

NAACCR Call for Data Data Retention Overview

Definitions:

- original data files – the xml files submitted by registries to the CFD submission web site.
- raw data files – files prepped for the recodes program. They were created from the original xml files (i.e. 12-month data removed, any edit/inter-record edit changes corrected, etc.).
- intermediate files – any file (and their derivatives) that have been run through the recodes program and are used to build SEER*Stat production and researcher databases.
- **NOTE:** Call for Data years reference the calendar year of submission, not diagnosis years.

Stage 1

For the current submission data year (i.e. 2025 CFD year, 1995-2023 dx years), the original xml data files submitted by registries, raw data files and all intermediate files are kept and maintained until August of 2028. And the original registry XML data files submitted for the upcoming submission year, 2026 CFD submission year, will be retained until August 2029.

- 2025 – we get data and have 2023 and 2024 raw data
- 2026 – we get data and have 2024 and 2025 raw data
- 2027 - we get data and have 2025 and 2026 raw data
- Etc.

For each new submission, we need to retain the raw data from the last 2 submission years (so, CFD 2026, we keep 2024 and 2025 raw data).

Registry submission files from the previous Call for Data submission (CFD-2025, 1995-2023 dx years) that include census tract are deleted.

Stage 2

Submissions prior to five years from the current (e.g, if current CFD is 2026, then five years back would be CFD-2021, 1995-2019) will have their intermediate data files deleted. So, if current CFD year is 2026, we would be deleting CFD-2020, 1995-2018 dx years. Existing SEER*Stat databases and the files used by SEER*Stat will be kept, but all data files used to build the databases (production and researcher db's) will be deleted (**no database with data from that cfd year will be able to be created or recreated**). The exception to this is the intermediate files used to build the delay databases. These will be retained until notified that it is acceptable to delete and would allow any given CFD delay database to be rebuilt.

Once Stage 1 and 2 Completed

Example: If current CFD is 2026

- CFD-2026 = all data files
- CFD-2025 = all data files
- CFD-2024 = all data files
- CFD-2023 thru CFD-2021
 - Raw data files deleted
 - Intermediate data files kept
- CFD-2020 and prior
 - Raw data files deleted (previously deleted)
 - Intermediate files deleted