Original research article

Is reporting “significant damage” transparent? Assessing fire and explosion risk at oil and gas operations in the United States

Benjamin D. Blair*, Lisa M. McKenzie, William B. Allshouse, John L. Adgate

Colorado School of Public Health, University of Colorado Denver, United States

ARTICLE INFO

Keywords:
Fire
Explosions
Oil
Gas
Risk
Mandatory reporting

ABSTRACT

Oil and gas (O & G) operations are periodically accompanied by fires and explosions. The frequency of these incidents is not well known, particularly at modern sites that increasingly use advanced techniques, such as horizontal drilling and hydraulic fracturing. The objective of this work is to determine the rate of fires and explosions at O & G sites in Colorado and Utah and apply this information to evaluate the proximity of these incidents to residences. Between 2006 and 2015, a total of 116 fires and explosions in Colorado (0.03% of active wells) and 67 fires or explosions in Utah (0.07% of active wells) were reported at O & G operations. The higher percentage of fires or explosions per number of active wells in Utah compared to Colorado (Rate Ratio = 2.49, p < 0.01) is likely influenced by the mandatory self-reporting requirements in Utah and the more lenient self-reporting in Colorado. The average number of residences within 1609 m of the reported incident was 31 (median = 3, SD = 131) in the Denver Julesburg Basin and 4 (median = 0, SD = 10) in the Piceance Basin of Colorado. To our knowledge, this is the first systematic analysis of fires and explosions at O & G sites and offers insight into the rate and reporting of these events.

1. Introduction

Technological advances, such as horizontal drilling and high volume hydraulic fracturing, fostered the extraction of oil and gas (O & G) from previously inaccessible resources, such as shale and tight sands, which resulted in a boom of United States O & G development at the beginning of the 21st century. This rapid expansion in development raised public concern about the health and safety consequences for populations living near O & G operations [1,2]. Fires and explosions at O & G operation sites have caused employee injuries, economic loss, environmental damage, and loss of human life [2–7], but the reported frequency of these events and their proximity to residences has not been well documented.

Numerous flammable and explosive materials and potential ignition sources are present at O & G operation sites. The hazardous materials can include chemicals and industrial supplies such as the petroleum products (i.e., oil, condensate, natural gas), diesel fuel, shaped charges, primer cord, detonators, and various chemicals [7,8]. Ignition can be caused by sources such as electrical shocks, mechanical friction causing sparks, lightning, and open flames such as ground fires, pilot lights, welding, cutting torches, lighters, or cigarettes [3,4,9,10]. Issues surrounding maintenance, operation, and equipment errors may result in conditions that could cause a fire or explosion at O & G operation sites [4]. A review of 77 fire and explosion incidents reported to the U.S. Occupational Safety and Health Administration found that O & G operations resulted in 42 deaths and 87 injuries between 2010 and 2014 [6]. Haley et al. [2] reports that second degree burn blisters can occur at the 350 foot Colorado outdoor recreational setback distance, the distance from a well to an outdoor recreational area such as a playground, after 22 s of exposure to an O & G operation site fire at a typical gas well. In addition to the risk caused by the fire or explosion, these incidents can also result in the release of hazardous materials [2,11]. To illustrate the potential risk from fires at O & G operation sites, a broken hydraulic line sprayed fluid on machinery that resulted in a fire at a well pad in Clarington, Ohio that caused an estimated 30 explosions with shrapnel and required a 1609 m (1 mile) evacuation of approximately 25 residences [7].

A wide range of organizations and government agencies in the United States collect data about fires and explosions at O & G operations sites, however, a comprehensive database regarding these incidents does not exist within the United States [12]. Therefore, the goals of this analysis are to use state level data to evaluate the occurrence and location of fires and explosions at O & G operation sites in the States of Colorado and Utah, provide a state level comparative analysis of these
incidents, and evaluate the number of residences potentially impacted by these incidents in two O & G basins in Colorado. Colorado and Utah were selected for comparison in this article due to the notable O & G operations in these states as Colorado is the 7th and Utah is the 11th largest state producer of crude oil in the U.S. [13], the availability of fire and explosion reports, and the different levels of stringency regarding state level regulatory elements surrounding O & G operations [14]. In this work, we advance the methodical approaches to assessing fires and explosions by creating a database of incidents through content analysis of incidents as reported by O & G operators in Colorado and Utah and consider the proximity of residences to these incidents. We also evaluate the reporting requirements and regulations between these two states and explore the impact of mandatory reporting on the rates of reported incidents.

1.1. Utah and Colorado O & G fire and explosion reporting and setback law

A notable difference is present in the reporting requirements for fires and other undesirable events at O & G operation sites in Utah and Colorado: Utah requires self-reporting of all fires and explosions whereas Colorado requires self-reporting of incidents that have caused harm to a member of the general public or fires and explosions that caused significant damage [15,16]. In Utah, rule R649-3-32 of the Utah Oil and Gas Conservation Act states that “The division shall be notified of all fires, leaks, breaks, spills, blowouts, and other undesirable events occurring at any oil or gas drilling, producing, or transportation facility, or at any injection or disposal facility” [16]. In Colorado, rule 602.b of the Colorado Oil and Gas Conservation Commission (COGCC) requires accidents to be reported to the COGCC Director within 24 h and an official written accident report submitted within 10 days for “any accident or natural event involving a fire, explosion, detonation, or release of pressure that results in: 1) injury to a member of the general public which requires Medical Treatment and/or 2) significant damage to equipment or the well site” [15].

Regulatory exclusion setback zones (“setbacks”) are used to define the distance between an O & G operation site and an adjacent property, structure, area or feature of interest, or building. These rules are generally designed to protect the general population from potential harm and nuisances from these operation sites. Colorado changed the setback requirement from 45.7 m (150 ft) in rural areas and 106.7 m (350 ft) in urban areas to 152.4 m (500 ft) for all residences in 2013 [15,17]. In Utah, no setback distances are set at the state level, however, setbacks can be regulated at the county level [18]. Previous research has questioned whether current setbacks are adequate to protect human health from fires or explosions at O & G sites [2]. In addition, it is not uncommon for residences to be built within the exclusion setback zone, such as in Colorado [19]. To date, an understanding of the proximity of fires and explosions at O & G operation sites to homes is lacking.

1.2. Targeted transparency and potential underreporting of O & G fires and explosions

“Targeted transparency” policies can compel information disclosure as an alternative to other policy tools and is “often used to introduce new scientific evidence of public risks into market choices” ([12], p. 1410). These policies differ from right-to-know policies as this also requires the collection and release of information that is standardized, factual, and comparable to inform public decision making [21]. This form of mandatory reporting can be viewed as more light-handed than command-and-control regulation and uses information to encourage informed decision making regarding the potential risk to the environment or public health from an activity [20–22]. Without targeted transparency policies, firms may under disclose information to the public, such as accidents, thereby creating information asymmetry [20]. Targeted transparency is often politically controversial and can be costly, time-consuming, become out-of-date or erode over time, and can prioritize some risks over others [20].

Public concerns surrounding O & G operations have led to targeted transparency efforts by policymakers and regulators in some states, such as the disclosure of chemicals within hydraulic fracturing fluids [22]. Evaluating the potential underreporting of fires and explosions due to vague mandatory reporting requirements is critical for evaluating risk rates for these incidents at O & G operation sites. In addition, while industrial accident reporting rates are higher with mandatory reporting compared to voluntary reporting [24], in Colorado the role of interpreting whether an incident results in “significant damage” is left to the discretion of the O & G Operator. The “pragmatic interpretation” of whether accidents or incidents are reportable may also influence the decision to report an incident [25], particularly when mandatory reporting or targeted transparency requirements are unclear. To the best of our knowledge, no empirical evaluation of the influence of differing reporting requirements on O & G operation site fires and explosions has been completed and published in the peer-reviewed literature.

2. Methods

2.1. Data collection, compilation, and analysis

In Colorado, accident reports (COGCC form 22) are used to describe significant damage from fires or from incidents that involve injuries to the general public. Using the search terms “form 22” and “accident report”, documents were downloaded from a publicly available COGCC database. Accident reports were evaluated for incident dates ranging from 2006 to 2015. Briefly, the accident reports were identified through a targeted SQL query keyword search on the title text of well documents from the COGCC database. Once identified, the accident reports; in either PDF or TIF file format; were downloaded using a custom Ruby script as the documents were not easily downloaded from the COGCC database. The complete accident report download methods are shown in Appendix A.

The Colorado accident reports were manually coded for the presence of a fire or explosion, the date, a short description of the incident, the American Petroleum Institute (API) number (a unique identifier given to each well in the U.S.), the etiology or mitigating factors surrounding the incident, the latitude and longitude of the well or facility, and the presence or absence of a fire department extinguishing or managing the incident (e.g., the decision to let the fire extinguish itself while managing the scene). All of the accident reports were coded twice by the same coder for the presence of fires or explosions and any identified discrepancies were reconciled. Each accident report was coded manually by a single incident coder as the factors included did not require complex coder interpretation.

Release reports are used in Colorado to describe incidents that involve spills at O & G operation sites and these reports include information on fires and explosions. Release report documents were collected from the COGCC publicly available data with reporting dates ranging from January 1 2006 to December 31 2015. The release reports were obtained from COGCC’s COGIS Database [23] by using the Inspection/Incident Inquiry Search function. Depending on the date of the incident, report text was extracted from either scraped HTML or from downloaded PDF file text. Both extraction methods employed automated Ruby scripting. Report text was stored in a Postgres database and report data elements were parsed by using SQL scripting. See Appendix A for the complete methods regarding the release reports. A total of 7706 reports were downloaded and reviewed. Due to the large number of spill reports, two readers were used to evaluate the release reports and the same coding criteria as the accident reports were used.

In Utah, the annual incident event reports and number of active wells were received in a Microsoft Excel document from the State of Utah Oil and Gas Program within the Division of Oil, Gas and Mining.
The Utah incident reports included fires, spills, fatalities, injuries, blowouts, and other undesirable events. The reports included a dichotomous variable regarding the presence or absence of a fire along with descriptions of the events. Two thousand incident reports were reviewed between 2006 and 2015. Two reports did not indicate the presence of a fire, but had reported explosion or fires in the narratives and these incidents were included in the analysis. As with the Colorado accident reports, each incident report was coded manually by the same coder and any discrepancies in coding results were reconciled.

Duplicate entries in the database were found by evaluating the dates of the incidents and the corresponding API well number or location. If an incident had an identical or similar date and the same well API or location, the written descriptions were reviewed to identify and remove duplicate incidents. A total of 13 duplicate entries were found in Colorado and 10 were found in Utah and these duplicate entries were removed from the analysis.

The number of active wells as classified by the COGCC were found in Colorado using COGCC reported values [26]. In Utah, the number of active wells were found by using the number of producing oil and gas wells from the Utah well status document (data received via email communication with the Utah Oil and Gas Program). We use the number of active wells as a proxy for O & G operation activity, although incidents can occur at other sites such as compressor stations, evaporation pits, and underground injection wells.

The local regression in Fig. 1 and the rate ratio (RR) were calculated using R v3.2.2. The RR calculation compared the rate of the total number of fires or explosion over the period of analysis per average number of active wells over the period of analysis between Utah and Colorado.

2.2. Case study on the proximity of fires or explosions to residences in Colorado

Addresses were obtained for residential properties in the Denver Julesburg and Piceance basins in Colorado from DataQuick, a dataset used to collect residential addresses for the counties within these basins. The case study area for resident proximity consists of ten counties in two basins: (1) Adams, Arapahoe, Boulder, Broomfield, Larimer, Logan, Morgan, and Weld counties in Colorado’s Denver Julesburg basin; and (2) Garfield and Mesa counties in the Piceance basin. We selected these two basins because they are the most active areas of O & G operations in Colorado [19]. The addresses were geocoded and rooftop accurate latitude and longitude coordinates were found for 94% of the addresses using the Google Maps Application Programming Interface; those without this level of accuracy were removed from the analysis. Only residences built in the same year or previous years to a specific fire or explosion were used. Incidents with an API number or latitude and longitude coordinates listed on the accident report were included. Incidents that included the quarter-quarter section grid information and did not have an API or geocoordinates were omitted. Using latitude and longitude coordinates from the accident report or based on the well API, the number of residences within 1609 m (1 mile) of a reported fire or explosion at an O & G operation site were calculated with ArcMap v10.2.2. A 1609 m radius was used based on the evacuation distance of the reported fire at an O & G operation site in Clarington, Ohio [7], which was a well documented example of an incident at an O & G operation.

![Fig. 1. Rate of fires or explosions per well in Colorado and in Utah from 2006 to 2015. Solid line are local regression lines.](image-url)
3. Results

3.1. Summary of fires and explosions

From 2006 to 2015, 116 unique fires or explosions were identified at O & G sites in Colorado and 67 fires or explosions in Utah. A summary of the number of fires or explosions and active wells by year are displayed in Table 1.

Colorado has a lower rate of reported fires or explosions per active well than Utah, as displayed in Fig. 1. In Colorado, fires or explosions were reported annually at 0.03% of active wells. By comparison in Utah, fires or explosions were reported annually at 0.07% of active wells. Fig. 1 also includes a local regression line for each state, which is a nonparametric measure of fit of the rate over time. Based on the total number of fires and explosions and the average number of active wells from 2006 to 2015, Utah was significantly more likely to have reported fires and explosions at O & G operation sites than Colorado (RR = 2.49, p < 0.01).

Lightning strikes were reported as the cause or suspected cause of the fire or explosion for 14% (n = 16) of the reported incidents in Colorado and 9.0% (n = 6) in Utah. Welding, torch cutting, or the use of grinding equipment at O & G operation sites were reported as a cause for 7% (n = 8) of the fires or explosions in Colorado and 7.0% (n = 5) in Utah. These incidents involved cutting piping, lines, pit liners, or tanks causing residual liquids or vapors to ignite resulting in fire or explosions. Other causes were documented as operator error, equipment error, static electricity, vandalism, other, and unclear/unknown, as shown in Table 2. Reported causes were unclear, unknown, or under investigation at the time of the report submission for 42% of the reported fires or explosions in both Colorado and Utah.

Fire department response and assistance were often needed at fires and explosions that occurred at O & G operation sites. Of the reported fires or explosions, 40% (n = 46) in Colorado and 37% (n = 25) in Utah indicated that the local fire departments were needed to extinguish or manage the incident.

3.2. Case study of proximity of fires to residences in the Denver Julesburg Basin

To explore the proximity of explosions or fires to residences, we conducted a case study on the number of residences within a 1609 m radius from the fires and explosions at O & G sites in the Denver Julesburg and Piceance Basins, as summarized in Table 3. Eighty-one of the 116 reported incidents in Colorado had latitude and longitude coordinates or API numbers listed on the accident reports and were within the two basin study area. Of these incidents, the average number of residences within 1609 m of the reported fire or explosion was 31 (median = 3, SD = 131) in the Denver Julesburg Basin and was 4 (median = 0, SD = 10) in the Piceance Basin.

The shortest distance of a reported fire or explosion to a residence was 115 m, as shown in Fig. 2. This figure highlights the close proximity of a flash fire at an O & G operation site to residential areas in this instance. The distance of 115 m is less than the current Colorado regulatory setback of 152.4 m, but greater than the previous setback distance of 45.7 m. The maximum number of residences within a 1609 m radius was 819, which occurred in the more populated Denver Julesburg Basin. In contrast, the greatest number of residences within 1609 m of a reported fire or explosion in the Piceance Basins was 51. Across both basins in Colorado, 18 incidents in Colorado were reported with more than 10 residences within 1609 m of the fire or explosion.

4. Discussion

4.1. Study strengths and limitations

Examining the reported rate, cause, and proximity to homes of fires and explosions at O & G operation sites are critical components needed to inform the management and oversight of these incidents. In Colorado, fires or explosions were reported at 0.03% of active wells, while in Utah fires or explosions were reported at 0.07% of active wells. Utah reported fires or explosions at a rate of 2.49 times greater than Colorado during this time period and this result is likely a function of the mandatory reporting requirements in Utah compared to the less stringent reporting requirements in Colorado. It is clear from these results that the reporting requirements affect reported incidents, although the total impact that different reporting requirements has on reported rates is uncertain.

A content analysis of self-reported accidents is a valuable method to understand the fires and explosions at O & G operation sites at a state level. However, the self-reporting by operators in both states regarding these incidents is also a notable limitation. For example, non- or underreporting of incidents at these sites is possible, even in the case of an emergency response. Similarly, incidents at inactive sites, such as abandoned or plugged wells, may not be reported. Previous research has also found that dishonesty and underreporting is prevalent with self-reporting of industrial accidents [24,25]. In addition, the accident reports require a short written description of the incident and did not require any specifics such as cost of property damage, severity of incident, occurrence of explosions or extreme hazards, or damage to
surrounding property. As a result of limited instructions for the operators to complete the accident reports, the information describing the fire or explosion were generally brief. To illustrate the under-reporting of incidents in Colorado, a lightning strike led to a tank battery fire with multiple explosions propelling the tanks into the air, caused an estimated $1.5 million dollars in damage, evacuation of homes in the area, and took many hours to control by the fire department, as reported by the news media [27]. In contrast, the entirety of the accident description on the accident report stated “Due to a Lightening [sic] strike on Friday 4/17/15 there was a loss of the Tank Battery. An Investigation of the extent of loss is in progress. The Facility is shut in and there were no injuries.” This highlights how the incidents are required to be reported initially, but follow up reports, which may also include more information on the cause, may have been submitted but not through a form that met our search criteria. Other data sources also exist to evaluate these incidents, such as the National Fire Incident Reporting System, yet the publicly available version of this database does not include written narratives or clear classification categories specific to only oil and gas operations. Thus, it was not possible to improve upon the COGCC self-reporting to verify rates that could help validate the results. Furthermore, the count of active wells may not include all operations, such as refinery, compressor, pipeline, or wastewater injection facilities.

An additional limitation is that the exact location of the incidents were not known as the geolocations used were often based on the API number, which locates the wellhead. Also displayed in Fig. 2, the wellhead location may not be indicative of the equipment and material storage at these sites and the proximity of a fire or explosion may be closer to a residence than the regulatory setback. The geocoding of addresses may also not represent the proximity of the nearest edge or corner of the residence to the well site or the reported incident. In addition, 30% of the reported incidents in Colorado did not have clear locations or were outside the two basin study area and further evaluation of these incidents is needed for the proximity to residence calculations. These limitations imply uncertainty in the distance between the fire and residences and the incident could have been nearer or further from residences than reported.

4.2. Policy recommendations

A primary recommendation from this work is based on the need for consistent reporting requirements of fires and explosions at O & G operation sites. At this time, no comprehensive database exists [12] and the need for standardized reporting to better understand the safety risks from these hazards would assist in informing public decision-making. In the U.S., additional reporting requirements may best be considered by state or local jurisdiction due to the concentration of areas with notable O & G operations. Further information is also needed regarding the potential explosion hazard under worst case scenarios. We find that fires or explosions are reported at up to 0.11% of active wells in a year (the reported rate in Utah in 2014), thereby creating a potential risk to those that reside near O & G operation sites. Moreover, a recent trend in drilling operations is fewer single well and more multi-well pads, with an average of five wells located on a multi-well pad [28]. The increased size and complexity of operations at multi-well pads will likely continue in the future, particularly in urban and suburban areas, increasing the potential risk for those living and working in close proximity to these sites.

Complete information regarding industrial incidents is needed for risk assessments, to determine risk mitigation strategies, and to consider the acceptable levels of risk. Based on the incomplete information in many reports, we recommend that the minimal data required on a fire report should document the fuel source and estimated volume, cause, damage to other property, and injuries to workers and citizens. Other information that could inform risk assessments to be considered for inclusion in the reporting process include the duration,
intensity, fuel source, exact location of the incident, proximity to residences or other buildings, inventory of combustible or explosive chemicals and supplies on site at the time of the incident, an estimate of economic costs, and emergency management response. The data on fires and explosions may be improved through reports with required data fields and increased reporting requirements. Similar to the recommendations made to modify FracFocus, the database used to disclose hydraulic fracturing chemicals, the use of automatic field population with prompts to enter information could offer reduced error rates and more detailed and systematic reporting of these incidents [29]. In addition, consistent terminology should be developed and utilized for reporting key information such as the fire intensity, fire severity, and burn severity [30].

An additional policy recommendation is that states with notable O & G activity should consider the implementation of mandatory reporting of all fires and explosions at O & G operation sites. At present, Colorado does not have mandatory reporting of all incidents, while Utah requires mandatory reporting of fires or explosions at O & G operation sites. Based on these findings, it is our judgment that many incidents are likely not being reported to the COGCC in Colorado. Utah reported an incident rate that is on average 2.49 times greater than Colorado for these incidents per well and this potentially represents the magnitude of the unreported events in Colorado. Likewise, small “near-misses” could have resulted in much greater incidents and these types of fires and explosions may warrant reporting and investigation. The required reporting is important to prevent more serious accidents as it is important to be able to analyze, quantify, and learn from fire and explosion incidents to help prevent larger more catastrophic events.

The reporting of fires and explosions at O & G operation sites should also be available to the general public in easy to access formats. The Colorado accident and release reports were primarily available for use in this study due to the custom SQL and Ruby script technology that allowed documents to be downloaded in large quantities for analysis, which is a technology that is often not readily available to the general public or decision makers. However, the forms are made available by the COGCC through their website [23] thereby making this information available to the public, but the forms are not available for bulk download and the number of documents exceed the search limit. Likewise, manual content analysis of accident reports is a time consuming and costly endeavor. By comparison, Utah made the reports available in a compiled spreadsheet, which increases transparency and made it easier to evaluate than the Colorado reports, which were not compiled and available for download in PDF or TIF documents. Furthermore, to more closely align with the concept of targeted transparency, the information in both states could be organized and communicated in a way that is perceived as useful by stakeholders and could be implemented into public discourse and policy making [21].

Based on the results of our analyses, the current setback regulations to address fire and explosion risks at O & G operation sites need further scrutiny. Resident evacuations were reported and those incidents may increase stress for those living in close proximity to industrial sites [31]. To further discuss the concern that current setbacks may not adequately protect from fire and explosions, an incident in Utah reported an explosion where the “explosion blew the top of the water tank approximately 500 ft (152 m) down into the draw below location.” Also, a blowout in the Texas Barnett Shale led to an explosion that produced a 750 foot crater [as reviewed by [2]]. These incidents demonstrate that residences at the 500 foot distance can be impacted by accidents at these sites.

Previous research found the average number of homes displaced during blowout or fire incidents at O & G operation sites was 149 (with a maximum of 500) per reported event with evacuations [2]. We found that 18 fires or explosions occurred within the two basin study area in Colorado with more than 10 residences within a 1609 m radius. While the 1609 m radius used in this study is an acceptable area of evacuation for some incidents, such as the Clarington, Ohio well pad fire, residences at this distance are likely to have limited risk for smaller or more controlled incidents. Future research and recommendations may focus on site specific evacuation radii based on the operation conditions of an individual operation.

In the Denver Julesburg Basin and Piceance Basin, it has been estimated that in 2012, 355,781 (19% of the total population) and 12,467 individuals (7% of the total population), respectively, live within 1609 m of an O & G well [19]. These individuals may need to be evacuated or be at risk from O & G operation fires or explosions that are similar in size as the Clarington, Ohio well pad fire. Furthermore, Colorado residences can and are being constructed within the 152.4 m (500 ft) regulatory setback once a well has been established [19]. The potential then exists for fires and explosions to occur closer to residences in the future due to the lack of restrictions on residences being built within established setback distances once wells are operating. This calls into question the utility of residential setback distances across the nation for protecting near-by residents from fires and explosions.

5. Conclusion

To our knowledge, this is the first state-level analysis of fires and explosions at O & G operation sites and the first comparing the occurrence of these incidents in states with different reporting requirements. Two state level datasets were compiled to compare the rate of fires and explosions at oil and gas development sites in Colorado and Utah over a 10 year period. Our judgement is that the observed differences are likely explained by the varying reporting requirements for these incidents. In Colorado, we also found areas with significant numbers of residences within 1609 m of a reported fire or explosion, which shows the potential for evacuation or direct risk from these incidents for populations living in close proximity to O & G operations. Based on these findings, we present policy recommendations, including the need for standardized and transparent reporting requirements. Given the frequent occurrence of fires and explosions at O & G operation sites, these incidents are a concern for populations living and working in close proximity to these operation sites. This study provides a baseline for understanding the reporting of these events and provides insight into the fire and explosion risks at both urban and rural O & G operation sites.

IRB statement

This research was considered non-human subject research by the Colorado Multiple Institutional Review Board.

Funding statement

This work was conducted as part of the AirWaterGas Sustainability Research Network funded by the National Science Foundation under Grant No. CBET-1240584. Any opinion, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Acknowledgments

We thank Troy Burke for his technical assistance in downloading the documents of analysis and Lindsay Taylor for her assistance in reviewing the release report.

Appendix A. Colorado Oil & Gas Conservation Commission (COGCC) Accident and Release Reports Data Collection.

Overview

Accident and Release Reports were collected from the Colorado
Oil & Gas Conservation Commission (COGCC) website. Accident Reports were identified through a targeted SQL query on the title text of well documents. Document title is an attribute of well document metadata collected by using custom Ruby scripts to scrape HTML from the COGCC website. Once identified, the Accident Reports, in either PDF or TIF file format, were downloaded using a custom Ruby script. Accident reports not associated with a well were downloaded manually from the COGCC website.

Release Reports were obtained from COGCC’s COGIS Database by using the Inspection/Incident Inquiry Search function. Depending on the date of the incident, report text was extracted from either scraped HTML or from downloaded PDF file text. Both extraction methods employed automated Ruby scripting. Report text was stored in a Postgres database and report data elements were parsed by using SQL scripting.

Accident Reports

Accident Reports were obtained through COGCC’s Document WebLink Service. Document searches can be performed either through theImaged Documents search form or by searching each well’s document index that is linked from the COGIS Well Information Scout Card. Imaged Documents Search URL:
http://cogcc.state.co.us/wellData2.html#/images

As part of our COGCC data collection efforts, we cataloged the API well number, id, number, name, date, and URL of each well document. This metadata was collected for over 2.3 million documents from approximately 108,000 wells.

Using keyword searches on the document name value, we were able to identify accident reports. Documents were identified based on the following criteria:

- `document_name ilike '%%report%'
- `document_name ilike '%accident%report%'
- `document_name ilike '%report%accident%'

Due to the large number of documents, the report download process was automated using a custom Ruby script incorporating the Mechanize gem package. By replacing the value for the DocumentId variable in the Sample Accident Report URL (see above), we submitted an HTTP request for each document and saved the file response as either a PDF or TIF file, depending on file format.

Release Reports

Release Reports are obtained through COGCC’s COGIS Database using the Inspection/Incident inquiry search. Reports are structured in either HTML or PDF format (starting in early 2014, COGCC stopped providing reports in HTML format).

Incident Search Form URL:
http://cogcc.state.co.us/cogis/IncidentSearch.asp

Using the Incident Search form, we submitted a query for Spill/Release incident types and the specified study range. The search form results are limited to 2500 records. The number of reports for the study date range exceeds 2500 records. By using a custom Ruby script and the Mechanize gem, we were able to circumvent the record limit by replacing a higher value for the record limits field, thus returning all reports for the specified date range.

After recording the report metadata (submit date, document number, and facility id) from the search result records, we used automated Ruby scripts with the Mechanize gem to submit an HTTP request for each report. Depending on report format, HTML or PDF, different steps were needed to extract report data.

Reports in HTML format were parsed using Ruby scripting and the nokogiri gem. This custom script parsed an HTTP request for each report using the document number value. The HTML code returned from the request was parsed and each distinct report element was saved to a Postgres database.

PDF Reports

The download of PDF reports was automated using a custom Ruby script incorporating the Mechanize gem. The PDF text was extracted to plain text format using additional Ruby scripting and the PDF-Reader gem package. Once extracted, the text was saved in a Postgres database. SQL scripting was used to parse the saved text and find each distinct report element.

Well Document Metadata

The API Well Numbers for all Colorado oil and gas wells were extracted from a GIS file download available on the COGCC website. Well metadata was read from the shapefile download and saved to a Postgres database by using the pgShapeLoader (PostGIS Shapefile Import/Export Manager) application.

Well Surface Location Data:
http://cogcc.state.co.us/documents/data/downloads/gis/WELLS_SHP.ZIP

Using a custom Ruby script, an HTTP request was placed for each API well number. The request returned HTML code listing all documents associated with a well.

Example well document list:
http://ogccweblink.state.co.us/results.aspx?id=12314699

The custom Ruby script automatically paged through the document index to ensure all documents were cataloged. For each document, the URL, date, id, number, and title were saved to a Postgres database.

References


