Data Preservation in the DOE Comprehensive Epidemiologic Data Resource (CEDR)

RHSP – Celebrating 30 Years of Scientific Achievements

ICRP

Comprehensive Epidemiologic Data Resource (CEDR)

- **Created by SPEERA 1990 Report**
  - Secretarial Panel for the Evaluation of Epidemiologic Research Activities
- **Program Goals**
  - Increase transparency for worker & public health concerns
  - Advance science by making data available for validation and further evaluation
  - Support academic communities – data for students & teachers
- **Managed by ORISE for EHSS13 since 2002**

https://OriseApps.orau.gov/cedr/

CEDR

- **What?**
  - **Free, de-identified** data repository
  - **1.5M+** workers
  - **400+** datasets
  - **4500+** citations
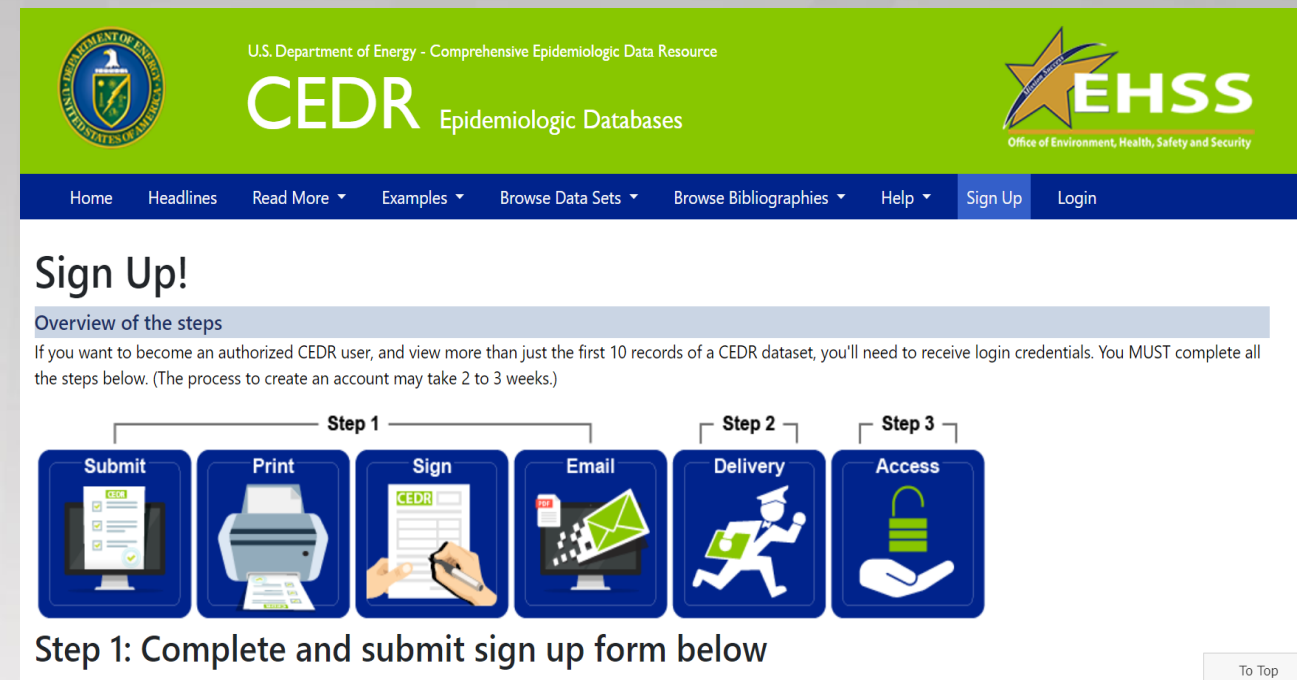 - **200+** historical documents
- **Who?**
  - **108** Current users
  - Average monthly visitors: **>5,500** from **39** countries, **32** states
  - Submitters: researchers using DOE data (or funds) for health-related studies
  - Users: researchers (public, private), teachers, students, historians, federal employees/policy makers

OAK RIDGE INSTITUTE
FOR SCIENCE AND EDUCATION

Access and Data Security

- Account required for full access
- All account requests are reviewed/approved by DOE
- Access is re-verified annually with new password assignment
- Additional enhancements on-going:
  - Implementation of Data Use Agreement
  - Tiered access based on account approval

- Data are de-identified
  - Unique study IDs, can be randomized
- Dates are masked: 15th of mon/year
- Data or variables may be suppressed
  - Health outcomes <5-10 cases
  - Request of PI
- Data can be categorized or summarized

U.S. Department of Energy - Comprehensive Epidemiologic Data Resource

CEDR Epidemiologic Databases

EHSS
Office of Environment, Health, Safety and Security

Home | Headlines | Read More ▾ | Examples ▾ | Browse Data Sets ▾ | Browse Bibliographies ▾ | Help ▾ | Sign Up | Login

Sign Up!

Overview of the steps

If you want to become an authorized CEDR user, and view more than just the first 10 records of a CEDR dataset, you'll need to receive login credentials. You MUST complete all the steps below. (The process to create an account may take 2 to 3 weeks.)

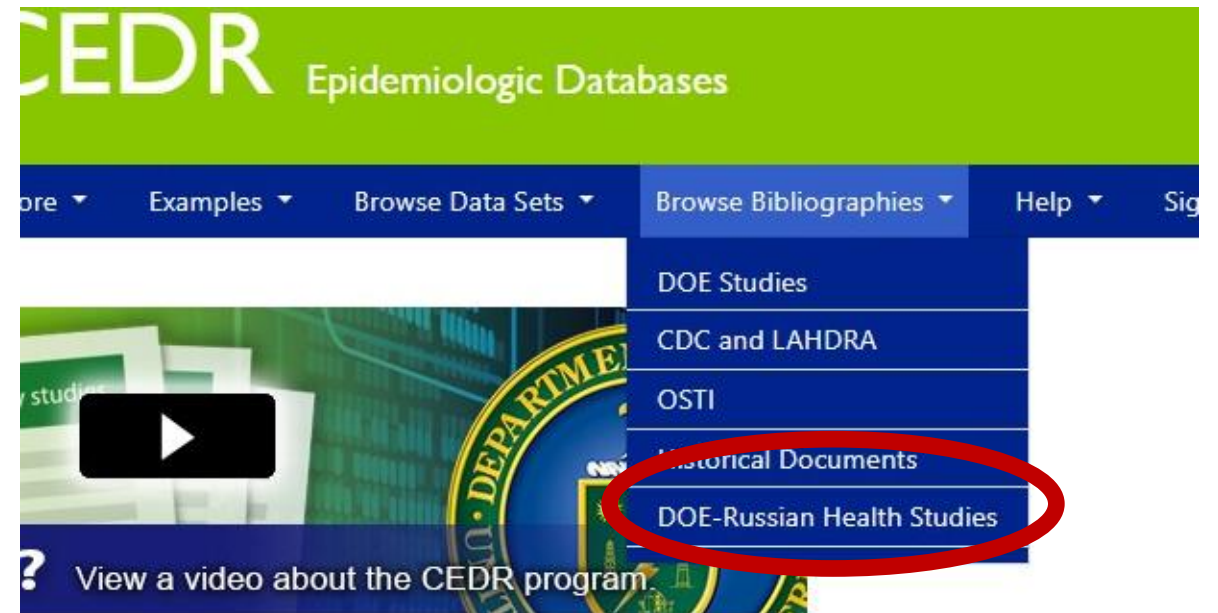Step 1: Submit | Print | Sign | Email   Step 2: Delivery   Step 3: Access

Step 1: Complete and submit sign up form below

To Top

Phase 1 – Archiving Publications

- **Make RHSP publications available for download through CEDR**
- **380 citations/manuscripts received**
  - 17 citations are missing PDFs
  - 111 PDFs (including 7 missing) are in Russian only
    - English versions available?

- **Add RHSP section**
- **Add tags for searching**
- **Add 508 compliance**
- **Update EHSS13 library**
  - Non-public access

Epidemiologic Data Resource (CEDR)

Phase 2 – Archiving Data for Future Use

- **Archive RHSP data, making available through CEDR as appropriate**
  - Archive section will be non-public
  - Considerations for which datasets can be made "public"
    - What approvals are needed?
    - What protections (masking, suppression, categorization, person-time tables, etc.)?
- **ORISE Approach**
  - Provide PIs examples of templates for data files and data dictionaries.
  - Discuss and implement data protections for each specific dataset with PIs.
  - Tag datasets and data dictionaries to ensure searchability.
  - Upload all files and connect to search functions
  - Upload of abstract (or study description) and citations associated with data.
  - QA/QC with study PI to ensure all elements are accurately displayed.

ICRP

CEDR – Tiered Access Approach

Ease of Access

Easiest

Hardest

Tier Access Level 1 - Access to all to low risk datasets for educational purposes
- All individuals are deceased (ex: Radium dial painter data)
- Aggregate data (ex: person-time tables)
- Simulated data
- Demographic variables (sex, race, marriage status, etc.) are removed.
  - Only year of birth included, unless 5 or few individuals are born in a given year
- Only 3 characters provided for ICD codes

Tier Access Level 2 – Access to low and medium risk datasets for research purposes
- IRB determination required
- Access limited to requested datasets and variables
- Users may have access to all demographic data within specific dataset(s), if required

Tier Access Level 3 – Higher risk/concern datasets that require additional attention.
- IRB protocol required **and** approval by specified "owner" of dataset(s)
- Access limited to requested datasets and variables
- Example: The Russian Health Studies Program or the Former Worker Program data

What's the Goal?

- **Preserve the excellent work of the RHSP**
  - Highlight the valuable research and results achieved
  - Ease of future re-start, or use, of the program data and information
  - Make data and information available to other researchers

- **Assist EHSS13 in complying with Executive Order (EXEC-2022-006008) from OSTP by 2025**
  - *"Ensuring, Free, Immediate and Equitable Access to Federally Funded Research"*

ICRP

www.icrp.org