

Links: [Navigator Page](#) | [Chemical Index](#) | [State Index](#) | [Operator Index](#)



Open-FF Data Dictionary



Open-
FF

This file was generated on October 09, 2025
from data repository: openFF_data_2025_10_07.

Sponsored
by
**FracTracker
Alliance**

Description of the contents of the final data files generated by Open-FF from the FracFocus data.

Pulling repo tables from: G:\My Drive\production\repos\openFF_data_2025_10_07\pickles

Acceptable use of FracFocus data

One requirement for using the FracFocus data is stipulated on the FracFocus website:

"Downloaded data may be aggregated or combined with other datasets, but the FracFocus data may not be altered in any way."

Please read the entire "Terms of use" at <http://fracfocus.org/data-download>.

The work in this project maintains the original FracFocus data as is reported in the bulk download. The field names used in the original are kept: All of these original names begin with an upper-case letter and can be identified in that way. Fields generated by this project or from external data sources will begin with a lower case letter (for example, `CASNumber` is the original field, `bgCAS` is the generated field. Note there are two exceptions: `DTXSID` and `MI_inconsistent` are NOT original with FracFocus.)

In the zipped bulk download from FracFocus, a data dictionary is provided in the '**readme.txt**' file. (This zipped download is in the /sources or /data directory and we rename it as 'currentData.zip') This file gives some information about many of the fields found; however, it is written for the SQL database version of the bulk download, not the CSV version which we use in this project. Further, some important fields are not mentioned in that readme.txt file; they are described below. In the descriptions of all fields below, we cite the FracFocus text from a June 2021 bulk download.

Descriptions of fields in the output data sets

Explanation of columns in the table below

column	what it is
fieldName:	The name of the field or column in the data set. All field names that are capitalized are from the original FracFocus downloaded data. Lower-case names are generated by Open-FF. tables: Which Open-FF internal tables that are used to construct output data sets have this field
FracFocus description:	Description of the (original) field given by FracFocus in the bulk download file, <i>readme.txt</i> .
Open-FF description:	Our description of the field
source:	is this field a direct copy of the original FracFocus data or is it generated by Open-FF, or pulled from an external data set?
Num:	the number of non-empty values in the field
Unique:	the number of unique types (including NaN) in the field
Data_type:	the python/pandas data type for the field

Search:

field Name, [tables]	FracFocus description	
APINumber ['disclosures']	The American Petroleum Institute well identification number formatted as follows xx-xxx-xxxxx0000 Where: First two digits represent the state, second three digits represent the county, third 5 digits represent the well.	T th ir F
CASNumber ['cas_ing', 'chemrecs']	The Chemical Abstract Service identification number.	
CountyName ['disclosures']	The name of the county were the surface location of the well resides. Calculated from the API number.	
DTXSID ['bgCAS']		E
DisclosureId ['water_source', 'disclosures', 'chemrecs']	Key linking to the RegistryUpload table.	
FFVersion ['disclosures']	A key which designates which version of FracFocus was used when the disclosure was submitted.	M cl u th
FederalWell ['disclosures']	True = Yes False = No.	(t N cl b F u
IndianWell ['disclosures']	(definition not provided)	(t A C d ir W *,
IngredientComment ['chemrecs']	Any comments related to the specific ingredient.	L cl
IngredientKey none		F S
IngredientName ['cas_ing', 'chemrecs']	Name of the chemical or for Trade Secret chemicals the chemical family name.	
IngredientsId	Key index for the Ingredients data	

field Name, [tables]	FracFocus description
['chemrecs']	
JobEndDate ['disclosures']	The date on which the hydraulic fracturing job was completed. Does not include site teardown.
JobStartDate ['disclosures']	The date on which the hydraulic fracturing job was initiated. Does not include site preparation or setup.
Latitude ['disclosures']	The lines that circle the earth horizontally, running side to side at equal distances apart on the earth. Latitude is typically expressed in degrees North/ South. In the FracFocus system these lines are shown in decimal degrees and must be between 15 and 75.
Longitude ['disclosures']	The lines that circle the earth vertically, running top to bottom that are equal distances apart at the equator and merge at the geographic top and bottom of the earth. Longitude is typically expressed in degrees East/ West. In the FracFocus system the number representing these lines are shown in decimal degrees and must be between -180 and -163 Note: Longitude number must be preceded by a negative sign.
MI_inconsistent ['disclosures']	
MassIngredient ['chemrecs']	Mass in pounds of the ingredient used on Job
OperatorName ['disclosures']	The name of the operator.
PercentHFJob ['chemrecs']	The amount of the ingredient in the total hydraulic fracturing volume in % by Mass.
PercentHighAdditive ['chemrecs']	The percent of the ingredient in the Trade Name product in % (Top of the range from MSDS).
Projection ['disclosures']	The geographic coordinate system to which the latitude and longitude are related. In the FracFocus system the projection systems allowed are NAD (North American Datum) 27 or 83 and UTM (Universal Transverse Mercator).
PurpCount ['purpose']	
PurpYears	

field Name, [tables]	FracFocus description	O
['purpose']		
Purpose ['chemrecs']	The reason the product was used (e.g. Surfactant, Biocide, Proppant).	
StateName ['disclosures']	The name of the state where the surface location of the well resides. Calculated from the API number.	
Supplier ['chemrecs']	The name of the company that supplied the product for the hydraulic fracturing job (Usually the service company).	
TVD ['disclosures']	The vertical distance from a point in the well (usually the current or final depth) to a point at the surface, usually the elevation of the rotary kelly bushing.	M
TotalBaseNonWaterVolume ['disclosures']	The total volume of non water components used as a carrier fluid for the hydraulic fracturing job (in gallons)	
TotalBaseWaterVolume ['disclosures']	The total volume of water used as a carrier fluid for the hydraulic fracturing job (in gallons).	
TradeName ['chemrecs']	The name of the product as defined by the supplier.	
UploadKey none		T d
WellName ['disclosures']	The name of the well.	W w d
api10 ['disclosures']		A s l n A
auto_carrier_type ['disclosures']		f fi
bgCAS ['cas_ing', 'bgCAS', 'chemrecs']		t C v a
bgCountyName ['disclosures']		C 'C
bgDensity		T

field Name, [tables]	FracFocus description	O
['disclosures']		
bgDensity_source ['disclosures']		C Ir
bgFederalLand none		A la g a cl (c
bgIngredientName ['bgCAS']		th re si se
bgLatitude ['disclosures']		C o A n c
bgLocationSource ['disclosures']		Ir T C v C
bgLongitude ['disclosures']		C p c a lc b
bgNativeAmericanLand none		A d o w cl fi H
bgOperatorName ['disclosures']		o si xl d
bgPurpose ['chemrecs']		
bgSource none		ir ir

field Name, [tables]	FracFocus description	O A l a t P ` C 'S o tr xl d n c T T a (F (th N b T N T t T T T N T T F d is ir (i V fi V
bgStateLand none		
bgStateName ['disclosures']		
bgSupplier ['chemrecs']		
calcMass ['chemrecs']		
calcMass_unfiltered ['chemrecs']		
carrier_density_MI ['disclosures']		
carrier_density_from_comment ['disclosures']		
carrier_mass ['disclosures']		
carrier_mass_MI ['disclosures']		
carrier_percent ['disclosures']		
carrier_problem_flags ['disclosures']		
carrier_status ['disclosures']		
cleanMI ['chemrecs']		

field Name, [tables]	FracFocus description	O
commercial_status ['bgCAS']		T
d_flags filter flag		ir s p d
data_source ['disclosures']		a st d st fc
date ['disclosures']		Ir C a
date_added ['disclosures']		th lr
density_from_comment ['chemrecs']		T
disc_ID ['disclosures']		d f d c a
dup_rec ['chemrecs']		T a st
earliest_poss_date ['disclosures']		T g d ir
eh_Class_L1 ['bgCAS']		C a C U
eh_Class_L2 ['bgCAS']		S a C U
eh_function ['bgCAS']		P C F P

field Name, [tables]	FracFocus description	O
eh_source ['bgCAS']		S H O G
eh_sub_class none		E
eh_subs_class ['bgCAS']		T V a
epa_pref_name ['bgCAS']		T C W a
epa_reg_flag ['bgCAS']		T T C
first_detected ['disclosures']		T w cl
has_TBWV ['disclosures']		T T C
has_water_carrier ['disclosures']		T w cl
huc10 ['disclosures']		
huc10_name ['disclosures']		
huc2 ['disclosures']		
huc2_name ['disclosures']		
huc8 ['disclosures']		
huc8_name ['disclosures']		

field Name, [tables]	FracFocus description	O
in_std_filtered filter flag		T st n
ingKeyPresent ['chemrecs']		d
ingredCommonName ['cas_ing', 'chemrecs']		C fo
is_duplicate ['disclosures']		T F C d
is_on_AQ_CWA ['bgCAS']		Ti E
is_on_CWA ['bgCAS']		ir co
is_on_DWSHA ['bgCAS']		ir a
is_on_HH_CWA ['bgCAS']		Ti E
is_on_IRIS ['bgCAS']		Ti E
is_on_NPDWR ['bgCAS']		Ti N
is_on_PFAS_list ['bgCAS']		Ti E
is_on_TEDX ['bgCAS']		Ti E
is_on_TSCA ['bgCAS']		Ti th
is_on_UVCB ['bgCAS']		Ti th p

field Name, [tables]	FracFocus description	O
is_on_diesel ['bgCAS']		Ti E th
is_on_prop65 ['bgCAS']		T
is_valid_cas ['chemrecs']		Ti o a
is_water_carrier ['chemrecs']		T
iupac_name ['bgCAS']		T
job_mass ['disclosures']		T th
job_mass_MI ['chemrecs']		T M
latlon_too_coarse ['disclosures']		T L - c w
loc_name_mismatch ['disclosures']		Ir o C si th
loc_within_county ['disclosures']		Ti th
loc_within_state ['disclosures']		Ti th
mass ['chemrecs']		n q ir re
massComp ['chemrecs']		ir N b

field Name, [tables]	FracFocus description	O
massCompFlag ['chemrecs']		ir
massSource ['chemrecs']		w a
max_d_warning filter flag		
max_r_warning filter flag		
no_chem_recs ['disclosures']		T
perc_gw_high_TDS ['water_source', 'disclosures']		R a
perc_gw_low_TDS ['water_source', 'disclosures']		R a
perc_other_high_TDS ['water_source', 'disclosures']		R h
perc_other_low_TDS ['water_source', 'disclosures']		R l
perc_pw ['water_source', 'disclosures']		R w
perc_sw_high_TDS ['water_source', 'disclosures']		R a
perc_sw_low_TDS ['water_source', 'disclosures']		R a
primarySupplier ['disclosures']		(a b s o d a c
pub_delay_days		N ir

field Name, [tables]	FracFocus description	O
['disclosures']		J tr
r_flags filter flag		
reckey ['chemrecs']		a
rq_lbs ['bgCAS']		T w
stLatitude ['disclosures']		a t e b
stLongitude ['disclosures']		a t e b
str_len ['purpose']		
total_percent_all_records ['disclosures']		a d
total_percent_of_valid ['disclosures']		a ir re
tsca_chem_definition ['bgCAS']		T "
within_total_tolerance ['disclosures']		T t
ws_perc_total ['disclosures']		S re

Showing 1 to 139 of 139 entries

Carrier detection sets:

Among the filters below, **s1** finds the majority of water carriers. However, there is no single set of criteria that can be used to identify the water carrier record(s) for all FracFocus disclosures. Therefore the other filters are employed to catch many other disclosure patterns without needing to curate each by hand.

Set name	description	Criteria to be detected
s1	Primary filter; most recent disclosures are detected with this	<ul style="list-style-type: none"> - Only one record whose <code>Purpose</code> is "carrier" (or related) - <code>bgCAS</code> is '7732-18-5' - at least 50% <code>PercentHFJob</code> - total % of disclosure is $95\% > x > 105\%$
s2	<p>More than one record as the carrier; covers situations, for example, where there are two water records (fresh and produced) and where other chemicals are also labeled as part of the carrier.</p> <p>It is important to include all water carrier records to avoid underestimating carrier mass</p>	<ul style="list-style-type: none"> - More than one record whose <code>Purpose</code> is "carrier" (or related) - at least one <code>bgCAS</code> is '7732-18-5' - total of water records is at least 50% - <code>PercentHFJob</code> - total % of disclosure is $95\% > x > 105\%$
s3	No carrier records labeled; but clear water record with typical percentage	<ul style="list-style-type: none"> - <code>bgCAS</code> is '7732-18-5' - at least 40% <code>PercentHFJob</code> - <code>IngredientName</code> contains phrase "including mix water" - total % of disclosure is $95\% > x > 105\%$
s4	Like s3, but CAS number missing; still obvious water record	<ul style="list-style-type: none"> - <code>CASNumber</code> is empty - at least 60% <code>PercentHFJob</code> - <code>IngredientName</code> contains phrase "including mix water" - total % of disclosure is $95\% > x > 105\%$
s5	Like s1 but no carrier records are labeled;	<ul style="list-style-type: none"> - <code>bgCAS</code> is '7732-18-5' - at least 50% <code>PercentHFJob</code> - total % of disclosure is $95\% > x > 105\%$
s6	<code>CASNumber</code> missing but clear carrier label	<ul style="list-style-type: none"> - <code>bgCAS</code> is ambiguousID - single record with a carrier <code>Purpose</code> - <code>IngredientName</code> is either "carrier" (or related) or has "water" in it - <code>TradeName</code> has "water" in it - $50\% < \%HFJob < 100\%$ - total % of disclosure is $95\% > x > 105\%$
s7	<p>Like s1, but for "salted" water</p> <p>Note that even though the record is labeled with the salt CAS number, the predominant mass is water</p>	<ul style="list-style-type: none"> - Only one record whose <code>Purpose</code> is "carrier" (or related) - <code>bgCAS</code> is either '7747-40-7' (kcl) or '7647-14-5' (nacl) - at least 50% <code>PercentHFJob</code> - total % of disclosure is $95\% > x > 105\%$
s8	Common pattern in the older disclosures (incl. SkyTruth archive)	<ul style="list-style-type: none"> - <code>bgCAS</code> is ambiguousID or 7732-18-5 - <code>IngredientName</code> is MISSING

Set name	description	Criteria to be detected
s9	Common pattern in the older disclosures (incl. SkyTruth archive)	<ul style="list-style-type: none"> - Purpose is "unrecorded purpose" - TradeName has either "water" or "brine" - can be one or two records in each disclosure - $50\% < \text{sum of PercentHFJob of these records} < 100\%$ - total % of disclosure is $95\% > x > 105\%$
s10	<p>A pattern seen in later disclosures: the carrier is only reported in the top part of the systems approach section under the "Listed Below" CASNumber . The actual PercentHFJob value isn't even reported in the PDF version, but is in the bulk download.</p>	<ul style="list-style-type: none"> - bgCAS is ambiguousID or 7732-18-5 - IngredientName is MISSING - Purpose is one of the standard carrier words or phrases - TradeName has either "water" or "brine" - can be one or two records in each disclosure - $50\% < \text{sum of PercentHFJob of these records} < 100\%$ - total % of disclosure is $95\% > x > 105\%$

Disclosures with detected problems for determination of water carrier ID

code	description
0	Disclosure has no valid chemical records.
1	TotalBaseWaterVolume is empty or 0 gallons.
2	None of the chemical records have non-zero PercentHFJob .
3	The sum of PercentHFJob values for valid CAS records is larger than limit (105%)
4	The sum of PercentHFJob values for all records excluding SystemApproach is larger than limit (105%)
5	PercentHFJob of all "proppant" records is greater than 50% (not used after v16)
6	The sum of PercentHFJob values for all records is less than 90% - a partial disclosure
7	PercentHFJob of Nitrogen or Carbon Dioxide records is greater than 50% (so carrier will be smaller) (not used after v16)
8	PercentHFJob of Chlorine dioxide records is 100% (it is typically an additive to the water; not a replacement) (added 3/2023, after v16).

code	description
9	PercentHFJob of Nonwater carrier record too large (>50%) (added 3/2023, after v16).