived variables specific to the form, and the naming convention is to add derv in the filename. Visit information is within the file. Ex: ncsderv61_np.zip	Follows the same convention Ex: ncsderv91_np
	Derived Visit-Based	Varies, based on Visit close	Cohort Studies → Current and Archived Visit Documents → Derived/Lab/RC Codebooks	The filename includes the visit number and version number. For example Version 1 of the derived dataset for Visit 5 is DERIVE51, or DERIVE*#, where * indicates visit number and # indicates the version. Ex: derive54_np.zip STATUS*# is another derived dataset for the	Follows the same convention. Ex: derive54_np

				cohort that uses such a naming convention. Ex: STATUS91_np.zip	
Visit Longitudinal	Derived Visit Longitudinal	Varies, based on Visit close	Cohort Studies → Current and Archived Visit Documents → Derived/Lab/RC Codebooks	The Visit numbers included in the dataset are given in the filename. MRI datasets have MRI in the name. Ex: V2_V9CNFA.zip	Follows the same convention. Ex: v2_v9_cnfa_np
Ancillary Studies	N/A	Varies, but typically quarterly to PI	Zipped with the dataset file	The ancillary study number is given, along with a clarifying name. Refer to codebook(s) zipped with the data for more filename information. Ex: AS2015.34(Echo).zip	Refer to codebook(s) zipped with the data for filename information.
Contributor-Derived	N/A	Varies	If available, zipped with the dataset file	Varies, but we ask Investigators to follow our general naming conventions.	Varies