

APPENDIX A

Data Processing Information (information on FracFocus, USGS, and spill data)

PURPOSE AND SCOPE

This appendix is intended to provide instructions on processing data that are included in the Water Support Document. For each dataset described in this appendix, various data processing applications may be used to process the data, depending on user preference (e.g., Excel or R statistical software [R]). Additionally, there are multiple approaches within each application to generate the same information (e.g., in Excel, the use of pivot tables, copying data into new tabs to use the Remove Duplicates button, or using filters; in R, various functions to aggregate and summarize data). Therefore, these instructions provide basic aggregation rules and specific column names in the datasets to accommodate different user preferences and styles of approaching data management.

In the water support document, some counties span multiple field offices. In that instance, it is possible that a county can be associated with multiple field offices. Data for the county will be reported in full for each field office that it overlaps.

FRACFOCUS DATA PROCESSING INSTRUCTIONS

Data Acquisition and Preparation

1. Download FracFocus data from <https://fracfocus.org/data-download>
2. The Water Support Document analyzes data from 2014 onward. The 2021 Water Support Document considers FracFocus data from 2014 to 2020. The 2022 Water Support Document will consider data from 2014 to 2021, and so on.
3. Isolate data for desired year using column heading JobStartDate, which is the “date on which the hydraulic fracturing job was initiated” (FracFocus 2021) and state (e.g., New Mexico). One Excel sheet for each year is helpful for data management and will also reduce file size.
 - a. The file named readme.txt in the data download packet is the FracFocus data dictionary and should be retained with the original downloads.

Data Screening and Quality Control

The following data checks are intended to evaluate and validate the consistency, completeness, and uniqueness of FracFocus data. In this process, records that do not meet the following data quality criteria are flagged and are rejected from analysis. Flagged records are not deleted but are marked as flagged in a new column in the data and are not included in further data aggregation.

1. For the purpose of this analysis, a drilling activity (a job) is defined as the job start date and the well name. Data users should check that the same well does not have multiple job start dates within the same year. If so, the job should be appended to differentiate the two jobs so that the two different water volumes are counted separately. These records should be included in data aggregation. This can be corrected by creating a new column in the data using the date alone (and not the date and time) to form a new primary key in the data. A new, unique primary key can be generated by combining the date and well name.
2. Any row where TotalBaseWaterVolume = 0 gallons should be flagged and rejected from analysis based on the assumption that all drilling activities require water. Therefore, if a well reports 0 gallons of water use, it is likely erroneous data and should be rejected from analysis.
3. American Petroleum Institute (API) well identification numbers are assumed to be a unique identifier in the data, and there should be a 1:1 relationship between API number and well name. Differing well names having the same API number should be flagged and rejected from analysis, as this indicates a non-unique API number (e.g., a 1:2 relationship). Similarly, if the same well

name is given two different API numbers, these records should be flagged and rejected from analysis.

4. Federal well designation should be mutually exclusive. A well can either be federal or non-federal but not both. Wells that are given both designations should be flagged and rejected from analysis.
5. TotalBaseWaterVolume is “the total volume of water used as a carrier fluid for the hydraulic fracturing job (in gallons)” (FracFocus 2021). For each job (e.g., a unique job start date and well name), there should be one reported total base water volume for all rows. Any jobs that have differing total base water volumes should be flagged and rejected from analysis.

Unit Conversions

Water use in FracFocus is reported in gallons and water use in the Water Support Document is reported in AF. The following conversion factors can be used to convert from gallons to AF and vice-versa:

$$\begin{aligned}1 \text{ AF} &= 325,851 \text{ gallons} \\1 \text{ gallon} &= 3.0689 \times 10^{-6} \text{ AF}\end{aligned}$$

Data Aggregation

To present the summarized information in tables summarizing water use by oil and gas wells for hydraulic fracturing in the state of New Mexico from 2014 through 2020, FracFocus data are processed and aggregated by various factors such as year and water use by both federal and non-federal wells. The following instructions describe the process by which the summarized totals are obtained. The data totals do not include the records that were flagged in the Data Screening and Quality Control section of this document.

For each job (note that a job is the well name and job start date) in the FracFocus data, there are many rows to document the various ingredients and chemicals used in the drilling activity, and the total base water volume is duplicated in each row for the specific ingredient. Therefore, the duplicate entries for total base water volume will need to be removed to generate totals. Duplicate rows can be removed by one of the following methods:

- A. Creating a new primary key using JobStartDate and WellName, copying the data to a new tab, and then removing duplicates based on the new primary key.
- B. Aggregating values in an Excel pivot table using average, minimum, or maximum, but not sum.

Federal Water Use: For each year (2014–2020), federal water use is the sum of total base water volumes reported for each job where the well is identified as a federal well. Duplicate total base water volumes for the various chemicals are not included in the total. More specifically, there should be one total base water volume for each job. An example pivot table setup is included in Figure A-1. The filters remove the flagged data and non-federal wells. Total federal water use is the sum of the total base water volumes for each job in AF.

Filters SWCA_QAQC_Status FederalWell		Columns	
Rows WellName		Σ Values Average of TotalBase...	

Figure A-1. Example pivot table setup to generate federal water use data.

Non-Federal Water Use: Similar to federal water use, non-federal water use is generated for each year of data (2014–2020). Non-federal water use is the sum of total base water volumes reported for each job where the well is identified as a non-federal well. Duplicate total base water volumes for the various chemicals are not included in the total. The example pivot table setup for non-federal water use is the same as for federal wells (see Figure A-1), however, the filter should be set to show FederalWell = FALSE. The units are acre-feet.

Total Water Use: Federal water use plus non-federal water use for each year of FracFocus data (2014–2020), reported in AF.

Federal Water Use Percentage: Obtained by dividing the federal water use by the total water and multiplying by 100. Federal water use percentage is calculated for each year of FracFocus data.

Federal Cumulative Water Use: For any given year in the FracFocus data, the federal cumulative water use is that year's federal water use plus the sum of all previously reported federal water use estimates. For example

$$2020_{FCWU} = 2020_{FWU} + 2019_{FWU} + 2018_{FWU} + 2017_{FWU} + 2016_{FWU} + 2015_{FC} + 2014_{FC}$$

Where FCWU is federal cumulative water use and FWU is federal water use

Total Cumulative Water Use: Similar to federal cumulative water use, the total cumulative water use is that year's total water use plus the sum of all previously reported total water use estimates.

Average Water Use Per Well: The average water use for all wells in the data (both federal and non-federal).

Total Number of Federal Wells: The number of federal wells drilled in any given year.

USGS DATA PROCESSING INSTRUCTIONS

The following processes are intended to guide obtaining data from the USGS 2015 data file to include in the state and field office water use tables throughout the Water Support Document.

Data Acquisition and Preparation

1. Download *Estimated Use of Water in the United States County-Level Data for 2015* (Dieter et al. 2018) from <https://www.sciencebase.gov/catalog/item/get/5af3311be4b0da30c1b245d8>
 - a. File name: usco2015v2.0.xlsx All Data XLSX

Data Aggregation

To present the summarized water use data in tables throughout the Water Support Document, USGS data are processed and aggregated by state and county. The instructions below describe the process by which the summarized totals are obtained.

State of New Mexico Water Use: For each county in the USGS data, there are many columns to document the various types of water usage. The total water use is listed per county in each state, so total water use per category for the state must be manually generated. Follow the steps listed below to generate totals for the state of New Mexico.

1. Isolate data for the state of New Mexico using the column titled STATE, copy the data to a new tab, and generate state grand total values (in AF). Grand total values should be a sum of all county values for each water use. A pivot table in Excel can be used to do this, and an example pivot table setup is included in Figure A-2.
 - a. Columns selected for values can include all columns
 - b. Retain the Excel data dictionary with the original data using the DataDictionary tab in the downloaded data file.
2. Once grand totals are calculated, copy and paste data into a new tab as values, making sure to transpose the data.
 - a. It is helpful to set up a definition lookup table for the abbreviated column names by using the translations found in the DataDictionary tab in the original data during this step.
3. Filter all data so the column tag can be filtered to fill out data tables later.

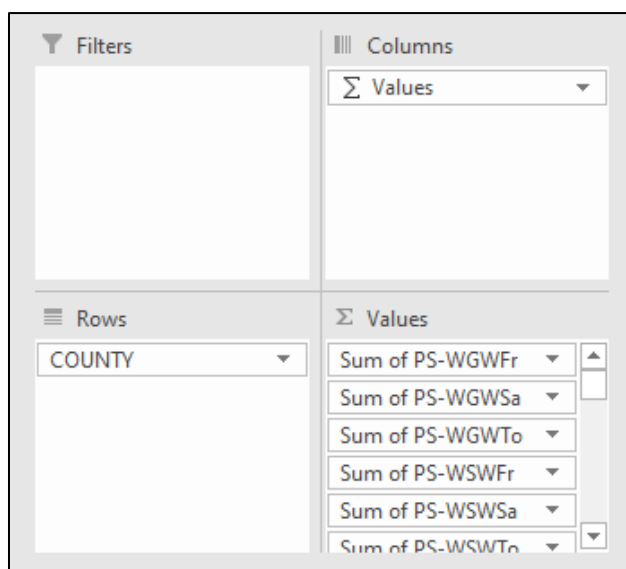


Figure A-2. Example pivot table setup.

County Water Use: Similar to state water use, county water use is generated for each county in each field office. The following steps are used to prepare the data for the County Water Use data table entry.

1. Isolate data by filtering the original data by STATE = NM. For each county, use the column titled COUNTY to filter data further.
2. Create a new row to generate totals per field office and use the =sum Excel formula to sum each county value per water use.

Unit Conversions

Water use in the USGS data is reported in million gallons per day, and water use in the Water Support Document is reported in AF. The following conversion factors can be used to convert from gallons to AF and vice-versa.

$$\text{Grand total in AF per year} = (\text{Grand Total [MGD]} \times 1.121) \times 1,000$$

Data Tables

To present the summarized information in Table 2-1 in the Water Support Document (*State of New Mexico Water Use by Category in 2015*), and individual county water use data included in each field office chapter, USGS data are processed and aggregated. The following instructions describe the process by which the summarized total water use values are obtained. These instructions provide specific column names in the USGS data to guide data entry.

State of New Mexico Water Use: State water use for each water use category is included in the USGS data. Each entry in Table 2-1 in the Water Support Document corresponds with a specific column header in the USGS data. For each category and water use (surface water, groundwater, and total withdrawals), refer to Figure A-3 when pulling data from the USGS data. Text in Figure A-3 and associated data dictionary terms in Table A-1 refer to the specific column tag that should be used for each data entry.

Category	Surface Water				Groundwater				Total Withdrawals					
	Fresh	Saline	Total	Total Use (%)	Fresh	Saline	Total	Total Use (%)	Fresh	Total Use (%)	Saline	Total Use (%)	Total	Total Use (%)
Aquaculture	AQ-WSWFr	AQ-WSWSa	AQ-WSWTo	AQ-WSWTo / TO-Wtotl	AQ-WGWFr	AQ-WGWSa	AQ-WGWTo	AQ-WGWTo / TO-Wtotl	AQ-WFrTo	AQ-WFrTo / TO-Wtotl	AQ-WSaTo	AQ-WSaTo / TO-Wtotl	AQ-Wtotl	AQ-Wtotl / TO-Wtotl
Domestic	DO-WSWFr	---	DO-WSWFr	DO-WSWFr / TO-Wtotl	DO-WGWFr	---	DO-WGWFr	DO-WGWFr / TO-Wtotl	DO-WFrTo	DO-WFrTo / TO-Wtotl	--	DO-WFrTo / TO-Wtotl	DO-WFrTo	DO-WFrTo / TO-Wtotl
Industrial	IN-WSWFr	IN-WSWSa	IN-WSWTo	IN-WSWTo / TO-Wtotl	IN-WGWFr	IN-WGWSa	IN-WGWTo	IN-WGWTo / TO-Wtotl	IN-WFrTo	IN-WFrTo / TO-Wtotl	IN-WSaTo	IN-WSaTo / TO-Wtotl	IN-Wtotl	IN-Wtotl / TO-Wtotl
Irrigation	IR-WSWFr	---	IR-WSWFr	IR-WSWFr / TO-Wtotl	IR-WGWFr	---	IR-WGWFr	IR-WGWFr / TO-Wtotl	IR-WFrTo	IR-WFrTo / TO-Wtotl	--	IR-WFrTo / TO-Wtotl	IR-WFrTo	IR-WFrTo / TO-Wtotl
Livestock	LI-WSWFr	----	LI-WSWFr	LI-WSWFr / TO-Wtotl	LI-WGWFr	---	LI-WGWFr	LI-WGWFr / TO-Wtotl	LI-WFrTo	LI-WFrTo / TO-Wtotl	--	LI-WFrTo / TO-Wtotl	LI-WFrTo	LI-WFrTo / TO-Wtotl
Mining	MI-WSWFr	MI-WSWSa	MI-WSWTo	MI-WSWTo / TO-Wtotl	MI-WGWFr	MI-WGWSa	MI-WGWTo	MI-WGWTo / TO-Wtotl	MI-WFrTo	MI-WFrTo / TO-Wtotl	MI-WSaTo	MI-WSaTo / TO-Wtotl	MI-Wtotl	MI-Wtotl / TO-Wtotl
Public Water Supply	PS-WSWFr	PS-WSWSa	PS-WSWTo	PS-WSWTo / TO-Wtotl	PS-WGWFr	PS-WGWSa	PS-WGWTo	PS-WGWTo / TO-Wtotl	PS-WFrTo	PS-WFrTo / TO-Wtotl	PS-WSaTo	PS-WSaTo / TO-Wtotl	PS-Wtotl	PS-Wtotl / TO-Wtotl
Thermoelectric	PT-WSWFr	PT-WSWSa	PT-WSWTo	PT-WSWTo / TO-Wtotl	PT-WGWFr	PT-WGWSa	PT-WGWTo	PT-WGWTo / TO-Wtotl	PT-WFrTo	PT-WFrTo / TO-Wtotl	PT-WSaTo	PT-WSaTo / TO-Wtotl	PT-Wtotl	PT-Wtotl / TO-Wtotl
Totals	TO-WSWFr	TO-WSWSa	TO-WSWTo	TO-WSWTo / TO-Wtotl	TO-WGWFr	TO-WGWSa	TO-WGWTo	TO-WGWTo / TO-Wtotl	TO-WFrTo	TO-WFrTo / TO-Wtotl	TO-WSaTo	TO-WSaTo / TO-Wtotl	TO-Wtotl	TO-Wtotl / TO-Wtotl

Figure A-3. Abbreviated column names for water use tables.

Table A-1. Data dictionary terms and associated abbreviated column names for water use data.

Dictionary Term	Abbreviated Column Name
Aquaculture, groundwater withdrawals, saline	AQ-WGWSa
Aquaculture, groundwater withdrawals, total	AQ-WGWTot
Aquaculture, surface-water withdrawals, fresh	AQ-WSWFr
Aquaculture, surface-water withdrawals, saline	AQ-WSWSa
Aquaculture, surface-water withdrawals, total	AQ-WSWTot
Aquaculture, total withdrawals, saline	AQ-WSaTot
Aquaculture, total withdrawals, total (fresh+saline)	AQ-Wtotl
Aquaculture, total withdrawals, fresh	AQ-WFrTot
Domestic, self-supplied groundwater withdrawals, fresh	DO-WGWFr
Domestic, self-supplied surface-water withdrawals, fresh	DO-WSWFr
Domestic, total self-supplied withdrawals, fresh	DO-WFrTot
Industrial, self-supplied groundwater withdrawals, fresh	IN-WGWFr
Industrial, self-supplied groundwater withdrawals, saline	IN-WGWSa
Industrial, self-supplied groundwater withdrawals, total	IN-WGWTot
Industrial, self-supplied surface-water withdrawals, fresh	IN-WSWFr
Industrial, self-supplied surface-water withdrawals, saline	IN-WSWSa
Industrial, self-supplied surface-water withdrawals, total	IN-WSWTot
Industrial, self-supplied total withdrawals, saline	IN-WSaTot
Industrial, self-supplied total withdrawals, total (fresh+saline)	IN-Wtotl
Irrigation, groundwater withdrawals, fresh	IR-WGWFr
Irrigation, surface-water withdrawals, fresh	IR-WSWFr
Irrigation, total withdrawals, fresh	IR-WFrTot
Livestock, groundwater withdrawals, fresh	LI-WGWFr
Livestock, surface-water withdrawals, fresh	LI-WSWFr
Livestock, total withdrawals, fresh	LI-WFrTot
Mining, groundwater withdrawals, fresh	MI-WGWFr
Mining, groundwater withdrawals, saline	MI-WGWSa
Mining, groundwater withdrawals, total	MI-WGWTot
Mining, surface-water withdrawals, fresh	MI-WSWFr
Mining, surface-water withdrawals, saline	MI-WSWSa
Mining, surface-water withdrawals, total	MI-WSWTot
Mining, total withdrawals, fresh	MI-WFrTot
Mining, total withdrawals, saline	MI-WSaTot
Mining, total withdrawals, total (fresh+saline)	MI-Wtotl
Public Supply, groundwater withdrawals, fresh	PS-WGWFr
Public Supply, groundwater withdrawals, saline	PS-WGWSa
Public Supply, groundwater withdrawals, total	PS-WGWTot
Public Supply, surface-water withdrawals, fresh	PS-WSWFr

Dictionary Term	Abbreviated Column Name
Public Supply, surface-water withdrawals, saline	PS-WSWSa
Public Supply, surface-water withdrawals, total	PS-WSWTo
Public Supply, total withdrawals, fresh	PS-WFrTo
Public Supply, total withdrawals, saline	PS-WSaTo
Public Supply, total withdrawals, total (fresh+saline)	PS-Wtotl
Thermoelectric, groundwater withdrawals, fresh	PT-WGWFr
Thermoelectric, groundwater withdrawals, saline	PT-WGWSa
Thermoelectric, groundwater withdrawals, total	PT-WGWTo
Thermoelectric, surface-water withdrawals, fresh	PT-WSWFr
Thermoelectric, surface-water withdrawals, saline	PT-WSWSa
Thermoelectric, surface-water withdrawals, total	PT-WSWTo
Thermoelectric, total withdrawals, saline	PT-WSaTo
Thermoelectric, total withdrawals, total (fresh+saline)	PT-Wtotl
Thermoelectric, total withdrawals, fresh	PT-WFrTo
Total groundwater withdrawals, fresh	TO-WGWFr
Total groundwater withdrawals, saline	TO-WGWSa
Total groundwater withdrawals, total (fresh+saline)	TO-WGWTo
Total surface-water withdrawals, fresh	TO-WSWFr
Total surface-water withdrawals, saline	TO-WSWSa
Total surface-water withdrawals, total (fresh+saline)	TO-WSWTo
Total withdrawals, fresh,	TO-WFrTo
Total withdrawals, saline	TO-WSaTo
Total withdrawals, total (fresh+saline)	TO-Wtotl

County Water Use: Similar to state water use, county water use is generated for each county in each field office. Using the specific county data of interest and Figure A-3, data can be entered into tables in each field office section of the Water Support Document.

Total Water Use Percentage: Total water use % is generated individually by dividing the total water use of a specific category in either surface water, groundwater, or total withdrawals by the total water use for the state, county, or field office of interest. See Figure A-3 and Table A-1 for guidance.

SPILL DATA ANALYSIS METHODOLOGY

Spill data are available for download from the NMOCD spills database located at <https://wwwapps.emnrd.nm.gov/ocd/ocdpermitting/data/Spills/Spills.aspx> (NMOCD 2021). The entire spills database contains records with incident dates ranging from 1900 to 2021 (at the time this update to the report was written). For each update to this report, spill data will be analyzed for the year of the report revision. For example, the 2020 Water Support Document will summarize records in the spills database with incident dates in the year 2020.

A spills data dictionary from NMOCD is not available to accompany the data. Therefore, several assumptions and definitions were made about the data that are summarized below. These data checks are intended to evaluate and validate the consistency, completeness, and uniqueness of spill data. In this process, records that do not meet the data quality criteria are flagged and are rejected from analysis.

Flagged records are not deleted but are marked as 'Flagged' in a new column in the data and are not included in further data aggregation.

Data Acquisition and Preparation

1. Downloaded spill data for all spill reports in all counties and filtered out the 2020 records (1,556 records)
2. Created a primary key in the data to identify and remove duplicates. *Primary key=Incident Number_Spilled Material*
3. Flagged duplicate records, records where Volume Spilled was 0 and where the entry for 'Unit of Volume' was missing or non-volumetric (185 records). The remaining record count for 2020 is 1,371.
4. Created a new column for Percent Loss, which is defined as follows:
$$\text{Percent Recovery} = (\text{Volume Lost/Volume Released}) * 100$$
5. Converted gallons to bbl using the following conversion factor:
$$1 \text{ gallon} = 0.02381 \text{ bbl}$$
6. Created a new column in the data to identify field office or district based on county. The Pecos District = Leah, Chaves, and Eddy Counties.

Data Screening and Quality Control

- For the purpose of this report, a *spill* is defined as the loss of a measurable volume of a material on the same day.
- The incident number is not unique, and for any one incident number, there may be many spill materials.
- Incidents where Volume Released is 0 are flagged and rejected from further analysis because these records are not in alignment with the definition of a spill, where a measurable volume of material has been released.
- Incidents where Unit of Volume is not volumetric (e.g., pounds [lbs]) are flagged and rejected from analysis.
- Records where the spill material type is Natural gas (methane) or Natural gas liquids should be reported in mcf and not bbl. Records where Spilled material is Natural gas (methane) and Natural gas liquids and Unit of Volume is bbl were rejected from analysis. Records where the material is natural gas flared or vented are not counted in the spills summary on the assumption that these are lost to the air.

Data Aggregation

To present the summarized information in the spill tables throughout the Water Support Document, spill data are processed and aggregated by various factors such as field office and material. The following instructions describe the process by which the summarized totals are obtained. The data totals do not include the records that were flagged in the Data Quality Review and Assumptions section of this document.

An example pivot table setup is included in Figure A-4. Using a pivot table or other data aggregation method, the values for the data descriptions below can be generated.

Filters <div>FieldOffice ▼</div>	Columns <div>Σ Values ▼</div>
Rows <div>Material ▼</div>	Σ Values <div>Count of Material ▼</div> <div>Sum of FinalVolumeR... ▼</div> <div>Sum of FinalVolumeL... ▼</div> <div>Average of FinalVolu... ▼</div> <div>Average of PercentLoss ▼</div>

Figure A-4. Example pivot table setup to generate volume released, volume lost, and percent lost.

Spill Count: Spill count is the number of spill records within a field office for a particular material.

Volume Spilled: Volume released is a sum of the volume released for all spills of a particular material within a field office. The data should be filtered to remove the flagged data.

Volume Lost: Volume lost is a sum of the volume lost for all spills of a particular material within a field office. The data should be filtered to remove the flagged data.

Average Spill Volume: Average volume spilled is an average of the volume released for all spills or all types of material within a field office. Data should be filtered to not include flagged data.

Percentage Lost: Percentage lost is the percentage of the volume spilled that was also lost. An average of percentage lost for all material spilled in a field office can be used to calculate the average percent of volume lost in spills across the entire field office for all spills.

LITERATURE CITED

Dieter, C.A., M.A. Maupin, R.R. Caldwell, M.A. Harris, T.I. Ivahnenko, J.K. Lovelace, N.L. Barber, and K.S. Linsey. 2018. *Estimated Use of Water in the United States in 2015*. U.S. Geological Survey Circular 1441. Available at: <https://pubs.er.usgs.gov/publication/cir1441>. Accessed November 2021.

Barber, N.L., and Linsey, K.S..2018. *Estimated Use of Water in the United States in 2015*. U.S. Geological Survey Circular 1441. Available at: <https://doi.org/10.3133/cir1441>. Downloaded November 2021. Accessed November 2021.

FracFocus. 2021. FracFocus data download. Available at: <http://fracfocus.org/data-download>. November 1, 2021.

New Mexico Oil Conservation Division (NMOCD). 2021. 2020 Spill Data. Available at: <https://wwwapps.emnrd.nm.gov/ocd/ocdpermitting/data/Spills/Spills.aspx>. Accessed November 3, 2021.

This page intentionally left blank.