Quality Assurance Project Plan (QAPP) for

Analysis of Data Extracted from FracFocus

A. PROJECT MANAGEMENT

This section addresses project management, including project background and purpose, roles and responsibilities, and key research questions and objectives.

A1. TITLE AND APPROVAL SHEET

QA Category: 1

Date Original QAPP Submitted: August 2, 2012
Number of Pages: 12
Revision No.: 0

Signatures indicate approval of this QAPP and commitment to follow the applicable procedures noted:

/s/	8/2/2012
Susan Burden, Hydraulic Fracturing (HF) Data Analysis Technical Research Lead	Date
/s/	8/2/2012
Jeanne Briskin, HF Study Coordinator	Date
/s/	8/8/2012
Stephen Watkins, Quality Assurance Manager, Office of Science Policy	Date
/s/	8/13/2012
Mimi Dannel, Deputy Director, Office of Science Policy	Date

A2. Table of Contents

A3. DISTRIBUTION

This QAPP will be distributed to the staff members of the U.S. EPA as listed in Table 1.

TABLE 1. QUALITY ASSURANCE PROJECT PLAN DISTRIBUTION LIST

Name	Title	Contact Information
Jeanne Briskin	HF Study Plan Coordinator	(202) 564-4568 briskin.jeanne@epa.gov
Stephen Watkins	Office of Science Policy Quality Assurance Manager	(202) 564-3744 watkins.stephen@epa.gov
Susan Burden	HF Data Analysis Technical Research Lead, Chemical Mixing Writing Lead	(202) 564-6308 burden.susan@epa.gov
Ross Geredien	Data Analyst	(202) 566-1466 geredien.ross@epa.gov
Stephen Kraemer	Water Acquisition Writing Lead	(202) 564-0307 kraemer.stephen@epa.gov

A4. PROJECT ORGANIZATION

Project organization for the FracFocus data analysis is depicted below. Ross Geredien will be responsible for the secondary data collection, analysis, and presentation, and will thus be responsible for ensuring that the quality of work meets the requirements of the EPA's *Study of the Potential Impacts of Hydraulic Fracturing on Drinking Water Resources*. He will work under the direction of the HF Data Analysis Technical Research Lead, Susan Burden, and will keep the Quality Assurance (QA) Manager, Stephen Watkins, advised of any quality problems that arise in this project. The QA Manager will be responsible for maintaining QA activities and the official, approved QAPP throughout the course of the project.

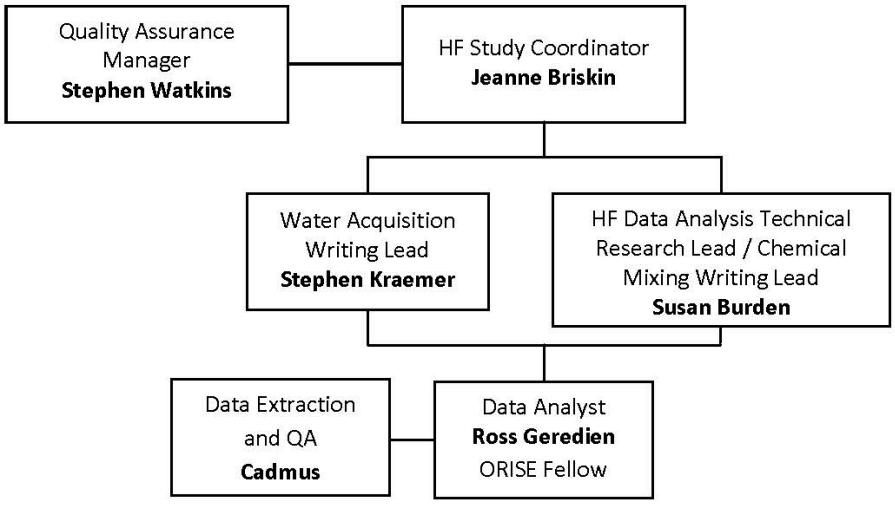

FIGURE 1. ORGANIZATION CHART FOR THE FRACFOCUS DATA ANALYSIS PROJECT

The FracFocus data analysis project will provide information relevant to two stages of the HF water cycle: Water Acquisition and Chemical Mixing.[1] Therefore, the Water Acquisition and Chemical Mixing Writing Leads—Stephen Kraemer and Susan Burden, respectively—will be involved in designing analyses (and reviewing products) related to data compiled from FracFocus.

A5. Problem Definition and Background

A5.1. Background

Hydraulic fracturing is a technique used to increase production of one of the nation's key energy resources, natural gas. Hydraulic fracturing increases the permeability of a geologic formation by pumping a pressurized fluid into the formation and creating fractures in the rock that allow gas to be extracted. These fluids typically contain a mixture of water, chemical additives, and proppants.

In response to the growing use of HF in the United States, Congress requested the EPA to study the relationship of HF and drinking water in Fiscal Year 2010. EPA responded to Congress' request by producing the *Draft Plan to Study the Potential Impacts of Hydraulic Fracturing on Drinking Water Resources* in February 2011. The draft study plan was reviewed and commented on by EPA's Science Advisory Board, and the final *Plan to Study the Potential Impacts of Hydraulic Fracturing on Drinking Water Resources* was completed in November 2011.

At the time the draft study plan was being reviewed, the Groundwater Protection Council and the Interstate Oil and Gas Compact Commission developed a new national registry for chemicals used in HF, called FracFocus. This registry is an online repository where oil and gas well operators can upload information regarding the chemical compositions of HF fluids used in specific oil and gas production wells. The registry also contains spatial information and information on well depth and water used. As of early 2012, this registry contained information on over 12,000 HF fluid disclosures (GWPC 2012).

This project will assess the quality and utility of the data in FracFocus to determine whether it can be analyzed in support of the current study. If data are of sufficient quality, they will be summarized and analyzed.

A5.2. Statement of Key Questions and Objectives

The information found in the FracFocus registry may help to inform the EPA's answers to the following research questions, originally posed in the study plan:

- How much water is used in the hydraulic fracturing operations? (Water Acquisition)
- What are the identities and volumes of chemicals used in HF fluids, and how might this composition vary across the country? (Chemical Mixing)

The project will address the following key questions with respect to data available through FracFocus:

Assessment of FracFocus Data.

- FracFocus data summary: how many well disclosures are reported in FracFocus?
- What is the geographic distribution of well disclosures in FracFocus?
- Other summary statistics include well depths and type (oil/gas)

[1] Described in EPA's *Plan to Study the Potential Impacts of Hydraulic Fracturing on Drinking Water Resources* (EPA/600/R-11/122).

Water Acquisition.

- What is the total volume of water used in HF operations reported in FracFocus nationwide, by state, and by county?
- What are the different sources of water identified in FracFocus, and is it possible to determine the relative proportions by volume of different water types used in HF operations?

Chemical Mixing.

- What are the identities of chemicals used in HF fluids reported in FracFocus?
- Which chemicals are the most commonly reported chemicals in FracFocus (by occurrence and mass/volume)?
- What is the geographic distribution of the most commonly reported chemicals in FracFocus?

A6. PROJECT/TASK DESCRIPTION

This project will assess the quality of and analyze data from FracFocus, the national HF chemical disclosure registry, managed by the Ground Water Protection Council and Interstate Oil and Gas Compact Commission. FracFocus uses a publicly available website (www.fracfocus.org) to provide information to the public about the chemicals used in HF fluids at specific well locations. Data will be extracted from PDFs found on the website and then transformed and aggregated into Excel spreadsheets and an Access database for analysis.

The analyses for this project will be performed during August-October 2012. Subsequent report summaries will be developed during the fall of 2012. Tables and/or maps will be produced to illustrate the use of different chemicals and water volumes nationwide and/or by state, county, or other geographic area. Key tables and maps and other results will be incorporated into a report that will be supplied to the HF Study Coordinator and the Water Acquisition and Chemical Mixing Writing Leads.

A6.1. DATA ACQUISITION

The Cadmus Group, Inc., (Cadmus) extracted data from all PDFs available on the FracFocus website as of February 27, 2012.[2] These files represent all data reported by operators in the FracFocus database as of the extraction date. The data elements extracted from the PDFs are listed below. Cadmus has performed all data QA and preparation for this project. For more information on the quality assurance procedures taken for these step, see the *Supplemental Programmatic Quality Assurance Project Plan For Work Assignment 4-58: National Hydraulic Fracturing Study Evaluation of Existing Production Well File Contents Issued Under Contract No. EP-C-08-015*.

Well information:

- Fracture date
- State
- County
- API number
- Operator name
- Well name and number
- Longitude

[2] The QAPP for this process can be found at: http://www.epa.gov/hfstudy/WellFile_Supplement_QAPP%20supplement_508_km.pdf.

- Latitude
- Long/lat projection
- Production type
- True vertical depth
- Total water volume

Fluid composition information:

- Trade name
- Supplier
- Purpose
- Ingredients
- Chemical Abstract Service (CAS) number
- Maximum ingredient concentration in additive
- Maximum ingredient concentration in HF fluid
- Comments

A6.2. DATA ANALYSIS

Analyses that will be performed fall into three main categories: summary statistics, water acquisition, and chemical mixing. A summary of each category is given below.

Summary Statistics. Summary statistics will describe both the overall data extracted from FracFocus and the quality of that data. Descriptions of the dataset will include the:

- Total number of records entered into FracFocus as of the extraction date; and
- Number of unique records summarized by state, production type, operator name and fracture date

Unique records will be determined by the reported American Petroleum Institute (API) number. Records with duplicate API numbers will be flagged for further analysis to determine the validity of these records. Additional QA issues will be summarized as follows:

- Identification of the issue
- Number of records affected
- Action taken

Water Acquisition. An initial analysis will be performed on the total water volumes reported for individual wells. Total water volumes will be analyzed by state and county, well depth, production type (i.e. oil or gas), and operator.

In some cases, operators report different sources of water through the Trade Name, Purpose and Ingredients entries in FracFocus. Where possible, EPA will summarize different sources of water by volume, state, production type and operator. However, concentrations of water by source type are generally found in the "Maximum ingredient concentration in HF fluid" field, which is reported as a percentage by mass, not percentage by total water volume.

In some situations there will be enough information in FracFocus to calculate water volumes by type, whether fresh water (e.g., surface water) or non-fresh water (e.g. recycled/produced, saline, seawater or brine). Given the FracFocus reported total water volume (V_{H2O}^{total}) (US gallons) and assuming that volumes are additive[3],

$$V_{H2O}^{total} \cong \sum_{i=1}^{n} V_{H2O}^{i} \quad (1)$$

We can use the FracFocus reported maximum water concentration in the HF fluid (percent by mass for each water type, x_{H2O}^{i}). Assuming an average density for each water type (ρ_{H2O}^{i}) (lbs/US gallon), the volume of each water type is expressed as:

$$V_{H2O}^{i} = \frac{x_{H2O}^{i}}{\rho_{H2O}^{i}} m_{total} \quad (2)$$

With *n* equations and *n* unknowns represented by equations (1) and (2), where *n* is the number of water types, the unknown total mass of the HF fluid (m_{total}) (lbs) can be calculated:

$$m_{total} = \frac{V_{H2O}^{total}}{\sum_{i=1}^{n} \frac{x_{H2O}^{i}}{\rho_{H2O}^{i}}} \quad (3)$$

The volume of each water type (V_{H2O}^{i}) is back-calculated using equation (2).

This calculation can only be made in the situation where the density of the fluid is known or reported. For example, in the situation where the FracFocus ingredient carrier/base fluid is clearly labeled fresh (surface) water, a water density of 0.99821 g/mL at 20° C may be assumed[4]. Another example might be where the density for the carrier/base fluid is reported in the FracFocus comment field, such as the case of WPX Energy and their Garfield County, CO, wells and the use of recycled produced water.

Additional analyses related to water acquisition will include:

- Correlation analysis between total water usage and True Vertical Depth (TVD); and
- Summary statistics of the number of wells that use water from each water type or source.

Chemical Mixing. The EPA will use the data available in FracFocus to identify and summarize information regarding chemicals used in HF fluids. The dataset used for this analysis will be summarized in the following way:

- Total number of unique records included in the analysis
  - Number of records per state, including summaries from states required to use FracFocus as of Feb. 27, 2012;
  - Number of unique chemicals per well and how these data are distributed across all wells;
  - Number of records per production type; and
  - Number of operators represented.

[3] EPA recognizes that volume is not a conserved quantity and estimates that the error introduced by assuming that volumes are additive is, in this case, negligible when compared to expected volume and density reporting errors.

[4] Lide, David R., Ed. (2008). CRC Handbook of Chemistry and Physics, 89th ed. CRC Press: Boca Raton, FL, 2011; p. 6-4.

The EPA will also note when, why and how many records were not included in the chemical mixing analysis.

The EPA will prepare tables with the following information:

- Total number and lists of all unique trade names, suppliers and purpose for chemicals or products;
- Total number and a list of all unique CAS numbers;
- Total number of records with "proprietary, "trade secret," or "confidential business information" in the ingredient or CAS field (with corresponding trade name, supplier and purpose); and
- Number of unique proprietary or confidential chemicals, chemical ingredient records, and in some cases the total number of proprietary chemical ingredient records per well.

The list of unique CAS numbers will also include the following information:

- Number of times the CAS number appears in the dataset;
- Number of unique wells associated with the CAS number; and
- The range, average, median, mode and standard deviation of the values reported in the "maximum ingredient concentration in HF fluid" element.

The EPA will use the list of unique CAS numbers to identify chemicals commonly used in HF fluids. The EPA will create maps of chemical usage based on select subsets of these chemicals.

A7. Quality Objectives and Criteria

The EPA does not make any claims on the quality or accuracy of the data or information found in FracFocus. This QAPP aims to ensure that the existing data used for this evaluation are of sufficient quality necessary to achieve proper data analysis. This section addresses the quality criteria used to assess the adequacy of secondary data used in this project, as well as the uncertainty in the results derived from the use of these data sources. The elements listed below will be considered when assessing the quality of FracFocus data reviewed under this task.

A7.1. Data Accuracy

EPA will limit its sources of data in this analysis to information available online from the FracFocus website. Data from FracFocus will be converted into an Access database, and all accuracy of input data will be ensured by procedures outlined in the *Supplemental Programmatic Quality Assurance Project Plan For Work Assignment 4-58: National Hydraulic Fracturing Study Evaluation of Existing Production Well File Contents Issued Under Contract No. EP-C-08-015*. To check the accuracy of the data analyses (simple non-parametric frequency counts national and by state), results of individual queries will be cross-checked with results that are sorted and filtered in the database. Some results will also be spot-checked for accuracy.

A7.2. Data Precision

Sources of data will be limited to information available online from the FracFocus website. The records in the FracFocus system are individual records and not repeated measurements. Thus, evaluations of precision of the original dataset do not apply. Cadmus will, however, note any issues that emerge that could indicate a problem with precision as analyses proceed. Quality checks will be conducted as described under accuracy.

A7.3. Issues Associated With FracFocus Data

FracFocus data will be identified for different QA issues using separate data columns. Each analysis performed will include an assessment of which data will need to be excluded based on QA issues. When QA issues indicate

whether certain records are unsuitable for specific analyses, these records will be scrutinized for other field values to determine whether any systematic bias would potentially be introduced by their exclusion.

A7.4. Completeness

All data downloaded from the FracFocus website will be used for analysis except where QA issues deem certain records inappropriate (duplicate records, data inconsistencies, etc.). In addition, data that cannot be completely and properly extracted and transformed will not be used.

A7.5. Representativeness

All FracFocus well record files will be downloaded, so the data used for the analyses are expected to duplicate the information available in FracFocus (i.e., data will be fully represented). It is beyond the scope of this project to evaluate the quality or representativeness on a national scale of the data submitted to FracFocus by oil and gas operators. The data will not be assumed to be a complete or statistically proportional representation of all wells hydraulically fractured during the time period of January 1, 2011, until February 27, 2012, either nationwide or within a defined geographic area. However, because FracFocus contains several thousands of well disclosures distributed throughout the United States, the EPA believes that the data in FracFocus will be indicative of HF activities during the time period covered.

A8. Special Training Requirements/Certification

No special training is anticipated at the time of this writing.

A9. Documentation and Records

An electronic lab notebook will be kept by Mr. Geredien for all draft queries, test queries, final queries as well as draft and final results, notes and explanations. Dr. Kraemer will use an electronic lab notebook to track lab notes and queries using the Excel spreadsheet tool developed by Cadmus Group for the Water Acquisition section. Dr. Kraemer keeps his electronic records on a private ORD drive (L:\Priv\MEERT\WorkGroups\Energy and Environment\HydraulicFracturing) that is backed up daily and archived monthly.

B. Data Generation and Acquisition

This section addresses data acquisition and management activities, including the following elements identified by EPA:

- Element B5: Quality Control
- Element B9: Non-direct Measurements
- Element B10: Data Management

B5. Quality Control

All data used in this project will be evaluated using the criteria listed in Section A7 and will be examined to ensure that they meet these criteria at each stage of this project (data acquisition, manipulation, and analysis).

B9. Non-direct Measurements

All data are considered existing data and so are non-direct, or secondary data. All data will be evaluated for appropriateness for each analysis.

B9.1. LIMITATIONS

Since quantities of chemicals and water types are given in percentage by mass, these units are not consistent with total volume of water used. As a result, water volumes by type will have to be back-calculated as shown above. However, volumes are not additive in solution formation. It is therefore possible that a fluid mixture will influence the volume of solution. We are comfortable that the error introduced by assuming the volumes to be additive is minor compared to the errors in measuring or reporting the fluid volumes into Fracfocus.

In addition, there is some uncertainty about how correctly True Vertical Depth (TVD) has been reported in FracFocus, given the presence of some outliers. It is possible that TVD was confused with Total Depth (TD), which includes horizontal drilling distances. However, in the vast majority of cases where there are no clear outliers, EPA will assume that TVD is accurately reported.

Any other noted inconsistencies and QA issues will be identified and flagged (i.e., marked in separate QA data columns in the database). Data will then be excluded from analyses when QA inconsistencies make them unsuitable for specific analyses.

B10. DATA MANAGEMENT AND HARDWARE/SOFTWARE CONFIGURATION

B10.1. DATA MANAGEMENT

Cadmus will provide data to EPA in Microsoft Access and Excel formats, which will be used for analytical queries. Data will be sent via FTP from Cadmus to EPA and stored on desktop hard drives by the individual project analysts as well as on network drives. Working files will be stored on individual hard drives while in use and backed up on the project network directory:

O:\316-258_501a2_Project_Files-Electronic\HF-Draft_Draft_Report_2012 (316-258_501a(2))

Query results and reports will be stored in separate file folders. Cadmus will develop a data dictionary defining the different fields in the Access database. Any GIS vector or raster data downloaded will be stored on hard drive. GIS coordinates reported in FracFocus will be used to convert well data into shapefiles. Accuracy of the GIS coordinates will be flagged whenever they are inconsistent with the county or state reported in FracFocus, or when they are inconsistent with the county and state codes in the API numbers.

B10.2. HARDWARE/SOFTWARE CONFIGURATION

Microsoft Excel and Microsoft Access will be used to store and analyze the data from FracFocus. ESRI's ArcGIS 9.3.1 and ArcGIS 10.0/10.1 software will be used to produce any maps or to do any spatial analyses. No special hardware configuration will be required by this project.

C. Assessment and Oversight

C1. Assessment and Response Actions

The draft Access database will be reviewed upon submission to the EPA by Cadmus, and any outstanding QA issues will be identified and brought to Cadmus' attention. Cadmus will make necessary corrections to the database and resubmit it to EPA before the final analyses will begin.

C1.1. Data Quality Audit

The EPA and Cadmus will assess the accuracy of all the data in FracFocus to the extent possible. Cadmus will perform the majority of this work in consultation with the EPA as outlined in the *Supplemental Programmatic Quality Assurance Project Plan For Work Assignment 4-58: National Hydraulic Fracturing Study Evaluation of Existing Production Well File Contents Issued Under Contract No. EP-C-08-01.*Issues such as inconsistent or invalid data values will be identified and corrected whenever possible. In instances where no obvious errors or data quality issues exist, however, the EPA and Cadmus make no claim to the validity or accuracy of those data, which are reported by HF operators in the FracFocus database. All potentially valid data values will be accepted as reported.

C1.2. Technical Systems Audit

A technical systems audit will occur toward the beginning of the FracFocus analysis to ensure that the appropriate analytical methods and models are employed. The technical systems audit will verify that the procedures described in the QAPP are being followed properly and will be conducted by OSP's Quality Assurance Manager. Necessary deviations from the QAPP procedures will be addressed in revisions to the QAPP.

C2. Reports to Management

Mr. Geredien will supply a draft report on the products and findings of this analysis to the HF Study Coordinator and the HF Data Analysis Technical Research Lead for comment and will then incorporate their comments into a final report. Mr. Geredien will keep the study team involved through weekly technical progress updates in which he will describe any problems encountered and solicit feedback as necessary to ensure quality of the finished product.

D. Data Validation and Usability

This section addresses the quality of the completed final report to see if this product will conform to the objectives outlined in this QAPP, especially given this project's use of existing data sets.

D1. Data Review, Verification, and Validation

All of the extracted data will be transformed into Comma Separated Values (CSV) file format, where a series of scripts can be run to perform a variety of QA checks. Once QA'd, the data will be further transformed into Microsoft™ Excel and Access formats, where they can be analyzed systematically using a series of standard queries. Data that do not meet certain QA check standards will be flagged accordingly so that data can be filtered appropriately during the analyses.

All query results during the analyses will be reviewed for accuracy and validity. This is particularly important during the data transformation phases following data extraction, but will continue during the review of the final deliverables from Cadmus.

D2. Data Verification and Validation Methods

Query results will be evaluated for verification using database cross-check techniques such as sorting data values alpha-numerically and filtering data according to specific data values and parameters. Data will also be checked for validity using the Cadmus report of summary statistics of QA issues, prepared under the *Supplemental Programmatic Quality Assurance Project Plan For Work Assignment 4-58: National Hydraulic Fracturing Study Evaluation of Existing Production Well File Contents Issued Under Contract No. EP-C-08-015*. The number of records in the Access database, for example, with QA issues will be compared to the number reported by Cadmus. Inconsistencies will be analyzed for the causes of any discrepancies.

D3. Reconciliation with User Requirements

As described in C1.1, all data will be assessed for accuracy, validity, consistency, and generally the elements in Element A7. All data QA issues will be quantified, generally, as a percentage of records contained in or successfully extracted from FracFocus. Certain data that do not meet the QA criteria will be excluded from the analyses, and in each case the assumptions and reasoning behind these decisions will be presented clearly and in a transparent manner. Only data that are determined appropriate for specific analyses will be used for those analyses based on which criteria are met. Summaries of data quality issues will be presented in tabular and written text formats.

Revision History

Revision Number	Date Approved	Revision
0		New Document