OAK RIDGE INSTITUTE
FOR SCIENCE AND EDUCATION
Shaping the Future of Science

Introduction & Purpose

The Comprehensive Epidemiologic Data Resource (CEDR) is a United States (US) Department of Energy (DOE)-funded data repository for archiving historical epidemiologic data to ensure its usability for future generations of researchers and educators. In the interest of transparency, CEDR aims to describe the processes for data submission, access, security, and archiving. This document describes the aspects of the data lifecycle in which CEDR is actively involved. For additional information, review the frequently asked questions section on the CEDR website or contact the CEDR team directly at cedr@orau.org.

Data Access

Individuals requesting access to CEDR datasets should complete and sign the data request form found on the CEDR website. As part of this form, data users submit general information about themselves, their institution, the requested dataset(s), and the intended use of the data. Depending on the tier of data requested, data users may be asked to provide additional information such as an Institutional Review Board (IRB) determination letter to CEDR before proceeding.

All individuals are eligible to request data access through the CEDR submission process. The primary purpose of CEDR is to provide educators and scientists with access to data from historical epidemiologic studies to further scientific and educational goals. However, there are specific uses of the data which are not permitted in order to ensure the privacy, integrity, and ethical use of the information. These restrictions include uses for litigation, proof of employment, media/social media stories, or by worker advocacy groups. All applications and intended uses of the CEDR data must be approved by the DOE Program Manager, and they may approve or decline CEDR applications at their discretion. For specific questions pertaining to data requests, please contact the CEDR team at cedr@orau.org.

All CEDR users are expected to comply with the user responsibilities and guidelines specified in the applicable data use agreement (DUA). Unauthorized use of the CEDR data may result in the loss of access to CEDR data at the discretion of the DOE Program Manager.

Tiered Access Overview

CEDR's tiered-access system assigns data to one of three tiers. Each tier has requirements for access that are congruent with the sensitivity of the data they contain, ensuring that appropriate data privacy is maintained. An overview of the tiered-access system is provided below.

Tier	Data Types	Access Requirements	Accessible Data
Tier 1	• Aggregate • Exposure monitoring • Simulated • Non-human subjects' data[1]	• Submit initial Data Access Request Form • Fully executed Tier 1 Data Use Agreement	All data in tier
Tier 2	• Individual level • Demographic	• Submit initial Data Access Request Form • Submit Institutional Review Board determination letter or equivalent documentation • Fully executed Tier 2 Data Use Agreement	Only requested data
Tier 3	• Individual level • Demographic • Data with special considerations that may be sensitive to data sharing	• Submit initial Data Access Request Form • Submit Institutional Review Board determination letter or equivalent documentation • Fully executed Tier 3 Data Use Agreement • Additional information if requested by the data submission point of contact	Only requested data

[1]CEDR does not contain information from animal or cellular studies at this time. However, CEDR does contain data from studies that may receive a non-human subjects' research designation from the Institutional Reviews Board (e.g., when all study members are deceased).

Data Request Process

The data request forms are submitted through the CEDR website, and the application should be completed in English. An applicable DUA, corresponding with the tier of the data requested, will be sent to the data user for signature. The signed DUA will be returned to CEDR for archiving, additional signatures (if appropriate), and approval by the DOE Program Manager. The CEDR team will contact the data user directly if additional information is required for approval. If additional information is requested and no response is received, the CEDR team will send out three reminders as shown below.

- First Reminder – Sent two weeks after last contact if no response has been received.
- Second reminder – Sent two weeks after the first reminder if no response has been received.
- Final reminder – Sent two weeks after the second reminder if no response has been received.

If no response has been received after the three reminders, the initial data request form will be moved to inactive status. The data requests can be reactivated by contacting the CEDR staff.

Once the data request is approved by the DOE Program Manager, the data user will be given access to the requested data. As stated in the DUA, the data user agrees not to use the data for any purpose other than that stated in the data request. However, the data user may request to modify their stated purpose. If this occurs, the data user will notify CEDR of the proposed changes via email so approval for the request may be obtained from the DOE Program Manager. The fully executed DUA, data request form, and any modification to the data request will be maintained by CEDR in accordance with the appropriate federal records schedule. Additionally, any modifications to the original data request may also require submission of an IRB protocol modification approval letter when applicable.

Maintaining Access

CEDR conducts an annual review of active data users and their specified data use to ensure the data user still has a need for access to the CEDR data. If the data user no longer needs the CEDR data, the user access will be terminated. If the data user fails to respond to the annual review within a reasonable time (approximately 1 month), then access to CEDR will also be suspended. Data users with terminated access to CEDR data may be granted future access after approval of a new data request form. Data users may also relinquish their access to the CEDR data before the annual review by emailing the CEDR team to notify them of the data access closure.

Data Submission

CEDR relies on data submissions from researchers to continually grow the data repository, and ensure it remains beneficial for future educational and scientific uses. This section describes the process of submitting data to CEDR.

Submission Timeline

The timeline from data submission to publication on the CEDR website is variable and depends on the needs of the data submitter. If information and data are submitted simultaneously with no questions from CEDR staff, then it takes approximately three to four weeks to publish the data on CEDR. The data submitter will be notified if there is a delay in publishing beyond the specified timeframe. At the request of the data submitter, CEDR staff can temporarily delay the publication of the data on the website. This typically occurs when the data and information are submitted to CEDR in advance of manuscript publication. To temporarily delay the publication of the data on CEDR, the data submitter should notify the CEDR team either via email or in the data submission application process. If this occurs, the data will be withheld from publication on the website until the data submitter agrees to the data's publication. CEDR does not publish copyrighted manuscripts on the website, but CEDR will provide the citation to the associated copyrighted manuscript (if applicable).

Accepted Data Types

In accordance with its stated purpose, CEDR accepts data from environmental or health studies on DOE workers, contractors, subcontractors, former workers, and the communities surrounding DOE facilities, as well as DOE-funded health research. This includes a variety of study designs such as cohort, case-control, cross-sectional, and ecological studies with data covering multiple scientific areas including morbidity and mortality, dosimetry and health physics, industrial hygiene and occupational safety, and congenital defects.

Although biological samples may be associated with some of the CEDR datasets (i.e., Radium Dial Workers), CEDR does not maintain or accept biological samples. Typically, CEDR

does not accept data from cellular studies, animal studies, or from clinical trials. However, other types of data may be accepted at the discretion of the DOE Program Manager. CEDR accepts a broad range of data file types, though a csv and xlsx file are preferred. For additional information on the types or formats of data accepted by CEDR, please contact cedr@orau.org.

Submission Preparation

The team at CEDR recognizes that submitting data to a repository can be daunting and time consuming, especially as a study is concluding. To ease the burden of data submissions, CEDR leverages information that researchers generate as a normal part of the study process for the data submission. The table below provides the specific information needed for a data submission and provides additional details helpful for preparing the data submission. If questions persist, please contact CEDR directly at cedr@orau.org.

The information in the table represents a typical data submission for United States based research. While CEDR welcomes data from international collaborations, the program recognized that data sharing policies may differ across countries or regions, and the program requests that related materials (data dictionary, study summary, etc.) be provided in English. For those seeking to submit international data to CEDR, please contact the CEDR via email to ensure our international data requirements are met.

CEDR Submission Information	Explanation
Final analysis file(s)	Researchers can submit their data as a single data file or file set (i.e., multiple data files).
Data Dictionary	The data dictionary should contain a list of variable names, definitions, and important notes for each variable in each data file. For all studies using International Classification of Diseases (ICD) codes, researchers should provide both the ICD version and code(s) or code range for the variable creation. All quantitative values should include a unit of measure. Categorical variables should define each value (i.e., 1 = males, 2 = females, etc.).
Study Summary	A brief summary of the study with which the data are associated should be included. This is typically the abstract for published works or executive summaries for governmental reports.
'At a Glance' Information	This section provides prospective data users with a brief snapshot of the data. Include study title, total study size, study population race(s) and sex(es), disease(s) of interest (can be very broad, like cancer or non-cancer mortality), earliest exposure date (as applicable), latest exposure date (as applicable), exposure types (if applicable), exposure agent (if applicable), and covariates.
Publication Citation	The full citation for the published study or report should be included. If more than one publication resulted from the data, provide all citations so they can be linked appropriately to the dataset. If data are part of an unpublished manuscript or report, CEDR can hold the data until publication. If publication is not sought (i.e., unpublished work), CEDR will work with the researcher to provide the appropriate acknowledgement.
Data Masking or Suppressing Information	As part of its normal process, CEDR masks all dates and suppresses categories with less than 10 members. Additionally, data fields may be suppressed or masked at the request of the data submitter if there are additional privacy concerns. Typically,

	the masking and data suppression processes are performed by the CEDR team after the data are submitted, but the data submitter may request via email to perform the masking or suppression themselves prior to data submission. If this option is preferred, then the data submitter should notify CEDR via email of any masked or suppressed data fields so the appropriate indication can be added to the 'Study Summary.'
Indication of Analytic or Working Files	Files on CEDR are considered either analytic or working files. Researchers can submit both types simultaneously or submit only one type at a time. Analytic files are the final datasets used in the analysis, typically with all data cleaning and exclusions applied. Working files may include more variables than just those used in the final analysis, as well as data from individuals who were excluded at a later point in the study. When submitting data, please indicate the appropriate category in the documentation.
Principal Investigator or Corresponding Author Contact Information	Provide the contact information, specifically a working email address, for the principal investigator or corresponding author for the data submission. Occasionally, the CEDR team will receive additional questions or special requests from data users and will contact the principal investigator or corresponding author.

Submitting Data

To initiate the data submission process, an introductory email should be sent to cedr@orau.org that indicates the intent to submit data to CEDR and includes a very brief overview of the data. If possible, please provide the number of data files and a rough estimate of the file size. Someone from the CEDR team will respond within 3 business days to acknowledge the email, provide more information on the data submission, and answer any questions from the submitter. Data files are typically submitted to CEDR via encrypted email or encrypted USB drive (supplied by CEDR). For data that are emailed, please package multiple files in a password protected zip file, if possible, and send the password in a separate email to ensure data security while in transport. All other documentation for the data can be sent via email or included in the encrypted USB drive. If neither of these data submission options are feasible, please contact the CEDR team to request an accommodation (cedr@orau.org).

Data Deidentification and Masking

The CEDR team will ensure the data are sufficiently deidentified through a combination of data masking and data suppression. The standard for de-identified data as it pertains to CEDR follows the definition provided by the DOE Central Institutional Review Board, under HRP-001-SOP-Definitions. For the process of de-identifying data, CEDR follows the National Institute for Standards and Technology (NIST) SP 800-188 (De-Identifying Government Datasets: Techniques and Governance) which defines this process as, "a general term for any process of removing the association between a set of identifying data and the data subject." Data submitted directly to CEDR as deidentified data will be reviewed by CEDR team to ensure compliance with DOE and other applicable Federal requirements and generally accepted practices for digital data. Questions or concerns about de-identification or masking should be directed to the CEDR staff (cedr@orau.org).

Data Tier Assignment

After the submission is received, the CEDR team will assign the data to the appropriate access tier to ensure data privacy. The data submitter will have an opportunity to review the tier assignment and request changes. For more information on the CEDR tiered access system, please refer to the CEDR website. After reaching agreement regarding access tier and ensuring the completion of all required documentation, the data will be made available on the CEDR website unless the researcher requests a hold prior to publication. In this instance, the data and documentation will be held in development until the researcher notifies CEDR of the article's publication. If there are questions or concerns about the tier assignment process, then please contact the CEDR team directly (cedr@orau.org).

Post-Data Publication

Once the data are published on the CEDR website, the data submitter's engagement with CEDR will be minimal. The CEDR staff will maintain contact information (i.e., email address) for the data submitter in case questions arise about the data, but this information will not be shared with external CEDR users. Please contact the CEDR team directly if updates to the contact information are required.

Data submitters with data classified as Tier 3 can expect to be contacted for data access request approvals for their specific data by the CEDR team. Annually, CEDR will review and verify the contact list for the data submitters responsible for the data in Tier 3 to ensure the contact information is still active and the data should still be categorized as Tier 3. If the CEDR team is unable to contact the Tier 3 data submitter, then approval of the Tier 3 data will revert to the DOE CEDR Program Manager unless a suitable alternative is found. Examples of suitable alternatives are the funding organization or federal agency that has oversight of the research.

Data Security and Archiving

Data Security for CEDR Data Users

The data must be secured in compliance with the Federal Information Security Modernization Act of 2014 ((2014 Pub. L. 113-283) (FISMA) (44 U.S.C. § 101 note)). This Act requires that data users protect the confidentiality, integrity, and availability of the data using appropriate technical, administrative, and operational safeguards. The required safeguards, including specific security controls and compliance requirements, are described in NIST Special Publication 800-53, revision 5, "Security and Privacy Controls for Information Systems and Organizations" and in NIST Special Publication 800 -171r3 "Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations."

The following are required safeguards for the data:

1. Controls must restrict physical and electronic access to the authorized data users.
2. Electronic data stored on a file server must be on a secured file server and access authenticated using strong passwords or biometrics. Data stored on portable media, including laptops, must be protected by encryption meeting Federal standards (see above).
3. Physical copies of data should be kept to a minimum and must be physically secured while in use, locked while not in use, and shredded using a cross-cut or better shredder upon disposal. Physical copies may not be removed from the approved data use location.

4. Cryptographic mechanisms must be used to protect the confidentiality and integrity of information stored on digital media during transport outside of controlled areas. Data must not be transmitted unencrypted.
5. Derivative data created from these data must be managed in the same manner as the original data, unless otherwise approved by DOE CEDR Program Manager.
6. CEDR login credentials will not be shared with other individuals.

Data Security for Archiving Data at ORISE

Oak Ridge Associated Universities (ORAU), the contractor who manages the CEDR under the Oak Ridge Institute for Science and Education (ORISE) conducts both automated and manual security analysis of custom developed applications. Additionally, ORISE performs vulnerability scanning of the application hosting environment and uses a third party to conduct independent external vulnerability scanning. In addition to these mechanisms, ORISE has robust software development practices that include security engineering, peer review, and software quality assurance. ORISE fields multiple systems using NIST 800-53 and NIST 800-171 as common controls frameworks for our operating environment. These control frameworks govern information and cyber security, as well as privacy control implementations for all systems fielded by ORISE, including the solutions ORISE has developed and is managing for CEDR. Those solutions are developed in accordance with NIST 800-171 security and privacy control requirements for Federal information systems and include implementation of all baseline security controls commensurate with the Federal Information Processing Standard (FIPS) 199 system security categorization of Moderate. ORISE will protect data in CEDR in accordance with the NIST Moderate classification. Additionally, all data submitted to CEDR will follow the appropriate federal records schedule for data retention.