ReadMe file for National Energy Water Treatment and Speciation (NEWTS) Integrated Dataset

Database Name: NEWTS Integrated Dataset (version 1.0)

Citation: Nicholas Siefert, Lucy Romeo, Isabelle Pfander, Rachel Yesenchak, Madison Wenzlick, Sophia Bauer, Zineb Belarbi, Justin Mackey, Burt Thomas, NEWTS Integrated Dataset (version 1.0), 5/31/2024, https://edx.netl.doe.gov/dataset/newts-integrated-dataset-version-1-0, DOI: 10.18141/2368855

Description:

The National Energy Water Treatment and Speciation (NEWTS) Integrated Dataset v1.0 provides water researchers, community leaders, and regulators with a unified and standardized energy-related wastewater stream database. This resource is derived from 27 state and federal entities, and scientific publications, and contains more than 400,000 sample records, many of which also provide geospatial information. The dataset includes data for several different energy-related wastewater types including produced water, other oil and gas wastewaters, mine drainage, coal ash leachate, and power plant wastewater. The NEWTS Integrated Dataset was built to support environmentally prudent decision-making, explore treatment opportunities, and identify potential critical mineral sources. A subset of this novel resource is also featured on NETL NEWTS State-Level Database Dashboard. Additional data can be found in the NEWTS EDX Group and the NEWTS Federal Database Dashboard.

More information on the NEWTS Integrated Dataset can be found in an upcoming publication, and can be found in the following resources:

- ***NETL-Developed Online Database Brings Energy-Related Wastewater Stream Data to Public's Fingertip, 2023***
- ***Water-Energy Nexus News, 2022***
- ***Resource Sustainability Project Review Meeting, 2022***

Acknowledgements: This work was performed in support of the U.S. Department of Energy's Fossil Energy and Carbon Management through the National Energy Technology Laboratory's (NETL) Research & Innovation Center's Water Management for Power Systems Field Work Proposal, [DE-FECM1611080].

recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Date of publishing: 5/31/2024

Geographic coordinate system: WGS 1984

Points of Contact: Rachel Yesenchak, rachel.yesenchak@netl.doe.gov

NETL Reviewer: Nicholas Siefert, nicholas.siefert@netl.doe.gov

Resources list for data release:

1. **Geodatabase**
    a. Contains all spatially available data from **NEWTS_Integrated_Full_052024.csv**, with state, county, and Hydrologic Unit Codes 2 (regions) and 8 (subbasins) appended as additional columns.
2. **Integrated Dataset CSV Files**
    a. **NEWTS_Integrated_Text_052024.csv** – Integrated resources with no unit conversions completed, rather unit fields are added alongside each measured field and filled in with unit terms for transparency.
    b. **NEWTS_Integrated_Full_052024.csv** – Integrated resource with unit conversions and reformatting included.
3. **OLI Studio/Geochemist's Workbench Files**
    a. **GWB_processed_reordered.csv** - The full integrated dataset formatted for input into Geochemist's Workbench. Data must be transposed prior to input into software.
    b. **GWB_processed_reordered_1val_min.csv** - Integrated dataset records with at least 1 relevant measured value formatted for input into Geochemist's Workbench. Data must be transposed prior to input into software.
    c. **GWB_processed_reordered_4majors_min.csv** - Integrated dataset records with at least 4 major cations measured, formatted for input into Geochemist's Workbench. Data must be transposed prior to input into software.
    d. **GWB_4majors_min_15000MostComplete_Transpose.csv** - The 15,000 most complete integrated dataset records formatted for input into Geochemist's Workbench. Data has been transposed and some GWB output are included.
    e. **NEWTSintegratedDataset_GWB_15000MostComplete.gss** - Includes the same data from GWB_4majors_min_15000MostComplete_Transpose.csv, but in a .gss file
    f. **NEWTSintegratedDataset_GWB_template.gss** - GWB template to copy the transposed GWB formatted data in mg/L
    g. **OLI_processed_reordered.csv** - The full integrated dataset formatted for input into OLI Studio. Data must be transposed prior to input into software.
    h. **OLI_processed_reordered_1val_min.csv** - Integrated dataset records with at least 1 relevant measured value formatted for input into OLI Studio. Data must be transposed prior to input into software.

i. **OLI_processed_reordered_4majors_min.csv** - Integrated dataset records with at least 4 major cations measured, formatted for input into OLI Studio. Data must be transposed prior to input into software.
j. **OLI_processed_charge_balance.csv** - Charge balance calculations for data included in the OLI_processed_reordered.csv file.
k. **OLI_processed_charge_balance_1val_min.csv** - Charge balance calculations for data included in the OLI_processed_reordered_1val_min.csv file.
l. **OLI_processed_charge_balance_4majors_min.csv** - Charge balance calculations for data included in the OLI_processed_reordered_4majors_min.csv file.
m. **OLI_4majors_min_15000MostComplete_Transpose.xlsx** - The 15,000 most complete integrated dataset records, formatted for input into OLI Studio.
n. **oli_integrateddataset_template.oad** - OLI template to copy the transposed OLI formatted data in mg/L
o. **oli_inetegrateddataset_casestudy_NEWTSID2525.oad** - OLI Studio case study using the mine drainage data stream from NEWTS ID 2525
p. **oli_integrateddataset_casestudy_NEWTSID59744.oad** - OLI Studio case study using the mine drainage data stream from NEWTS ID 59744
q. **oli_integrateddataset_casestudy_NEWTSID240587.oad** - OLI Studio case study using the produced water data stream from NEWTS ID 240587

4. **NEWTS_Metadata_ReadMe.pdf** (this document) - List of references included are available in table on pages 4-5.
5. **NEWTS_FieldDictionary_052024.xlsx** – Field dictionary showing mapped attributes, which resources contained each attribute, and how the attributes are defined.

Processing and Integration Notes:

Processing steps and methods were unique to each original resource. As such, below are a list of methods applied, when relevant. After processing, dataset attributes were mapped to identify common and unique attributes. This attribute mapper was then used as a guide to integrate the datasets into one resource, using customized Python scripts.

- General Initial Processing Protocol
    - Filtering
        - Data were filtered to include only sample types of interest. For example, data are filtered for only water samples, leaving out gas and solids.
        - When applicable, data were further filtered to only include wastewater streams of interest (produced water, mine drainage, coal ash leachate, etc.) leaving out other sample types such as groundwater.
        - If maximum measurements were reported and equivalent to average reported measurements for an entire column, only average values were retained and the maximum column was removed.
        - If values were reported as negative (except for isotopic values), they were removed.
        - For larger datasets (e.g., U.S. Geological Survey's USGS Produced Water), only the top ~15,000 most complete records were integrated.
    - Data Reformatting and Cleaning
        - Units and analyte names were cleaned to have consistent formatting and spelling.

        - Commas were removed from field names to accommodate csv format.
    - Handling Detection Limits
        - If a measurement was below the reported detection limit, the value was changed to 0.
    - Unit Processing
        - When measurements were reported in multiple units per analyte, units were appended to analyte names and are separated into unique columns. For example, columns "Calcium_mg/L" and "Calcium_ug/L" were separated.
        - If alkalinity, hardness, or acidity measurements were reported only as mg/L and we could not confirm whether it was reported as CaCO3, HCO3, etc. it was separated into a new column.
    - Transposing Data
        - Data were transposed to produce a row for every sample.
        - When a sample ID was present, this is used as an index.
        - If the data did not contain a sample ID, relevant fields (date, location, sample type etc.) were used to create a unique identifier for every row.
    - The resulting dataset has every analyte and metadata field as a separate column.
- Attribute Mapping Protocol
    - Column names were matched to other column names with the same meanings and measurements, not units. Units were handled during integration and conversion.
    - Study-specific fields that would not be interpretable or useful in the context of an integrated dataset (e.g. activity codes) were not attribute mapped and are not included in the final dataset.
    - For each mapped attribute, the input format (e.g., double, long, text, date) and units (when applicable) were included when available in the original resource metadata or through communication with said resources point of contact.
    - When not specified if measurements were field or lab, measurements were assumed as lab.
    - When non-metallic element concentrations were labeled using the pure element symbol (Br, Cl), they were mapped to the ionic form (bromide, chloride).
    - Average values were mapped over a monitoring period to singular measurement values.
- Integration Protocol
    - If a measurement was reported with "<" it was set to 0.
    - If a measurement was reported with ">", the ">" was removed and the associated numeric value was maintained.
    - If original measured values did not have units reported in the metadata or through the resource, they were considered unfit for conversion and are not available in the integrated numeric version.
    - If one mapped final attribute had multiple representative columns in the original resource (e.g., two columns representing a single anion measurement), the integration script treats the multiple potential input fields as an ordered list and goes from first to last, using whichever has the first non-null value. If there was a value available it will use it, if not it moves to the next one and looks for a value.
    - If a spatial reference was not available in the original resource, but latitude and longitude were included in decimal degrees, the locations were defined in World Geodetic System 1984.

Table 1. The 28 resources from which data was acquired, curated, and included in the current version of the NEWTS integrated database, listed by Source Name ('newts_source_name' in the NEWTS database) and the full citation.

Source Name	Citation
USGS_Produced_ **Water**	Blondes, M.S., Gans, K.D., Engle, M.A., Kharaka, Y.K., Reidy, M.E., Saraswathula, V., Thordsen, J.J., Rowan, E.L., & Morrissey, E. A. (2019). U.S. Geological Survey National Produced Waters Geochemical Database v2.3 [Data set]. U.S. Geological Survey. https://doi.org/10.5066/F7J964W8
USGS_Produced_ Water_Dataset_ **V3**	Blondes, M.S., Knierim, K.J., Croke, M.R., Freeman, P.A., Doolan, C., Herzberg, A.S., and Shelton, J.L., 2023, U.S. Geological Survey National Produced Waters Geochemical Database (ver. 3.0, December 2023): U.S. Geological Survey data release, https://doi.org/10.5066/P9DSRCZJ.
CO_ECMC_ **Processed_Water**	Colorado Oil and Gas Conservation Commission. (2014). COGCC Oil and Gas Analytical Database [Data set]. https://ecmc.state.co.us/documents/data/downloads/environmental/ProdWellDownLoad.html
Cravotta_Brady_ **2015**	Cravotta III, C.A. and Brady, K.B. (2015). Priority pollutants and associated constituents in untreated and treated discharges from coal mining or processing facilities in Pennsylvania, USA. [Data set]. https://www.sciencedirect.com/science/article/pii/S0883292715000591
EIP_Coal_Ash_ **Groundwater**	Environmental Integrity Project. (2019). Coal Ash Rule Groundwater Monitoring Results Database [Data set]. https://environmentalintegrity.org/coal-ash-groundwater-contamination/
ME_Power_Plant_ NonContact_ Cooling_Water_ **Effluent**	Maine Department of Environmental Protection. (2023). Power plant non-contact cooling water effluent data. [Data set]. Acquired through email correspondence.
MD_Coal_Mines	Maryland Department of the Environment. (2023). Coal Mine Effluent [Data set]. Acquired through email correspondence.
MD_Generating_ **Stations**	Maryland Department of the Environment. (2023). Generating Stations Effluent [Data set]. Acquired through email correspondence.
MN_NPDES_ **Effluent**	Minnesota Pollution Control Agency. (2009). Discharge Monitoring Reports Data [Data set]. https://files.pca.state.mn.us/pub/file_requests/datasets/Wastewater/DMR/
NM_Coal_Mine_ **Water_Quality**	New Mexico Energy, Minerals and Natural Resources Department. (2023). Coal Water Quality Sample & Parameter Data [Data set]. https://catalog.newmexicowaterdata.org/dataset/coal-water-quality-sample-parameter/resource/a1a0a6af-abc3-44e5-8ea5-9cf0c818e993
NM_Chino_Mine	New Mexico Environment Department. (2023). Chino Mine Water Quality Monitoring Data [Data set]. Acquired through email correspondence.
Continental_ **Mine_Water**	New Mexico Environment Department. (2023). Continental Mine Water Quality Monitoring Data [Data set]. Acquired through email correspondence.
NM_Tyrone_ **Copper_Mine_1**	New Mexico Environment Department. (2023). Tyrone Mine Water Quality Monitoring Data [Data set]. Acquired through email correspondence.
NM_Tyrone_ **Copper_Mine_2**	New Mexico Environment Department. (2023). Tyrone Mine Water Quality Monitoring Data [Data set]. Acquired through email correspondence.
EPA_Leachate	Nguyen, Dan-Tam, Eastern Research Group. Sep 29, 2015. Ash Landfill Leachate Data. [Data set]. Analytical Database for the Steam Electric Rulemaking - DCN SE05359. https://www.regulations.gov/document/EPA-HQ-OW-2009-0819-5640
EPA_Gasification_ **Effluent**	Nguyen, Dan-Tam, Eastern Research Group. Sep 29, 2015. US EPA Gasification Effluent Database [Data set]. Analytical Database for the Steam Electric Rulemaking - DCN SE05359. https://www.regulations.gov/document/EPA-HQ-OW-2009-0819-5640
Nicot_et_al_2020	Nicot, J.P., Darvari, R., Eichhubl, P., Scanlon, B.R., Elliott, B.A., Bryndzia, L.T., Gale, J.F.W., Fall, A. (2020). Origin of low salinity, high volume produced waters in the Wolfcamp Shale (Permian), Delaware Basin, USA [Data set]. https://doi.org/10.1016/j.apgeochem.2020.104771
OH_Power_Plant_ **Energy_Station**	Ohio Environmental Protection Agency. (2023). Industrial NPDES Effluent (Mines, Powerplants, Sub-stations) [Data set]. Acquired through email correspondence.
OH_Groundwater_ 2000_2003	Ohio University. (2023). Ohio Groundwater Data Related to Abandoned Mines. [Data set]. https://watersheddata.com/GroundMap/map.aspx
OH_Surface_ Water_2000_ **2003**	Ohio University. (2023). Ohio Surface Water Data Related to Abandoned Mines. [Data set]. https://watersheddata.com/map/Map.aspx?WaterShed=OTHER

Pashin_et_al_**2014**	Pashin, J.C., McIntyre-Redden, M. R., Mann, S. D., Kopaska Merkel, D. C. (2014). Water Chemistry Data [Data set]. Geological Survey of Alabama. https://netl.doe.gov/files/oil-gas/fe0000888-final-report.pdf
WA_EIM	Raforth, R. L., Norman, D. K., & Johnson, A. (2004). Third Screening Investigation of Water and Sediment Quality of Creeks in Ten Washington Mining Districts, with Emphasis on Metals [Data set]. Washington State Department of Ecology. https://apps.ecology.wa.gov/eim/search/Detail/Detail.aspx?DetailType=Study&SystemProjectId=89574601
USGS_Brackish_**Water**	United States Geological Survey. (2020). National Brackish Groundwater Assessment 2017 [Data set]. https://dx.doi.org/10.7481/1787482
Quillinan_et_al_**2017**	University of Wyoming. (2017). Rare Earth Element Concentration of Wyoming Thermal Waters Update [Data set]. https://dx.doi.org/10.15121/1364771
VA_Power_**Facilities**	Virginia Department of Environmental Quality. (2023). Discharge Monitoring Report Data [Data set]. Acquired through email correspondence.
Silverton_CO_**Mine_Drainage**	Walton-Day, K., Runkel, R.L., Mast, M.A, and Qi, S.L. (2020). Water-quality and discharge data from draining mine tunnels near Silverton, Colorado 1988-2015 [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P9FE667O
WI_Power_Plant_**Outfalls**	Wisconsin Department of Natural Resources. (2023). Power Plant Discharge Monitoring Report Data [Data set]. Acquired through email correspondence.
WY_Produced_**Waters**	Wyoming Oil & Gas Conservation Commission. (2022). Water Analysis Records [Data set]. Acquired through email correspondence.