

Um caso de Democratização de Dados na Indústria de Óleo e Gás

Nathalia Marassi Cianni¹, Kellyne Marques Santos¹, Dean Pereira de Melo¹

¹Petroleo Brasileiro S.A. (Petrobras)
av. Henrique Valadares, 28 – 20231-030 – Rio de Janeiro – RJ – Brasil
{nathalia.cianni,kellyne,deanmelo}@petrobras.com.br

Abstract. *Making business decision more data driven led to an increase in the need of having data as available as possible. This work presents the data democratization process in an oil and gas company. This process was based on simplifying the data access policies and was organized into phases: from risk analysis to its implementation in the main databases and systems; its adoption was incremental. As a result, it increased the availability of data for the company's professionals and more time to do their analyses.*

Resumo. *As decisões de negócio estão cada dia sendo mais direcionadas pelos dados e há uma necessidade crescente em ter os dados o mais disponíveis possível. Este trabalho apresenta o processo de democratização do acesso a dados em uma empresa do ramo de óleo e gás, baseada na simplificação da política de acesso a dados. A democratização de dados foi dividida em etapas desde a análise de riscos até a sua implementação nas principais bases de dados e sistemas, sendo adotada de forma progressiva. A simplificação da política de acesso aumentou a disponibilidade dos dados aos profissionais e possibilitou um maior tempo para execução das análises em si.*

1. Introdução

A indústria de petróleo e gás tornou-se uma indústria massiva e intensiva em dados. Estudos e relatórios afirmam que engenheiros de petróleo e geocientistas gastam maior parte do seu tempo na busca, acesso e coleta de dados [Mohammadpoor, 2020], em alguns casos chegando a 84% do tempo total de um projeto [Melo et al., 2019]. Além disso, a difusão de práticas de ciência de dados e de tecnologias digitais trouxe à tona a necessidade das organizações de se transformarem: de uma gestão de dados mais tradicional para uma gestão de dados que apoie à tomada de decisões de negócio orientadas a dados [Samarasinghe, 2022].

Apesar do uso, compartilhamento e integração de dados por qualquer pessoa, em qualquer lugar, para qualquer finalidade ter se ampliado com os dados abertos [Open Knowledge Group, 2005] e esta ter se tornado uma realidade cada vez mais presente na nossa sociedade - como ocorre com a Lei de Acesso à Informação [Brasil, 2011] e o Portal Brasileiro de Dados Abertos [Brasil, 2012] - as organizações não têm seguido o mesmo compasso para facilitar o acesso aos seus dados internamente. A aplicação do conceito de dados abertos dentro de uma organização, atingindo o maior número de empregados

possível e considerando as limitações legais de confidencialidade e segurança, é chamada de democratização dos dados [Awasthi, 2020].

Perseguindo o objetivo de democratizar os dados, silos existentes nas organizações vêm sendo removidos de forma a permitir a abertura para usuários internos – uma vez que o uso dos dados democratizados por empregados aperfeiçoa o ecossistema de dados da organização e torna-o mais *user-friendly* [Samarasinghe, 2022]. A área de Exploração e Produção (E&P) da Petrobras iniciou o investimento na remoção de silos de dados em 2001, ao patrocinar a criação da Base Integrada de Dados de E&P (BDIEP) [SBGF, 2011]. Essa base unifica informações adquiridas externamente e geradas pelos principais sistemas de E&P desenvolvidos internamente na empresa, tanto em formato estruturado como não estruturado (arquivos), e armazena dados transacionais, informacionais (*data warehouse* e *data lakehouse*) e séries temporais (*near real-time*).

Além do pilar de dados abertos, a democratização de dados conta com outros dois pilares: a responsabilidade (*accountability*) compartilhada dos dados e o compartilhamento das habilidades em dados [Awasthi, 2020]. Para atender a esses dois últimos pilares, a Petrobras e suas áreas de negócio têm políticas e diretrizes que definem os atores envolvidos na gestão dos seus dados e atua na alfabetização dos seus empregados, através de trilhas de conhecimento e oficinas de gestão de dados, inclusive especializando esses cursos por segmento, como no caso das Geociências de E&P.

Embora a Petrobras tenha patrocinado a implementação de arquiteturas de dados para a remoção dos silos e promovido a alfabetização de dados dos seus empregados, por questões de segurança da informação e auditoria, o acesso ao dado necessita da autorização de atores como o gestor de dados – para acesso direto ao dado – ou o gestor de aplicação – caso o acesso seja realizado através de um sistema, e esse processo de acesso aos dados aparece como queixa recorrente desde os usuários de dados mais técnicos até os executivos da empresa.

O processo de concessão de acesso descrito nesse trabalho refere-se a quatro casos principais: a) acesso de um usuário em um sistema, concedido pelo gestor da aplicação; b) acesso de um sistema a dados, concedida pelo gestor de dados; c) acesso de um usuário ou um grupo de usuários a um conjunto de arquivos, concedida pelo gestor de taxonomia (um gestor de dados definido para um conjunto de arquivos relacionados); d) acesso de usuários a dados tanto em ambientes informacionais tradicionais (*data warehouse*) como na nuvem (*data lakehouse*). Em todos esses casos, há relatos de insatisfação em relação ao tempo necessário para um usuário ou sistema ter acesso às informações. Em 2019, foi realizado um levantamento do tempo médio para atendimento das centenas de solicitações de acesso recebidas e observou-se que demorava, em média, 22 dias para que a concessão acontecesse. Analisando esse processo, foi identificado que ele envolvia muitos atores e o resultado era quase sempre o mesmo, ou seja, todas as solicitações a dados eram aprovadas. As únicas exceções eram as solicitações que envolviam dados confidenciais ou que não fosse para uso interno à companhia. Além disso, os sistemas ainda aplicam uma segunda camada de acesso aos dados, podendo segregar os usuários por perfis associados a uma delimitação geográfica ou à unidade organizacional que trabalhe, recriando a visão de silos de dados.

A dificuldade no acesso à informação pode representar uma significativa perda de competitividade, tanto pelo impacto no tempo quanto na qualidade da tomada de decisão. Um exemplo disso é que alguns dados que não estão disponíveis à força de trabalho

podem ser adquiridos por empresas concorrentes através da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) após um período de sigilo definido por essa Agência [ANP, 2022].

A visibilidade das questões relacionadas a dados tornou-se ainda mais relevante quando a empresa sinalizou, como ambição de alguns dos seus programas estratégicos de E&P, a utilização de 100% dos seus dados em projetos exploratórios e o aumento das suas reservas de petróleo através do desenvolvimento dos melhores modelos de reservatório da indústria, modelos estes que são grandes agregadores de dados.

Por todas essas razões, as áreas de gestão de dados de E&P vêm promovendo, desde o final de 2019, a democratização através da simplificação da política de acesso a dados. As seções seguintes irão apresentar o processo de implementação da democratização de dados na Petrobras, passando pela submissão da proposta de simplificação da política de acesso, avaliação dos riscos envolvidos, análise de impacto para implementação nas principais aplicações até os principais resultados obtidos.

2. Implementando a Democratização de Dados no E&P da Petrobras

Em 2019, a equipe de governança de dados geológicos da Petrobras levou uma proposta de simplificação da política de acesso a dados internos para o Comitê de Dados de Geociências. Dados internos podem ser entendidos como dados e informações geradas pela companhia que não são públicos, confidenciais nem secretos, mas nos quais podem ser adicionadas barreiras de restrição de acesso internamente à companhia.

A proposta de simplificação da política de acesso submetida ao Comitê de Dados defendia que o acesso de leitura aos dados internos deveria ser facilitado de forma que seu uso por parte dos empregados não fosse mais limitado por área geográfica, gerência funcional e/ou cargo/ênfase. A proposta inicial restringiu-se aos dados disponíveis na Base de Dados Integrada de E&P (BDIEP). A proposta foi restrita a essa base devido ao seu rigoroso processo de gestão e governança de dados e pelo fato de ser o repositório de dados da maior parte dos sistemas de E&P desenvolvidos internamente.

Um subcomitê foi designado para realizar a análise de risco da simplificação da política de acesso. O trabalho constituiu-se em mapear os riscos e identificar o impacto e a probabilidade de ocorrência deles ao aumentar a abrangência do acesso aos dados. Os riscos mapeados referiam-se à classificação e uso incorreto da informação. Por falta de acesso ao histórico de incidentes de informação, foram realizadas entrevistas na comunidade técnica e gerencial. Nenhum incidente foi reconhecido e não foram identificados outros riscos além dos previamente mapeados. Realizou-se também a revisão da classificação das informações de forma a evitar a simplificação de acesso a dados que pudessem ser sensíveis para a empresa (dados confidenciais ou secretos). Adicionalmente, foi realizada uma análise de impacto da implementação da proposta de simplificação de acesso nos principais sistemas utilizados pelo segmento, não sendo identificados quaisquer impedimentos. Além disso, foi verificado que a decisão da simplificação da política de acesso a dados estava alinhada ao padrão corporativo de segurança de informação que prevê perfis de acesso padronizados.

De posse dessas informações, o Comitê de Dados oficializou a democratização do acesso a dados internos, que contemplou inicialmente um dos segmentos das Geociências de E&P e incluiu:

1. A concessão do acesso a dados e arquivos internos geridos pela Exploração (mantidos nos ambientes transacionais, informacionais, *real-time* e de arquivos da BDIEP) para profissionais dessa área, através dos principais sistemas internos, removendo os silos que haviam com a segregação de acesso de usuários por perfis;
2. A concessão do acesso direto a dados internos da Exploração em bases analíticas (*data warehouse* da BDIEP) e plataformas de dados (*data lakehouse*, uma evolução da arquitetura da BDIEP) para profissionais dessa área;
3. A concessão do acesso a dados internos da Exploração em todos os ambientes da BDIEP para os sistemas geridos por essa área.

Quase dois anos após o início oficial da democratização dos dados, foi realizado um diagnóstico para avaliar os *gaps* nos projetos de ciência de dados e o acesso a dados apareceu como segundo problema mais citado. Dentre outras razões, três se destacaram por estarem relacionadas ao processo de democratização:

1. Impressão de que os sistemas implementavam diferentes processos de concessão de acesso, que poderia estar relacionada ao fato de que eles não teriam se adaptado à simplificação da política de acesso ou não teriam divulgado essa adequação;
2. Dificuldade de acesso a dados de softwares de prateleira (bases comerciais), que não estavam no escopo inicial da democratização por não usarem a base integrada;
3. Dado o acesso a uma maior quantidade de arquivos, identificou-se que arquivos classificados como confidenciais no passado deixaram de ter informações sensíveis e, portanto, precisaria passar por uma reclassificação do grau de sigilo.

Foi então definido um Grupo de Trabalho (GT) com a participação de diferentes áreas para atuar nas questões acima. Seus resultados levaram a uma ampliação da democratização para dados internos disponíveis em softwares de prateleira, à definição de um processo periódico de revisão de classificação de arquivos - mais de 40 mil arquivos já foram reclassificados até o momento – e a uma divulgação ampla do processo de democratização.

Após a realização do GT, foi iniciado um trabalho para expansão da simplificação da política de acesso para outras áreas que geram dados de geociências devido à grande sinergia dos seus conjuntos de dados e necessidade de consumo entre segmentos. Nessa fase, as áreas de Reservatório e do Centro de Pesquisa embarcaram no processo de democratização dos dados, seguindo os passos iniciais do segmento piloto (Exploração) e adicionando ações, como, por exemplo, a reclassificação do sigilo dos dados mestres e de referência para torná-los abertos a todos empregados da empresa e não apenas para as Geociências de E&P.

A seguir são apresentados alguns resultados e benefícios obtidos após a implementação desse processo.

3. Resultados da Democratização dos Dados

A Tabela 1 apresenta a quantidade de dados e de sistemas cujo acesso foi simplificado através do processo de democratização.

Tabela 1. Resultados da democratização de dados das Geociências de E&P

	Dados geridos por Exploração + Centro de Pesquisa	Dados geridos por Reservatórios
Quantidade de arquivos armazenados na base integrada disponibilizados automaticamente para usuários do segmento	3,93 milhões	0,63 milhões
Quantidade potencial de sistemas com acesso previamente autorizado a dados na base integrada	247	123
Quantidade potencial de usuários que podem acessar os ambientes informacionais	1395	1120
Quantidade de objetos de dados abertos a todos empregados do segmento	5425	1706
Quantidade de objetos (principalmente dados mestres e de referência) reclassificados quanto ao grau de sigilo para ficarem abertos a todos empregados da empresa	264	224

A implementação da simplificação da política de acesso a dados permite a melhoria da disponibilidade das informações internas de E&P e, conseqüentemente, apresenta os seguintes benefícios:

1. Incremento na cultura de tomada de decisão orientada a dados (*data driven*);
2. Redução do risco de perda de competitividade, pelo fato de profissionais de outras empresas terem acesso a informações que ainda não estariam acessíveis aos empregados da Petrobras ou pela demora na tomada de uma decisão dada a dificuldade de se ter acesso às informações existentes;
3. Melhor compreensão das incertezas envolvidas nos projetos e tomadas de decisão, visto que poderão fazer uso de uma maior diversidade de informações;
4. Melhoria na eficiência da execução dos projetos e nas tomadas de decisão – haverá redução de tempo por busca de informações e, conseqüentemente, diminuição do tempo total de execução dos projetos (tradicionais, de análise e ciência de dados);
5. Redução do trabalho dos gestores de dados e dos gestores de aplicação, que não precisarão autorizar o acesso de leitura a dados internos para o público (sistemas e usuários) que está no escopo desta proposta; e das equipes da TIC e das áreas de negócio que atuam em problemas de acesso a dados;

4. Conclusão e Trabalhos Futuros

O trabalho apresentou o processo de democratização que vem sendo realizado na Petrobras, iniciado com uma proposta de simplificação da política de acesso a dados disponíveis na Base Integrada de Dados de E&P (BDIEP). A proposta foi implementada após a avaliação dos riscos envolvidos e da análise de impacto de implementação nos principais sistemas desenvolvidos internamente.

Atualmente, os profissionais de Geociências de E&P podem acessar dados estruturados e arquivos com classificação interna, através de sistemas ou ambientes

informacionais sem a necessidade de aprovação do gestor do dado, gestor da taxonomia ou da aplicação; o que facilitou bastante os projetos de análise e ciência de dados. Sistemas também têm autorização prévia para consumo de dados internos sem precisar passar por avaliação do gestor de dados. Um dos desafios é garantir que as principais aplicações utilizadas para dados de geociências implementem a simplificação da política de acesso, principalmente numa visão de sistemas comerciais que têm questões adicionais como o número e o custo de licenças.

A simplificação da política de acesso a dados estimulou a criação de mecanismos para alterar a classificação de informações, dentro do seu ciclo de vida, a partir de critérios como temporalidade, assim como tornou menos complexos e burocráticos alguns processos de trabalho relacionados à gestão dos dados das Geociências de E&P. Percebeu-se, assim, que o processo de democratização de dados patrocinou ações em outras áreas de conhecimento da gestão de dados [DAMA International, 2017], visto que foram necessárias ações relacionadas à governança de dados, segurança de informação, metadados, gestão de conteúdo e qualidade de dados, tornando o processo de gestão dos dados mais perene e contínuo.

Considerando que há interação entre as disciplinas de Geociências e que o ciclo de vida dos dados permeia os diversos processos de negócio, foi planejada uma avaliação da ampliação da abrangência dessa proposta para outras áreas que atuam no E&P e, com isso, há uma expectativa de que essas ações inspirem a criação de uma política ou de uma estratégia de democratização de dados para toda empresa.

Referências

- ANP (2022) Resolução nº 889. <https://www.in.gov.br/en/web/dou/-/resolucao-anp-n-889-de-7-de-outubro-de-2022-434893600>.
- Awasthi, P. and George, J. (2020) “A case for Data Democratization”, In: AMCIS 2020 Proceedings.
- Brasil (2011) LEI Nº 12.527 (LAI: Lei de Acesso à Informação). http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm
- Brasil (2012) Portal Brasileiro de Dados Abertos. <https://dados.gov.br/home>
- DAMA International (2017) “DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition)”. Technics Publications, LLC, Denville, NJ, USA.
- Melo D.P., Brum M., and Gonçalves, C.A. (2019). Can enterprise search be inspired by Google search? A case study in the context of oil and gas exploration. *Acad. J. Environ. Sci.* 7(4): 030-036.
- Mohammadpoor, M. and Torabi, F. (2020) “Big Data analytics in oil and gas industry: An emerging trend” In: *Petroleum* volume 6, issue 4, p. 321-328.
- Open Knowledge Foundation (2005) Opendata. <https://okfn.org/opendata/>
- Samarasinghe, S. and Lokuge, S. (2022) “Exploring the Critical Success Factors for Data Democratization” In: ACIS 2022 Proceedings.
- SBGF (2011) “Bases de dados na indústria do Petróleo e Gás” In: *Boletim SBGF – Sociedade Brasileira de Geofísica*. p. 12-13