Untapped Energy

Datathon For the Oil and Gas Industry
October 12 - 14, 2018

EXECUTIVE SUMMARY –
AGING INFRASTRUCTURE

Pod Members
Veronica Daigle
James Louie
Jamie Yulo
Jayachandran Ramachandran
Priya Sanjay
Victoria Kusherbaeva
Kevin Rush
Michelle Wan
Prashanth Sundaravadivelu
Heli Gong
Laura Li
Emma Liu
Steve Johnson
Monica Sippola
Dustin Symes
Tim Chan

**** special thanks to JuneWarren-Nickels Energy Group for generously providing valuable data that this pod used ****

1. INTRODUCTION

OVERVIEW

Alberta's long history of energy development, with oil and gas exploration and development activities beginning more than 100 years ago, have led to today's extensive industry with some 150,000 current producing oil and gas wells. Over time, thousands of inactive energy structures scattered across the province can be found in varying states of upkeep – these are collectively known as "aging infrastructure." This includes pump jacks that no longer pump, pipelines that no longer transport oil or gas, mines that no longer operate, and gas wells that are closed off. These all represent the possibility of adverse impacts should proper decommissioning not occur. To consider these potential impacts using data science, the Aging Infrastructure pod was formed.

OBJECTIVE

Using available datasets, the Aging Infrastructure pod sought to investigate these challenges by focusing on 2 objectives:

1. What factors contribute to FAILURE of aging infrastructure?
2. What factors contribute to the EFFICIENCY of aging infrastructure?

OUTCOMES

Due to the lack of completeness of data within the available datasets, analysis was primarily centred around wells rather than pipelines or other infrastructure facilities. Results of the analysis on the relationship between age and incidents is summarized in section 2, which uncovered that the highest number of incidents reported occurs within a well's first year and that a clear association between the advanced age of a well and reported incidents could not be seen. Section 3 considers an efficiency issue associated with aging infrastructure, summarizing the analysis of whether suspended legacy midstream infrastructure constrains production.

A number of gaps in data quality were observed, summarized with some examples in section 4. Lastly, in section 5, several recommendations are proposed to consider for future datathon events with the intent of continuous improvement to further the growth and positive impact of data science contributions to Alberta's oil and gas industry.

METHODOLOGY

Using the datasets, tools and technologies listed below, the team assessed several angles of the age of infrastructure and incidents. The results of these, including several summary graphs and charts, are available through the Untapped Energy website: www.untappedenergy.ca

EXTERNAL DATA

Data sources used by this pod:

- JWN Energy – Production and Injection tables
- AER Field Surveillance Incident Inspection List
- AER Pipelines Installations Shapefiles
- NEB Incidents
- NYS Abandoned Wells
- Pipeline Incident Data
- Pipeline Performance Measures
- Pipeline Ruptures

- Pipeline Throughput and Capacity Data
- NEB Compliance and Enforcement Information
- NEB Media Inquiries 2006-2018
- Well Characteristics
- Alberta Township Survey Shapefile
- AER Administrative Boundaries PDF (created Shapefile in QGIS)

TOOLS AND TECHNOLOGIES

Tools and Technologies used by this pod:

- Microsoft Excel – data conditioning and data wrangling
- Microsoft Power BI – data modelling and visualization
- Microsoft Azure – machine learning
- Spot Fire – data wrangling and visualization
- Tableau – visualization
- QGIS
- LSD: www.lsdfinder.com – mass convert LSD to GPS coordinates

KEY LEARNINGS AND TAKEAWAYS

The Aging Infrastructure pod used various data science techniques and tools to determine that:

- The expected correlation between well age and incident is not a strong as initially believed.
- Most spills for wells occur in the first year of production, and that most of these occur just post-drilling.
- Artificial intelligence / machine learning can be used to accurately predict, and visualize, when an active well should be abandoned.
- Voluntary well abandonment has a tendency to not occur, particularly when there is no incentive to do so. Without regulatory pressure/consequences, a well's active well status is not always updated to "suspended" when wells stop producing. These wells sit untouched for years presenting a risk, and there is no timeline on when a "suspended" well needs to deemed "abandoned".
- There might be an opportunity to coordinate disparate licensees in a way which would allow for strategic re-activation of key facilities that can help service the flow of oil and NGL.
- In the absence of complete or high quality data, or when reporting structures change after being established, the potential gains from data analytics cannot be fully realized.

2. AGE AND INCIDENTS

HOW DOES THE AGE OF WELLS AFFECT THE INCIDENT RATE?

Aging infrastructure was expected to result in an increase of releases. The rate of incidents resulting from wells was expected to increase with age. Incidents are not always the result of equipment failure; operator error and design flaws are common causes. The insights were expected to allow operators to prevent failures through appropriate planning and risk management. The investigation of the effect of age on well incident rates yielded some surprising results.

Data used for the analysis included:

- JWN Well Facts, which described well parameters such as drilling dates, depth and geology, and
- AER Field Surveillance Incident Inspection, which described well incident details since 1975.

When the incidents were joined to the wells and the age of the well at the time of incident was plotted, it showed the majority of incidents occurred in the first year, with a sharp downward trend afterwards, consistent with the downward trend of the max age of wells. This trend was held true for wells drilled in every decade for which data was available. This demonstrates a high risk of incident in the first year and steady risk as the well ages. When causes were investigated the majority of incidents in the first year were due to operator error, followed by equipment failure. Overall causes of incidents were most frequently equipment failure followed by operator error. Some data deficiencies were noted in that the cause data prior to 2003 was not collected and is represented by a null value.

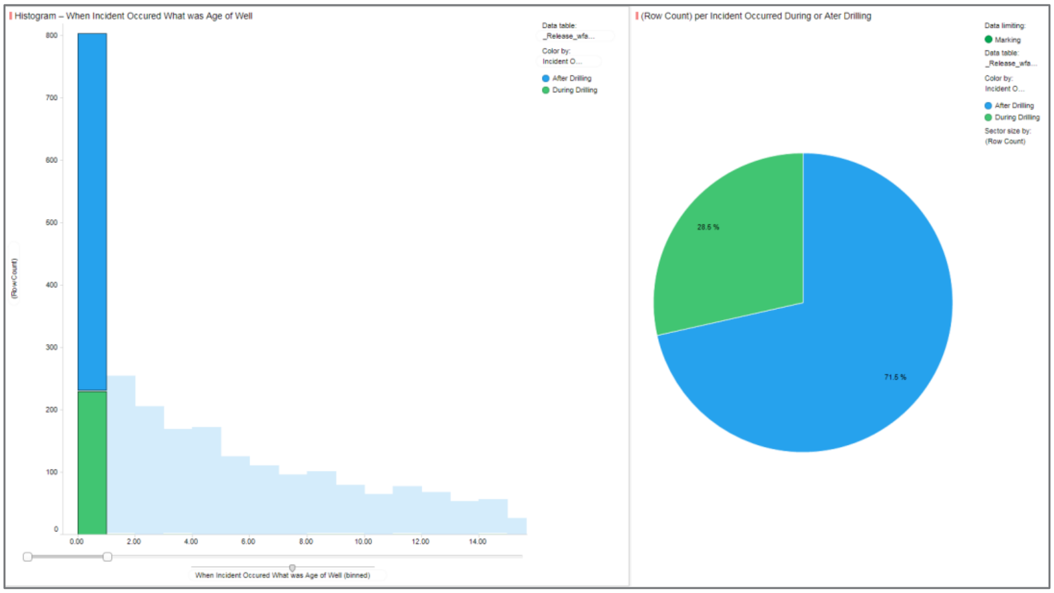

Fig. 1 – Drill down on year 1 of well spills; split by During Drilling vs. Post Drilling

An analysis of which types of well are most likely to have an incident was also completed. The percentage of each well type that had one or more incidents associated with it was plotted. Injection/disposal wells had the highest percentage of wells with one or more associated incidents.

Suggested further analysis may include analyzing recent spills since 2003 and filtering out incidents caused by human error and analyzing the rate of incidents with relation to age.

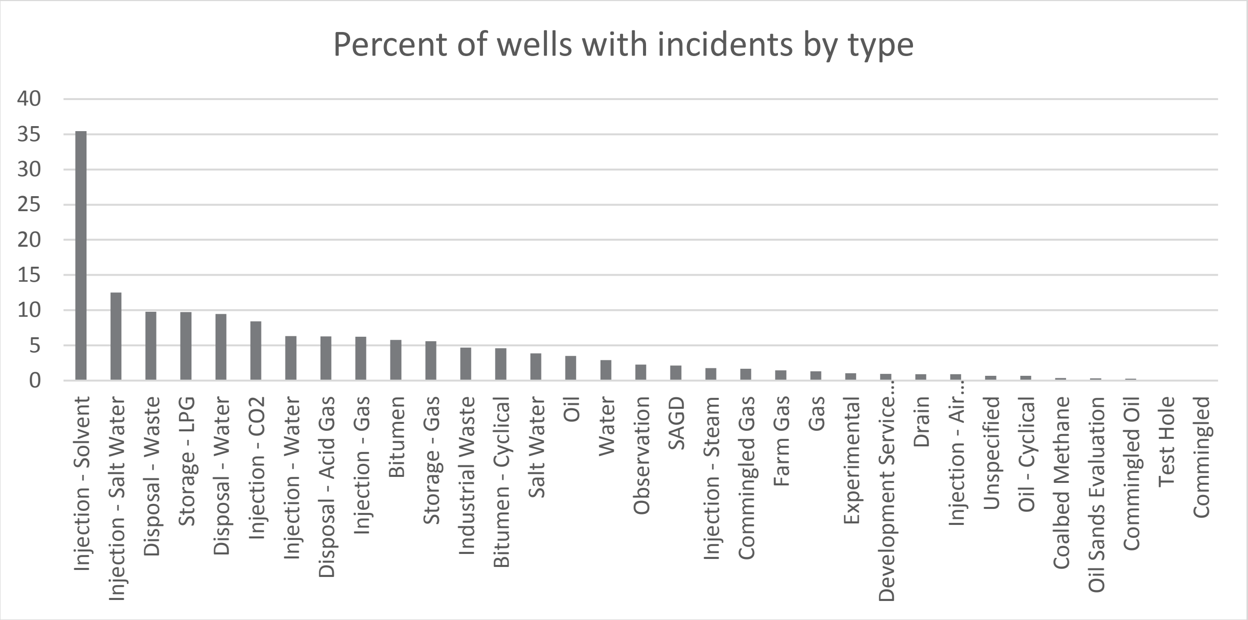

Fig. 2 – Percent of each well type with at least one associated incident

PIPELINES AGE AND INCIDENTS

Data for pipeline age was limited and analysis of the relation of age to incidents was not possible based on the available data. Construction dates were not readily available for many older pipelines due to a lack of collection by the regulator. An attempt was made to plot the pipeline ages on a provincial map, the extent of the null data can be seen in the resulting visual.

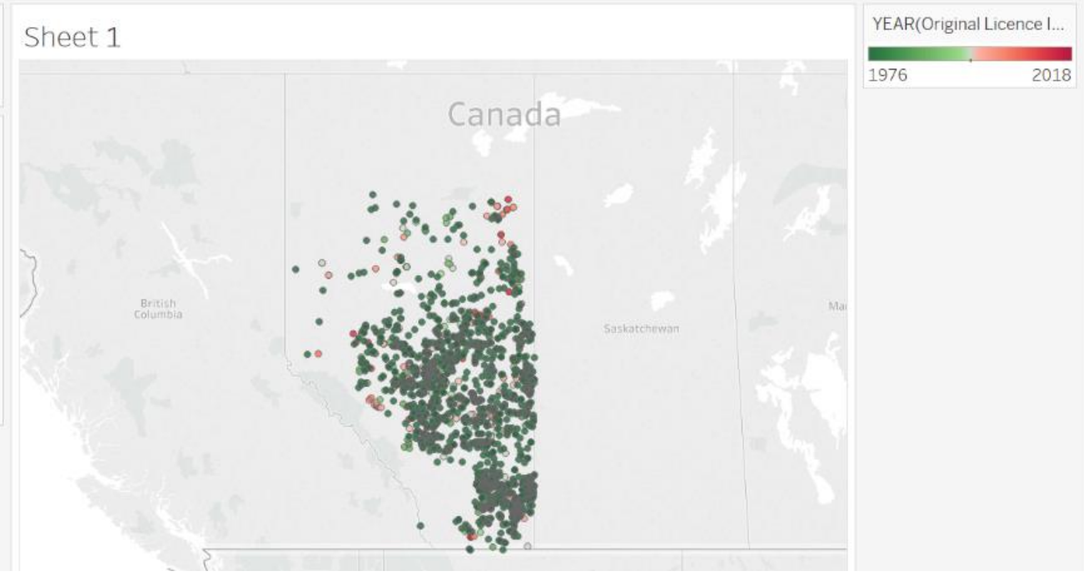

Fig. 3 – Original License Date, (grey is empty data)

Some National Energy Board pipeline spill data was available, but similar difficulties were encountered associating the spill data with pipeline age. The majority of NEB pipeline incidents were due to maintenance deficiencies.

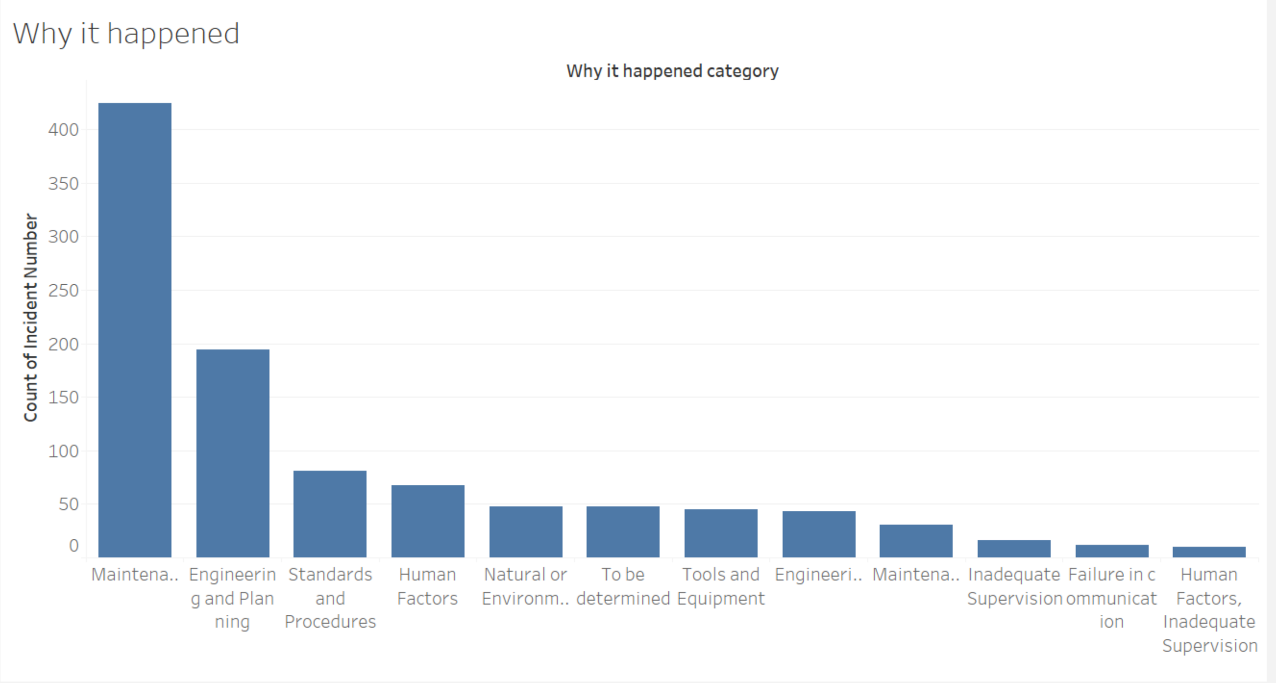

Fig. 4 – Incident by Cause Category

FACILITY AGE AND INCIDENTS

Data for facilities was sparse and inconsistent, while age data in particular was not available for facility licences. As such, an analysis of the relationship between facility age and incident rates was not possible. Due to the large scope of the aging infrastructure issue and the limited time available for analysis, analysis of incident types on facilities was not completed. This could provide an interesting avenue for future investigation and analysis.

3. AGE & EFFICIENCY

DOES SUSPENDED LEGACY MIDSTREAM INFRASTRUCTURE CONSTRAIN PRODUCTION?

Market access is a key challenge faced by Canadian E&P companies, as seen by the many suspensions/dispositions of midstream facilities in key operating areas. At the same time, other operators can be seen dedicating resources towards constructing new facilities to help manage market access. Does production activity in key regions support the re-activation of suspended facilities?

The data used looked at operational statuses of midstream facility infrastructure and their relationships to wells. A relationship map was constructed to reconcile facilities to their respective license numbers, which then ties facilities to the wells connected to them. Finally, the data was joined into the JuneWarren-Nickels well dataset, which provides a complete picture of well and facility details, allowing further analysis into the nuances of facility profiles.

To determine how much conventional well production volume was associated with suspended facilities, and how much potential well production is associated with suspended wells, Power BI's embedded DAX programming capabilities was used to manipulate the resulting large data set (~4.5 GB).

FINDINGS

The analysis suggests that there are substantially more wells in the AB2 region that are not explicitly tied into a facility than in any other region, where a significant number of Drilled/Uncompleted (DUC) wells are present[1]. The AB2 region includes the Montney/Duvernay region, which has been seen to face difficulties in market access and midstream capacity.

[1] DUC wells are generally associated with potential production that is being held sub-surface as inventory. A company will keep this production in storage as it waits for better prices, or midstream tie-in.

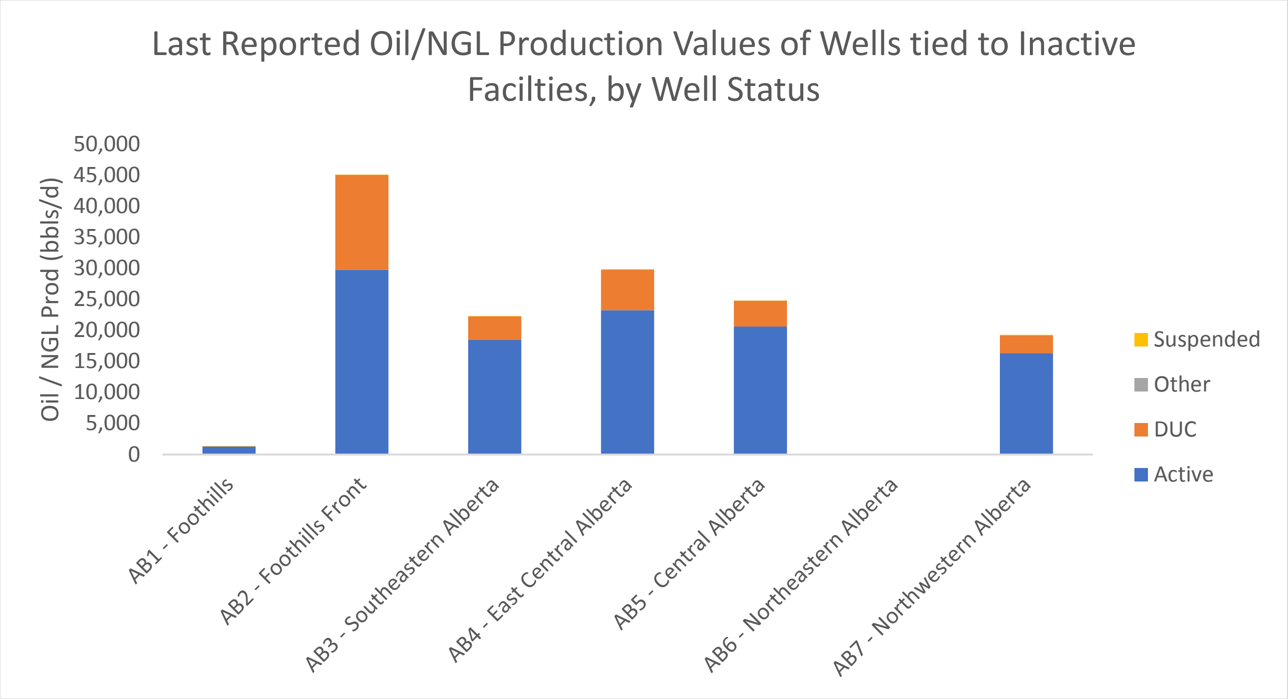

Fig.5 – Last reported production per AB region – Oil/Condensate Production

Based on a simplified assumption that production tied to suspended facilities is still making it to market, about 30,000 bbls/d of active AB2 Oil/Condensate production is not explicitly being tied into an active facility. Furthermore, this orphaned production would benefit from direct facility access, but market inefficiencies may be preventing this.

Further analysis then examined a distribution of the orphaned production that is attributed to its licensee, which revealed that there are 200 different licensees contributing to back-logged production, indicating that this is an industry pervasive issue.

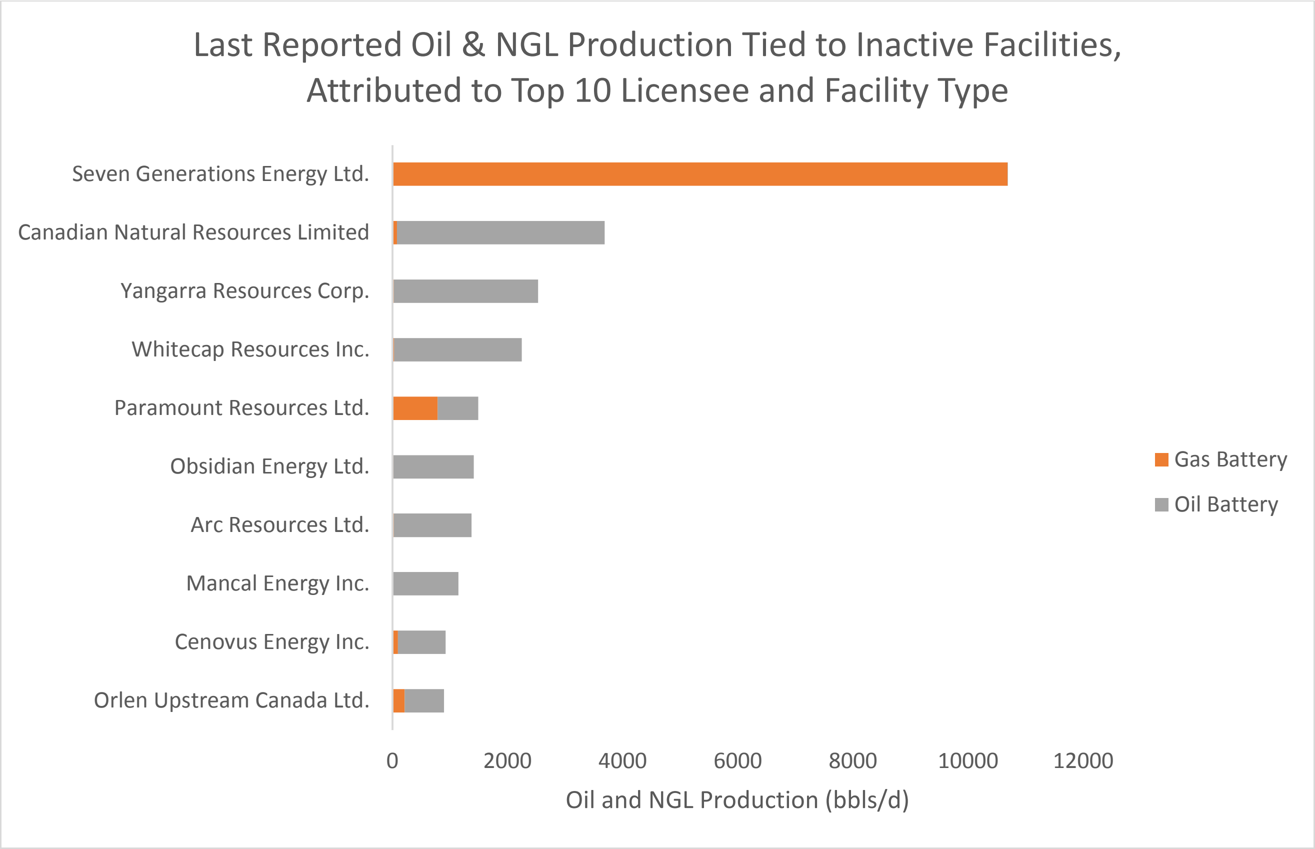

Fig. 6 – Last reported production per license and facility type

CONCLUSIONS

The analysis provides preliminary support of facility-related production constraints, particularly in the AB2 region, suggesting that facility shut-ins may be an issue that requires further research. The data suggests that there is a tendency for active wells to be tied into inactive facilities, which possibly inhibits market access. While it is likely that this production is still finding its way to market, there might be an opportunity to coordinate the backlog attributable to multiple disparate licensees in a way that allows for strategic re-activation of key facilities that can help service the flow of oil and condensates.

4. DATA GAPS

THE IMPORTANCE OF DATA QUALITY

Poor or incomplete data can lead to biased or unrepresentative conclusions. Much effort is needed to cleanse data to the point of where it can be used: identifying and correcting of incomplete, inconsistent, or inaccurate data. The validation of data then ensures that data matches expectation, and strengthens the data's validity and reliability – the Truth. Common dimensions that describe data quality include: usability, precision/resolution, timeliness, non-duplication, consistency and completeness[2].

The Aging Infrastructure pod used primarily openly available data sets for the Datathon, sourced from various government agencies: the National Energy Board and the Alberta Energy Regulator (via the Petrinex portal). Addition curated data was provided generously by JuneWarren-Nickels Energy. One data set was purchased directly from the AER.

During the data preparation stage, it was quickly discovered that data quality challenges would hamper the efforts to answer specific aging infrastructure questions. Key data fields were either missing or incomplete in the available regulatory reports, such as:

- Pipe material
- Coating material
- Year of Manufacture
- Year of Installation
- Facility data (throughput, current status, abandonment date)
- Location information (pipeline shape files show facilities and surface locations but not actual pipe segments)
- Date data - for example, the AER and NEB Pipeline data contains many "null" dates where the reporting operator or company did not provide any information.

Without appropriate date information, the age of equipment is difficult to determine and assessing whether incidents or failures can be attributed to its age, and possibly for consideration of preventative maintenance activities, cannot be determined.

[2] https://www.cebglobal.com/member/audit/assetviewer.html?filePath=/content/dam/audit/us/en/General/PDF/17/09/data-scoping-sourcing-and-cleaning.pdf&contentType=research&searchString=&pageContentId=201456955&pageRequestId=26d717e9-4615-4888-aa0f-0c79cc1fce04

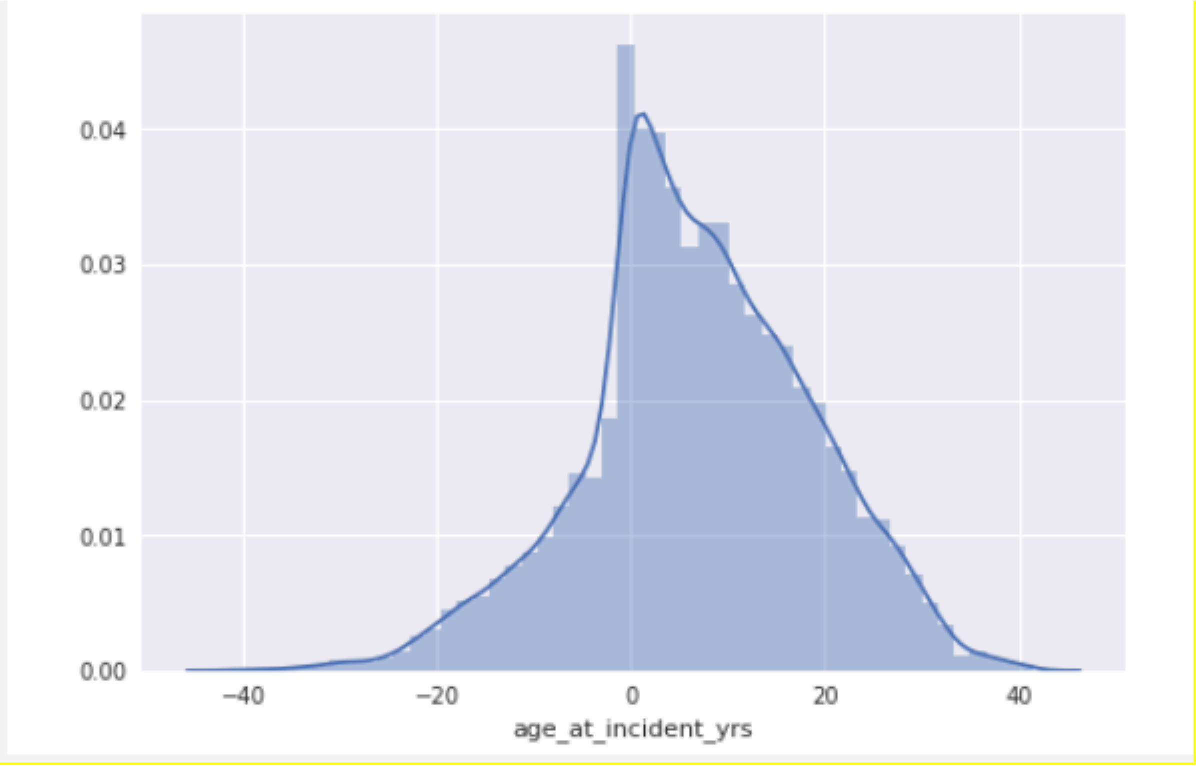

Fig. 7 – Example of incomplete data – well age in negative years

While analyzing the AER pipeline data, when the inferred "age" of a pipeline is compared to an incident date, negative ages appear as illustrated in the graph above. Obviously impossible, this is due to incomplete or incorrect data in the dataset, making it challenging to determine the correct age of infrastructure and possible correlations.

Data consistency contributes to data validity and reliability. The available data was observed to be inconsistent in the following ways:

- Pipeline incidents have many conversion errors (i.e. data could not be converted to a valid status)
- Different provinces use the same license numbers (i.e. cannot combine these data sets because the license number is not a unique identifier)
- The definition of "well status" is not the same from province to province (e.g. in BC and Saskatchewan, the status "active" is the same as "flowing" in the province of Alberta)

SPECIFIC EXAMPLE OF DATA INCONSISTENCY: MANDATED REPORTING REQUIREMENT CHANGES

Effective data analytics leads to effective decision making. This is underpinned by the underlying structure of how data is captured and stored. When changes to the data structure and collection occur, the opportunity to consider "apples to apples" over a historical period may be not be possible.

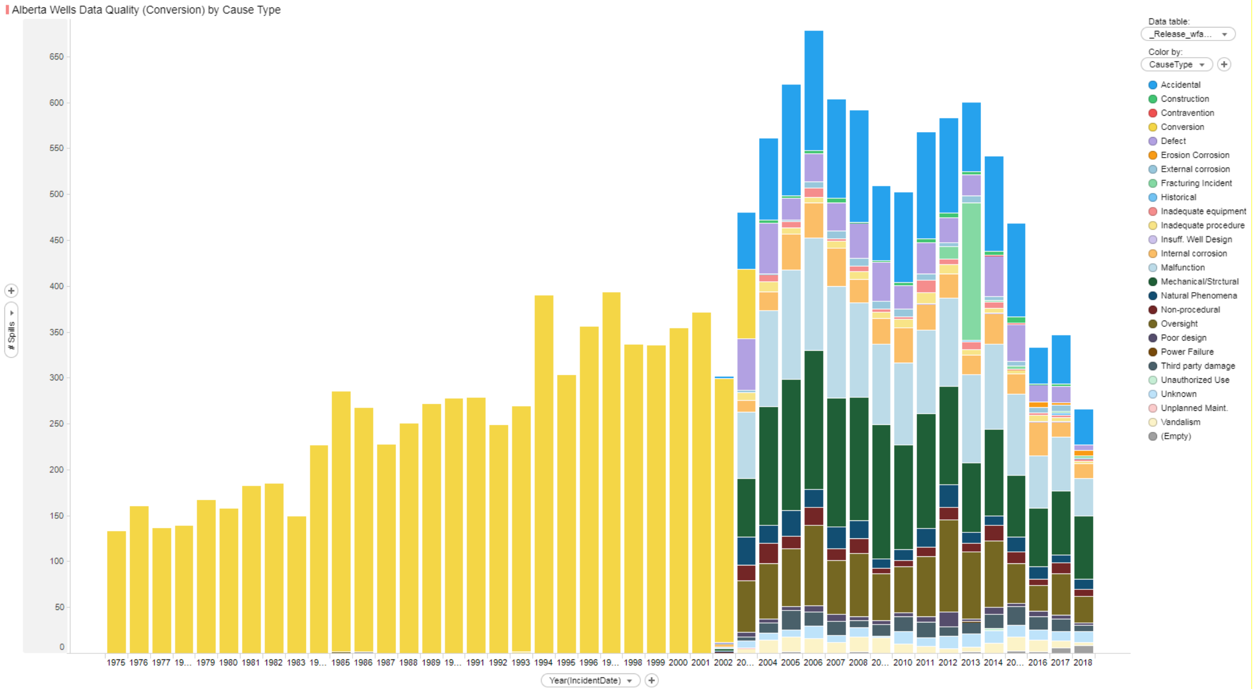

Fig. 8 – Alberta wells data quality (conversion) by Cause

In this example, the AER changed its reporting framework in 2002, which caused the required reporting to be more granular. This conversion can be seen clearly in the data. As such, analysis applied to data after 2002 is different from that prior to 2002, resulting in a fragmented data narrative.

Having an impassionate community of data practitioners gather to solve problems is a great starting point, which the Datathon succeeded in doing. However, the greater benefits that result from such a gathering are dependent on quality data sets that are openly available. Data sets currently available from government agencies are inadequate for driving insights for solving problems within a collaborative Datathon setting.

5. RECOMMENDATIONS

As the inaugural oil and gas datathon event, some observations were made, offering "lessons learned" for future oil and gas data analytics efforts to improve upon results. In the spirit of continuous improvement, a few suggestions are offered:

1. **Narrow down problem statement with subject matter experts.** Tackling a large topic of "aging infrastructure" is a significant effort with many different possible angles and data sources to assess the issue. Distilling the problem with a subject matter expert familiar with the both the issue and datasets would make it a more manageable bite-size to investigate over a short time period (such as a datathon weekend event) and likely lead to usable insights.
2. **Plan approach and algorithms with a data scientist.** Along with a subject matter expert, it's also recommended a data scientist participate ahead of the datathon event to jointly plan the approach and algorithms addressing the problem statement. This would contribute to the maximization of the value of the available data, identification of possible issues to troubleshoot pre-event and determination of analytical approaches and required skill sets.
3. **Request specific skill sets as "job postings".** During the datathon recruitment, inform the community of attendees of any specifically required skill sets, whether subject matter expertise, coding languages, visualization techniques, etc., that may be needed to tackle the team's topic.
4. **Pre=event dataset preparation.** To maximize efficiencies during the datathon event, it would be most beneficial for team members to familiarize themselves with the datasets and have any pre-work, data cleaning or preparation done ahead of the event.
5. **Improve dataset quality.** Some of the public datasets used had incomplete information, likely submitted "as is" by reporting companies. Without reasonably complete or consistent datasets, proper analysis and investigation into issues is hampered, leading to lack of conclusive insight. With the growing need for data science to support the industry find efficiencies, now is the opportunity for dataset owners to address these gaps in data submission.