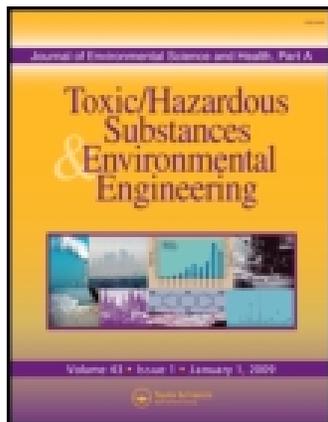


This article was downloaded by: [69.239.249.174]

On: 03 March 2015, At: 16:10

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Environmental Science and Health, Part A: Toxic/Hazardous Substances and Environmental Engineering

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/lesa20>

Data inconsistencies from states with unconventional oil and gas activity

Samantha Malone^{ab}, Matthew Kelso^a, Ted Auch^{ac}, Karen Edelstein^a, Kyle Ferrar^{ab} & Kirk Jalbert^{ad}

^a FracTracker Alliance, Pittsburgh, Pennsylvania, USA

^b Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

^c Department of Biological, Geological, and Environmental Sciences, Cleveland State University, Cleveland, Ohio, USA

^d Department of Science and Technology Studies, Rensselaer Polytechnic Institute, Troy, New York, USA

Published online: 03 Mar 2015.



[Click for updates](#)

To cite this article: Samantha Malone, Matthew Kelso, Ted Auch, Karen Edelstein, Kyle Ferrar & Kirk Jalbert (2015) Data inconsistencies from states with unconventional oil and gas activity, *Journal of Environmental Science and Health, Part A: Toxic/Hazardous Substances and Environmental Engineering*, 50:5, 501-510

To link to this article: <http://dx.doi.org/10.1080/10934529.2015.992678>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Data inconsistencies from states with unconventional oil and gas activity

SAMANTHA MALONE^{1,2}, MATTHEW KELSO¹, TED AUCH^{1,3}, KAREN EDELSTEIN¹, KYLE FERRAR^{1,2}
and KIRK JALBERT^{1,4}

¹*FracTracker Alliance, Pittsburgh, Pennsylvania, USA*

²*Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA*

³*Department of Biological, Geological, and Environmental Sciences, Cleveland State University, Cleveland, Ohio, USA*

⁴*Department of Science and Technology Studies, Rensselaer Polytechnic Institute, Troy, New York, USA*

The quality and availability of unconventional oil and gas (O&G) data in the United States have never been compared methodically state-to-state. By conducting such an assessment, this study seeks to better understand private and publicly sourced data variability and to identify data availability gaps. We developed an exploratory data-grading tool - Data Accessibility and Usability Index (DAUI) - to guide the review of O&G data quality. Between July and October 2013, we requested, collected, and assessed 5 categories of unconventional O&G data (wells drilled, violations, production, waste, and Class II disposal wells) from 10 states with active drilling activity. We based our assessment on eight data quality parameters (accessibility, usability, point location, completeness, metadata, agency responsiveness, accuracy, and cost). Using the DAUI, two authors graded the 10 states and then averaged their scores. The average score received across all states, data categories, and parameters was 67.1 out of 100, largely insufficient for proper data transparency. By state, Pennsylvania received the highest average ($\bar{x} = 93.5$) and ranked first in all but one data category. The lowest scoring state was Texas ($\bar{x} = 44$) largely due to its policy of charging for certain data. This article discusses the various reasons for scores received, as well as methodological limitations of the assessment metrics. We argue that the significant variability of unconventional O&G data—and its availability to the public—is a barrier to regulatory and industry transparency. The lack of transparency also impacts public education and broader participation in industry governance. This study supports the need to develop a set of data best management practices (BMPs) for state regulatory agencies and the O&G industry, and suggests potential BMPs for this purpose.

Keywords: Data transparency, environmental governance, oil and gas extraction, public participation, unconventional drilling.

Introduction

Energy extraction has been a topic of contentious debate in recent years in both the United States and internationally.^[1–3] The process of extracting natural gas and other hydrocarbons from difficult-to-access formations in the ground, herein referred to as “unconventional drilling,” embodies this controversy due to its potential costs and benefits,^[4] which are not discussed in detail here. Advances in hydraulic fracturing and horizontal drilling technologies made formerly difficult hydrocarbon extraction more feasible, but the generation, cataloging, and sharing of quality data about the industry have not kept up with such

innovations or the related permitting process. Although some unconventional drilling data are available that provide basic information (e.g., the location of wells drilled), the availability and quality of these datasets are not consistent across states.

For the purposes of this article, we define data transparency as the standardization and consistent publication of various types of unconventional oil and gas (O&G) data by state regulatory agencies. Some state regulatory agencies have improved O&G data transparency in recent years by releasing information and datasets on their websites. Ohio, for example, switched from reporting production data annually to quarterly as of January 2014. Unfortunately, many states have caught up with neither the pace of O&G activities nor the public’s interest in accessing data.

O&G data transparency is made complex for many reasons. Increasingly scarce resources within environmental regulatory agencies limit personnel capacity to manage data and build appropriate technical infrastructure such as

Address correspondence to Samantha Malone, FracTracker Alliance, 4600 Penn Ave. Floor 1, Pittsburgh, PA 15224, USA; E-mail: Malone@FracTracker.org

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lesa.

databases, websites, and other public information portals. Data accessibility also suffers due to the sheer number of entities responsible for collecting and sharing O&G data, as well as the rate at which unconventional drilling wells are permitted. Historical studies show that diffusion of responsibilities across agencies causes data to be aggregated using different standards, with specific parameters to suit the needs of particular regulatory agencies and/or administrative codes, in addition to the laws on which such efforts are designed to reflect.^[5] Although such data may eventually become public, many data portals and variable standards, rather than a state-wide or nationally standardized approach, poses a persistent problem that results in poor data-sharing arrangements among agencies, ultimately creating barriers to public accessibility.^[6-8]

Inadequacies in data sharing and quality can significantly affect the degree of public trust in the regulation of the O&G industry. Arguments for data transparency go well beyond the obvious need for better public understanding, however. Data transparency promotes broader participation in the democratic governance of the environmental “commons.”^[9-11] Access to state regulatory data also empowers local communities in mitigating potential industry impacts and allows non-governmental organizations (NGOs) to fill gaps in economic, social, health, and environmental knowledge created by lagging and/or incomplete regulatory oversight.

Furthermore, utilization of data provided by respected regulatory agencies allows academic and non-profit researchers to improve the accuracy of their findings and add credibility to their efforts when participating in scientific and political debates about the impacts of the O&G industry.^[12] In addition to opening doors to broader public participation, data transparency also benefits local government and other state governments that are seeking information to inform their own regulatory policies.^[13]

Some states have approached the issue of data transparency by requiring the O&G industry to share data through projects like FracFocus.org, a site designed to disclose chemicals used in unconventional drilling.^[14] The logic of this method is derived from success stories of similar industry databases, such as the Toxic Release Inventory, and projects like “Scorecard” that build on those inventories to make data more readable for non-expert audiences.^[15,16] Some argue, however, that the reporting of O&G data to third parties has diverted attention away from the responsibilities of regulatory agencies tasked with managing environmental resources and public health.^[17] Reporting data to non-regulatory entities also offers few assurances that data are accurate and complete.^[18] In light of this issue, we believe that in the case of O&G data, the states themselves need to undertake the tasks necessary to ensure data transparency and make possible cross-state data comparisons.

In recognition of the data gaps in the O&G industry, several NGOs and websites, including the FracTracker Alliance (www.FracTracker.org), were launched to better

collect and share unconventional drilling data with the public.^[19] Over the last three years FracTracker has developed a series of parameters to quantitatively measure data transparency and data quality within the field. FracTracker’s position outside of industry and government allows it to better understand and approximate the experience of the general public when obtaining and working with O&G data. Data tools and meaningful pathways between raw data and refined information promote increased public understanding and engagement.

This study documents and quantifies the variability we see in data provided by state regulatory agencies across the United States where active unconventional drilling activity is occurring as a way to estimate the barriers faced by the general public when trying to access this information. This article also provides initial recommendations regarding the data gaps identified during this process so as to improve O&G data transparency. Closing these access gaps will enable the public to participate more meaningfully in important scientific and political debates that determine how O&G extraction will be regulated in their communities.

Methods

Between July and October 2013, five different categories of data were sought from 10 U.S. states. These data were reviewed using an internally developed Data Accessibility and Usability Index (DAUI).

Data categories reviewed

The five data categories assessed in this article are wells drilled, violations, production (of profitable hydrocarbons), waste (non-profitable returns), and U.S. Environmental Protection Agency (EPA)-defined Class II disposal wells. Disposal wells do not include those that are used for enhanced oil recovery, hydrocarbon storage, annular disposal, or salt-solution mining injection. Permit data were not reviewed because many states release these combined with drilled wells in a single dataset. Information about the nature and location of compressor stations from state regulatory agencies is difficult to obtain. (Only in January 2014 did it become possible to obtain dated locations of compressor stations in the United States from the U.S. Energy Information Administration.)

Although disclosure of data about this aspect of the industry is critically important, it was not the primary focus of this study, which aimed to test the DAUI in situations where data availability was relatively robust. Data on current and proposed pipelines were also not reviewed, as their locations are difficult to obtain in the United States due to the contention that releasing on-shore pipeline location information could be a national security risk. (Pipeline incidents data are available from the U.S. Department

of Transportation's Pipeline and Hazardous Materials Safety Administration. Offshore pipelines are available from the Bureau of Ocean Energy Management. Of note, the government of British Columbia does release pipeline right-of-way data to the public.) Sand mining datasets were also not assessed in this study, but could feasibly be included in future reviews.

Data quality criteria

Eight data qualities were assessed using the DAUI for each of the categories of data: accessibility, usability, point location, completeness, metadata, agency responsiveness, cost, accuracy (on varying scales of 1–20, 1–15, or 1–10, depending on the importance of the metric). Inclusion of latitude and longitude information in the data is a key parameter; without this information, data cannot be easily mapped or statistically analyzed for research purposes. From the viewpoint of effective public information, raw—or tabular—data have far less power to inform or educate compared to data that can be mapped or visualized.^[5] (Table 1 describes the discrete scoring categories). Using this system, each of the five categories of data may earn up to 100 points. Each state can score a maximum of 500 points.

States included in the assessment

States were chosen for this study based on three factors: 1) degree of unconventional O&G activity along a gradient from somewhat active to very active, 2) geographic location to capture distribution across the United States, and 3) whether violations data are available. Experience in working with unconventional O&G data to-date shows that datasets about violations are rarely released. While not all of the states considered in the DAUI maintain datasets on violations, all states with known violation data as of the inception of the project were included in the index. These criteria resulted in our choice of the following states for this review: Arkansas (AR), California (CA), Colorado (CO), Nebraska (NE), New York (NY), North Dakota (ND), Ohio (OH), Oklahoma (OK), Pennsylvania (PA), and Texas (TX). Figure I presents selected states and U.S. shale plays.

Obtaining the data

Free data from the selected states were reviewed between July and October 2013 in order to provide a snapshot of data over that time period. ND data were later purchased in order to determine how much of ND's original score had been affected by its policy of charging for violation information. Scores presented in the results section for ND represent the original results from our review, not the assessment of this purchased data. Datasets were collected

by several of the study's authors, but all data were reviewed and graded by the same two authors in order to maintain scoring consistency. The final results presented comprise the average scores derived by these two authors.

The first step in the data gathering process involved visiting each regulatory agency's website. If datasets were provided free of charge and online, data review could begin. When data could not be assessed in this fashion, the agencies were contacted by email and phone to request the data. All unanswered requests were repeated a second time. If data could not be obtained after this point, a state received a score of zero in most, if not all, of the data quality categories. (Even if no data were available for review, a state could receive up to 10 points in the responsiveness category if the agency responded to our request in some way.)

Results

Scores received in each data category are shown alphabetically by state in Table 2. The combined average across all 10 states and categories is 67.1. By state, PA received the highest overall mean score ($\bar{x} = 93.5$) and the highest in all data categories except drilled wells. The lowest scoring state was TX ($\bar{x} = 44$), which is largely the result of its charging for certain data. ND was also docked for access to violations data. By data category, drilled wells ($\bar{x} = 87.8$), Class II disposal wells ($\bar{x} = 85.3$), and production data ($\bar{x} = 81.3$) received much higher averages state-to-state because of their wide availability than did violations and waste, at 33.8 and 47.5, respectively.

Discussion

The variability of O&G data regarding wells drilled, permitting and waste produced variability is substantial. This appears to be due to variation in corporate and regulatory structures from state-to-state. The structure of data provided to the public; however, need not be so varied. Some degree of consistency is required to discern even the most basic trends, as well as for making meaningful comparisons between states and between various shale plays.

Scores by data category

Below, two states' scores from each data category are reviewed, both to highlight variability in the data and to demonstrate how the DAUI captures such differences.

For the category of drilled wells with a mean of 87.8 points, 10 out of 10 states provided accessible data. OK scored a perfect 100 points and ranked first among the states in this data category. Compared to OK, AR scored 75 points and ranked 9th overall. In the DAUI scoring system, AR lost points because of limitations in

Table 1. Data Accessibility and Usability Index, a grading index with (scores) for each data category based on eight criteria.

<i>Accessibility (20)</i>	<i>Usability (10)</i>	<i>Point Location (15)</i>	<i>Completeness (15)</i>	<i>Metadata (10)</i>	<i>Responsiveness (10)</i>	<i>Accuracy (10)</i>	<i>Cost (10)</i>
Data not in digital form (0)	Agency only releases pre-packaged data (ex: .pdf, .jpg, or can only use their map interface; no download options) (0)	Point location not provided (0)	Data missing several important columns, such as date or operator, or only a portion of the records are available (0)	No metadata available (0)	Agency is unhelpful or obstructive (e.g., requiring a FOIA request for non-critical data) (0)	Major accuracy issues call reliability of dataset into question (0)	Data costs money (0)
Data is in digital form, but not available to public (5)		Point location provided, but in alternate system, such as State Plane or PLSS (5)	Data missing one important column, such as date or operator (5)	Metadata present, but incomplete (5)	Emails or calls are sometimes not returned (5)	Minor accuracy issues noted (5)	
Downloadable digital data available upon request (15)			Some minor data missing, but basic analysis is still possible (10)				
Downloadable digital data published on web (20)	Agency releases data that can be analyzed with free software (ex: .shp, .csv) (10)	Point location provided in decimal degree (15)	Data very complete (15)	Metadata very complete (10)	Agency is cooperative with data request (separate issue from whether or not they have the data in the first place) (10)	No known accuracy issues (10)*	All data free (10)

*Where data exists, the default score is 10. Points are deducted if we notice accuracy problems when the data are mapped. Abbreviations: FOIA, Freedom of Information Act; PLSS, Public Land Survey System.

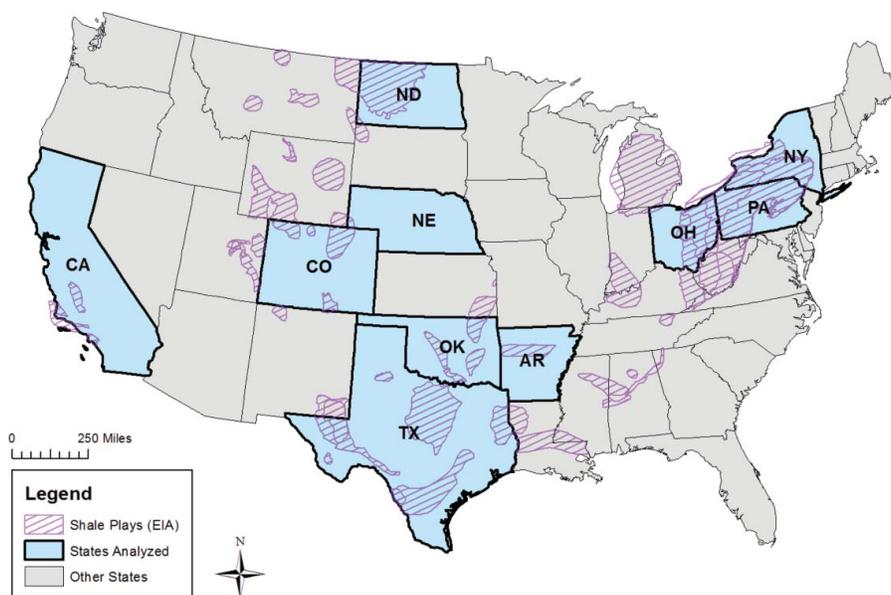


Fig. 1. States analyzed with the Data Accessibility and Usability Index in this study overlaid on continental U.S. shale plays (shapefile source U.S. Energy Information Administration).

its online data management system; AR allowed access to only the first 1,000 records. Full data were provided upon request in Microsoft Access form. Microsoft Access requires that users have more sophisticated computer skills compared to those needed for working with a simple spreadsheet. Therefore, some points were deducted because data were provided in a challenging file format.

Datasets for disposal wells were also commonly available, obtained from 10 out of 10 states. Among the states in this study, the mean score for disposal wells data was 85.3. CO scored 73 points and ranked 9th, due to usability issues with data released in HTML tables. CO's disposal well dataset was also missing point locations and metadata. NE received 80 points and tied for a rank of 6th due

an issue similar to the one described above for AR – requiring Microsoft Access to work with the data – and also because metadata were missing.

All 10 states provided production data, with an average score of 81.3 points. CA scored 95 points and ranked 3rd in this data category, while TX scored only 60 points and ranked last (10th) among the states. CA was docked for usability because it, too, released data in Microsoft Access format. Although TX does publish production data for wells, the information is solely available in online tables that cannot be downloaded and viewable only 15 records at a time. For this reason, TX scored poorly in both accessibility and usability. Additionally, the free TX data lacked location coordinates.

Table 2. Summary of scores on the FracTracker data index by state and data category.

Data Categories	State Scores										State \bar{x}	State σ
	AR	CA	CO	ND*	NE	NY	OH	OK	PA	TX		
Wells Drilled	75	85	90	95	90	92.5	90	100	95	65	87.8	10.4
Violations	75	10	65	15	10	55	10	0	77.5	20	33.8	30.6
Production	75	95	75	62.5	85	80	80	100	100	60	81.3	14.1
Waste	70	10	67.5	62.5	85	0	75	0	95	10	47.5	37.8
Class II Disposal	75	100	72.5	95	80	85	80	100	100	65	85.3	12.8
Category \bar{x}	74	60	74	66	70	62.5	67	60	93.5	44	67.1	12.7
Category σ	2.2	46	9.8	32.8	33.7	37.7	32.3	54.8	9.3	26.8	24.8	

*The format of ND's datasets and website navigation initially prevented us from finding the state's unconventional drilling waste and production data. When these data were located, however, we reevaluated ND on the DAUI in December 2013. ND's up-to-date scores for waste and production data are presented here.

Abbreviations: AR, Arkansas; CA, California; CO, Colorado; ND, North Dakota; NE, Nebraska; NY, New York; OH, Ohio; OK, Oklahoma; PA, Pennsylvania; TX, Texas; \bar{x} , Sample Mean; σ , Standard Deviation from the Mean.

Waste data could be acquired only from 6 of the 10 states that we analyzed in the DAUI, making the mean score 47.5 for all datasets and 75.8 for procured datasets. Immediately prior to submission of this article, the NY Department of Conservation responded to a Freedom of Information Law (FOIL) request for data on the disposition of drilling waste generated in NY. The data are not available in spreadsheet format. We were directed to an FTP site where the following information is catalogued: permit documentation (for general operation, not specific waste products), and annual reports in PDF form for incinerators (MWC), landfills, construction and demolition, household hazardous waste, vehicle dismantling facilities, and recyclables. In order to find O&G data, it is necessary to go through each folder manually. Documents are not searchable by keyword as they are scanned as images. Due to the long time it took for NY to respond (>3 months) and the poor format of the data when a response was received, the NY score was not amended on the DAUI. NY scored 0 points, thus tying for 9th in their management of waste data. OH scored 75 out of 100 and ranked 3rd. OH was docked for missing point location information regarding the nature and origin of the waste being reported. Metadata were insufficient, and some data seemed to be missing from brine totals.

Violations data were available from 4 out of 10 states at the time of our assessment, averaging 33.8 points per dataset, and 68.1 for those datasets that we could obtain. (OH has since released a violations dataset.) PA ranked 1st with a score of 78 for violations data, with point reductions due to insufficient metadata and missing location information. For the public to determine the location of an issued violation such as a spill, in PA, the user must link the violation report to a separate dataset that contains location data (e.g., drilled wells), using a unique API number. There were also some concerns regarding accuracy of the data. For example, in the past, several nearby wells were commonly assigned a single violation ID number.

In recent years, the ID coding methodology seems to have been improved by the PA Department of Environmental Protection (PA DEP). Nevertheless, there is a legacy of this older record-keeping practice manifesting itself in confusion over the number of violations for unconventional wells in the state. PA DEP appears to count violations by the number of unique violation ID numbers issued rather than actual incidents themselves. CO scored 65 points and ranked 3rd in quality and availability of their violations data. Problems with CO data included inconsistencies between dataset entries, spills, and Notices of Alleged Violations, as well as missing point locations and metadata. Data were accessible online in HTML table format only, with results limited to 2,500 records at a time. Compared to PA's system, where complete datasets can be downloaded from the website with one click, downloading data from CO in aggregate, and subsequently

analyzing the datasets, is very difficult. CO's Notices of Alleged Violations lacked not only location data but often API numbers, as well.

Reflections on methodology

State scores on the DAUI should be considered only in terms of the data and states that were reviewed because not all states with active unconventional drilling activity in the United States were reviewed, and only in terms of a sample of potential data categories. Among the states assessed, PA stands out due to its consistently high scores. Despite its faults as discussed previously, data provided by the PA DEP ranked 1st in every data category except for drilled wells, for which it tied for 2nd place. In addition, PA datasets did not vary much from one data category to the other, as indicated by a fairly low category standard deviation ($\sigma = 9.3$).

Several reasons may account for the relatively high scores of PA on the DAUI. First, FracTracker originated in PA, and reviewers were most conversant with the characteristics and sources of PA data. Familiarity with and expectations of O&G data likely shaped how the DAUI was structured and the data categories selected. A similar group in CO, for example, might have prioritized other data categories, therefore yielding different results. CO publishes data about impoundments, which are not readily available from the PA DEP website (although they do make some impoundment data available on the Pennsylvania Geospatial Data Clearing House).

Choosing to include impoundments as a data category in the analysis would have helped to weigh the results more highly for CO than for PA. Similarly, if providing a shapefile of the lateral portion of horizontal wells had been one of our criteria, PA would have scored poorly for not providing that data category, whereas ND and CO would have earned relatively higher marks. Under such circumstances, scores for specific states could be significantly different than those presented here. The DAUI's ability to highlight variances in the data, however, would be retained. Future versions of the DAUI and other data assessments would benefit from including additional data categories.

It might also be posited that the existing familiarity between FracTracker and the PA DEP gave PA a scoring advantage. This review included not only a search for data online, but also involved asking regulatory agencies for this information. Existing knowledge of PA's website and data availability could have helped to bolster the rating we applied in the "accessibility" category of the DAUI. However, it did not significantly aid PA's score in any of the other fields. Another scoring advantage for PA could have been that the PA DEP is more familiar with data requests from FracTracker, whereas other states might not have received as many requests in recent years from like groups for data that is appropriately geo-referenced.

It must be emphasized that these scores represent data acquisition in a singular point assessment at a moment in time. In general, the trend in PA over the last two years toward better data accessibility and quality points to the possibility that, when citizen groups and NGOs apply repeated pressure to regulators, state agency reporting practices improve. At the time of the launch of FracTracker in 2010, for example, comprehensive PA violation data were unavailable. Data on drilled wells were released only once per month. These datasets were incomplete, lacked location data, and were generally fraught with errors. In contrast, a current DAUI snapshot of Tennessee (TN) would have likely scored more poorly today than it would have several years ago. Personal correspondence from the TN Department of Environment and Conservation with FracTracker has verified our observation that O&G data for the state are altogether unavailable on the TN website. TN previously offered such data, however, and plans to reinstate this service in the future.

In situations where the exact data being collected are not already available online, there is a social aspect to data collection. In other words, the manner in which data requests are delivered could have significant impacts on the DAUI score beyond what is captured in the “responsiveness” category. For example, a regulatory worker whose function it is to help with the dissemination of data might respond in differing ways depending on whether he or she perceived particular requests as polite, demanding, or aggressive. This tendency could, in theory, cause over-worked regulatory agencies to selectively prioritize other requests if they felt that requests were rude, ill informed, or unimportant.

Of the 50 datasets analyzed in the DAUI, 3 returned scores of 0 because data were not available and our inquiries went unanswered. One of these is the aforementioned waste dataset from NY, while the other two are the waste and violation datasets from OK. Two different officials at the OK Corporation Commission Oil and Gas Division failed to reply to email requests that were sent to them. Inadequate staff resources and differing our approaches to data requests could have impacted these, and other, scores on the DAUI. OK should review its public inquiry process to ensure requests like ours are answered in the future.

The lowest scoring state on the DAUI was TX, with an average of 44 points across the data categories. These low scores were strongly impacted TX’s policy of charging for certain data, such as location information. Originally, the DAUI was designed to demerit states 10 points if they charged for data. Because paying for the data was not part of the initial study design, however, the two states that charged for data, ND for violations data and TX for most of its data, scored poorly in some data categories on the DAUI. Simply put, data quality could not be assessed if not procured. The overall impact of charging for data created a cascade effect—reducing scores by more than the intended 10-point demerit—even though some points

could still be obtained in the Responsiveness and Meta-data categories. This effect should be addressed in future versions of the DAUI.

A related shortcoming of the matrix was that there was no mechanism in the DAUI to account for the relative price of data. For example, a basic one-year subscription to ND’s data costs a relatively modest \$50 and provides the user with several types of valuable data. In contrast, one would have to purchase several datasets from TX to do a similar analysis, with estimated costs exceeding \$1,000. For the general public and small non-profit organizations, the financial burden of acquiring data from TX is much higher than for data from ND.

Although data purchases were not part of the original study design, ND’s violations data were purchased *post hoc* in order to determine how much the purchase would affect ND’s score in that category. Access to these new data improved ND’s violation score from 15 points to 55, considerably enhancing the state’s mean and standard deviation from 66.0 and 32.8 to 74.0 and 19.4, respectively. TX’s DAUI scores would also likely benefit from purchasing the data.

Best data management practices

Next, we list recommended data best management practices to improve O&G data transparency.

A universal data management system. Due to the nature of O&G data being processed and distributed at the state level of government, attempts to make assessments at the national scale are inherently difficult due to inconsistency. A recent Associated Press analysis of violations datasets^[20] concluded that it is very difficult to conduct cross-state comparisons because of the wide variability in the data collection process and in what determines a ‘violation’—if violations data are even available. The implementation of a national O&G data management system is one way to reduce these inconsistencies.

Such a system could be structured similarly to the U.S. EPA Toxic Release Inventory (TRI), which requires companies in every state to report emissions of certain toxic chemicals to TRI databases and then publishes the data online in a publicly accessible and downloadable format. FracFocus.org is a website originally designed to improve the transparency of unconventional drilling data. The site design has some inherent barriers to transparency, however; data are not available for download in aggregate formats, and reporting requirements are uneven across states. Additionally, the TRI is government-run, while the Groundwater Protection Council and Interstate Oil and Gas Compact Commission run FracFocus.^[21] These characteristics would severely limit the extent of compulsory disclosures by the states; hence the current FracFocus model is insufficient.

More stringent data and metadata. Even with a system like TRI in place, O&G data accuracy, transparency, and access issues could still arise because the data are often self-reported by industry. This issue makes clear the need for reporting requirements to carry threat of prosecution for deliberate false reporting or failing to report. Separately, due to resource limitations, the state regulatory agency might not be able to verify all of the incoming data, increasing the possibility for errors. For example, when a biannual waste report was released in PA in August 2012, the total waste amounts were greatly over reported by several operators who had reported amounts in gallons rather than the required unit of barrels.^[22] In addition to improving data quality, it is important to maintain basic metadata standards.

At a minimum, these standards should include descriptions of codes used in the various datasets that are maintained by a regulatory agency. Other information about the data is also important. For example, personal correspondence with the CA Department of Conservation Division of Oil, Gas, and Geothermal Resources (DOGGR) has verified that the regulatory agency was in the process of retroactively applying a new data column to old wells, in order to indicate whether they had been hydraulically fractured or not. As the process was incomplete, not knowing this crucial metadata could potentially lead those analyzing the data to reach faulty conclusions.

States can improve the quality and accessibility of existing O&G datasets by merely including latitude and longitude in each dataset. Consistently applied and comprehensive location information would help ensure data quality, especially over time, when wells and facilities will need to be located in order for reclamation to take place. Had there been proper record keeping and location information for historical O&G wells in PA, for example, PA would have been able to locate thousands of abandoned and orphaned traditional wells.^[23]

Augmenting cost of data management. Limited funding to support data management by state agencies remains a major hurdle to greater accessibility. Regulatory agencies must look to increase their funding in order to be able to ground-truth incoming industry data. Charging indiscriminately for access to the data reduces overall transparency, especially for smaller NGOs, academic institutions, and individuals that may not be able to afford the high cost. These funding models amount to asking citizens to “pay twice” for services they should rightly expect public institutions would offer.^[24] This barrier could be reduced if regulatory agencies charged companies a fee for data collection and cataloguing as a part of the permitting process, provided a tiered payment system for citizens or small groups that request data, or incorporated management costs into severance/*ad valorem* taxation.

Accessible technologies and standards

Many of the issues described here demonstrate the need for states to recognize that they are serving multiple audiences – not just government agencies and industry. Complicated and obtuse website navigation posed a serious impediment to obtaining data for this study. For example, while one author was looking for NE disposal well data and was not able to find it through the direct route recommended by the state’s web page, another author was later able to locate it while looking for data about wells that have been drilled. In light of this experience in regularly working with unconventional drilling data it is reasonable to conclude that the process of accessing these data would be only more difficult for a member of the general public. To improve user access to data on these sites, states should adopt search and navigation recommendations set forth by web-based user-interface research.^[25] Data and metadata, such as geolocation, should also be available for download in multiple formats, including CSV or Microsoft Excel spreadsheets, rather than sharing data through tedious HTML tables or intricate relational databases (such as Microsoft Access).

Conclusion

This review of unconventional O&G regulatory data highlights the need for more consistent data management practices across affected states. Answers to seemingly simple questions such as “How many unconventional oil and gas wells are there in the United States?” and “How often are these wells issued violations?” remain elusive due to inconsistent standards and data reporting among the states.

The DAUI was designed as a tool to identify the variability in data issued by differing regulatory agencies, as well as to provide a means to compare resulting datasets. By applying the index to 5 distinct data categories from 10 states, the former goal was met. Cross-state comparisons, while possible, must account for the variability outlined in the “Reflections of Methodology” section. Future analyses would benefit from some minor adjustments in the scoring mechanism, most notably with respect to data that is provided for a fee. This result is reasonably expected; much in the way that the index serves to highlight weaknesses in the data, applying real-world data to the DAUI provides an opportunity to further optimize the DAUI.

The best practices for data management outlined here, such as more stringent data standards, technologies designed to be accessible to multiple audiences, and alternative methods for augmenting the costs of data management, can help to improve data transparency issues revealed in the application of the DAUI. Furthermore, while there are certainly state-specific characteristics that show promise in how data is shared, data pertaining to

particular issues or geographies should not be compartmentalized. Waste disposal^[26] and hydraulic fracturing sand networks,^[27] for example, can often encompass industrial sites operating across state lines. The simplest solution to this problem is to adopt a universal, mandatory, and federal government system that supplements state-managed data systems.

Whether it be through practices suggested here, or by some other means, ultimately it is critical to find solutions to data accessibility and usability gaps that have been identified. Unconventional O&G extraction incites many criticisms, as well as support. Proponents argue that developing a greater capacity for O&G extraction across the US creates jobs in local economies and offers a viable way out of dependency on global energy markets. Many others, however, voice concerns about the possible impacts of unchecked development on transportation infrastructures, ecological habitats, public health, air quality, and water quality. Precedents set by states with existing O&G development may have lasting impacts on how we expect public institutions to respond to demands for transparency in the future. Without access to consistent and comprehensive O&G data, these debates will only continue to be muddled by a lack of adequate public knowledge about industry activities.

Acknowledgments

The research was originally presented at the 2013 Facing the Challenges: Research on Shale Gas Extraction Symposium hosted by the Center for Environmental Research and Education at Duquesne University.

Funding

This work was supported by the FracTracker Alliance and our various funders (see www.fractracker.org/about-us/#Funders).

References

- [1] Koch, W. EPA proposes strict emission limits on new power plants. USA Today, 2013. Available at www.usatoday.com/story/news/nation/2013/09/19/epa-limits-emissions-coal-power-plants-carbon-capture-technology/2838391 (accessed Oct 2013).
- [2] Williams, S. Total Looking At UK Shale Gas Opportunities – CEO. The Wall Street Journal. 2013. Available at <http://online.wsj.com/article/BT-CO-20131001-706027.html> (accessed Oct 2013).
- [3] Kenarov, D. Poland stumbles as shale gas industry fails to take off. McClatchy: Washington, DC, 2013.
- [4] Garavaglia, B.H. Hydraulic fracturing: Sources of law and information. Mich. Bar J. Librar. Legal R. **2013**, 92(9), 58–59.
- [5] Haklay, M. Public access to environmental information: past, present and future. Comput. Environ. Urban Syst. **2003**, 27(2), 163–180.
- [6] Harvey, F.; Tulloch, D. Local – government data sharing: Evaluating the foundations of spatial data infrastructures. Inter. J. Geograph. Info. Sci. **2006**, 20(7), 743–768.
- [7] Medeiros, F.; Kurtoglu, B.; Ozkan, E.; Kazemi, H. Analysis of production data from hydraulically fractured horizontal wells in shale reservoirs. SPE Reserv. Eval. Eng. **2010**, 13(3), 559–569.
- [8] Lynch, M. Shale gas production and high decline rates. Forbes Magazine: New York, 2013.
- [9] Fortun, K. Environmental information systems as appropriate technology. Design Iss. **2004**, 442(7105), 862–864.
- [10] Elwood, S. Grassroots groups as stakeholders in spatial data infrastructures: challenges and opportunities for local data development and sharing. Inter. J. Geogr. Info Sci. **2008**, 22(1), 71–90.
- [11] Burger, J.; Ostrom, E.; Norgaard, R.B.; Policansky, D.; Goldstein, B.D. (Eds.). Protecting the Commons: A Framework for Resource Management in the Americas; Island Press: Washington, DC, 2001.
- [12] Morello-Frosch, R.; Pastor, M.; Sadd, J.; Porras, C.; Prichard, M. Citizens, science, and data judo: leveraging community-based participatory research to build a regional collaborative for environmental justice in southern California; In: *Methods for Conducting Community-Based Participatory Research in Public Health*; Israel, B.A.; Eng, E.; Schulz, A.J.; Parker, E.A., Eds.; Jossey-Bass: San Francisco, CA, 2005; 371–391.
- [13] Adair, S.K.; Pearson, B.R.; Monast, J.; Vengosh, A.; Jackson, R. B. Considering shale gas extraction in North Carolina: Lessons from other states. Duke Envtl. Law Pol. F. **2012**, 22, 257–385.
- [14] FracFocus.org. GWPC and IOGCC launch www.fracfocus.org. 2011. Available at <http://groundwork.iogcc.org/topics-index/hydraulic-fracturing/iogcc-in-action/gwpc-and-iogcc-launch-wwwfracfocusorg> (accessed Oct 2013).
- [15] Fung, A.; O'Rourke, D. Reinventing environmental regulation from the grassroots up: Explaining and expanding the success of the toxics release inventory. Environ. Manage. **2000**, 25(2), 115–127.
- [16] Delborne, J.; Galusky, W. Toxic transformations: Constructing audiences for environmental justice. In: *Technoscience and Environmental Justice: Expert Cultures in a Grassroots Movement*; Ottinger, G.; Cohen, B.R., Eds.; The MIT Press: Cambridge, MA, 2011.
- [17] Shonkoff, S. Public Health Dimensions of Horizontal Hydraulic Fracturing: Knowledge, Obstacles, Tactics, and Opportunities. Berkeley, CA: University of California, 2012.
- [18] Konschnik, K.; Holden, M.; Shasteen, A. Legal Fractures in Chemical Disclosure Laws: Why the Voluntary Chemical Disclosure Registry FracFocus Fails as a Regulatory Compliance Tool. Harvard Law School, Environmental Law Program: Cambridge, MA, 2013.
- [19] Michanowicz, D.; Malone, S.; Kelso, M.; Ferrar, K.; Christen, C.; Volz, C.D. A Participatory Geographic Information System (PGIS) utilizing the GeoWeb 2.0: Filling the gaps of the Marcellus Shale natural gas industry. J. Syst. Cybern. Inform. **2012**, 10(3), 45–53.
- [20] Begos, K. Oil and gas drilling pollutes well water, states confirm. Associated Press, 2014. Available at http://usnews.nbcnews.com/_news/2014/01/05/22190011-oil-and-gas-drilling-pollutes-well-water-states-confirm (accessed Jan 2014).
- [21] FracFocus. About Us. 2014. Available at <http://fracfocus.org/welcome> (accessed Jan 2014).
- [22] NGI Staff Reports. Unconventional Gas Production Hits New Record in Pennsylvania, NGI's Shale Daily, 2013. Available at <http://www.naturalgasintel.com/articles/5157-unconventional-gas-production-hits-new-record-in-pennsylvania> (accessed Jan 2014).
- [23] PA DEP. Abandoned and Orphaned Well Program. Available at http://www.portal.state.pa.us/portal/server.pt/community/abandoned_orphan_well_program/20292 (accessed Jan 2014).

- [24] Tulloch, D.; Harvey, F. When data sharing becomes institutionalized: Best practices in local government geographic information relationships. *Urban and Regional Information Systems Association*, **2007**, *19*(2), 51–59.
- [25] Nielsen, J. Top 10 Information Architecture (IA) Mistakes. Nielsen Norman Group, 2009. Available at <http://www.nngroup.com/articles/top-10-ia-mistakes/> (accessed Jan 2014).
- [26] FracTracker Alliance. PA Unconventional Production and Waste (Jan-June 2013). Online map from FracTracker.org of data provided by the Pennsylvania Department of Environmental Protection, 2013. Available at <http://maps.fractracker.org/latest/?appid=1239a1b243a04a66b1ca813c9a2ac9be> (accessed Jan 2014).
- [27] FracTracker Alliance. US Frac Sands Locations and Silica Geology. Online map from FracTracker.org of data provided by several different organizations and states, 2013. Available at <http://maps.fractracker.org/latest/?webmap=2f382d5fcd748deba89e6104b59551d> (accessed Jan 2014).