

Metodologia colaborativa para captura de problemas de Qualidade de Dados no âmbito da indústria de Petróleo e Gás

Kellyne Marques Santos¹, Rodrigo Moreira Araujo¹

¹Petroleo Brasileiro S.A. (Petrobras)
av. Henrique Valadares, 28 – 20231-030 – Rio de Janeiro – RJ – Brasil

{kellyne, rodrigoma}@petrobras.com.br

***Resumo.** Esse trabalho apresenta o resultado da combinação de técnicas de ideação, oriundas do design thinking, com técnicas para composição de planos de ação, como o 5w2h, para identificar e priorizar os principais problemas de dados a serem tratados em uma empresa multinacional de Energia.*

Minibio dos Autores

Kellyne Marques Santos, MBA, é especialista em Governança de Dados na Petrobras. Tem bacharelado (UFS) e mestrado (PUCRS) em Ciência da Computação, pós-graduação em filosofia e cursou Ciência de Dados na Itália (UNITO). Já desempenhou diversas funções no setor de TI como: desenvolvedora de software, arquiteta de dados, modeladora de dados, especialista em ETL, analista de dados, analista de negócios, facilitadora de dinâmicas etc. Foi professora em duas universidades no Brasil e atualmente da Universidade Petrobras.

Rodrigo Moreira Araújo é consultor de Ciência de Dados e atua na área corporativa de Reservatórios da Petrobras. Formou-se engenheiro de computação em 2008 e ingressou na Petrobras em 2009. Em 2016, fez mestrado em Engenharia de Petróleo e, em 2020, concluiu MBA em Ciência de Dados. Nos últimos 15 anos, também trabalhou como projetista de completações, engenheiro de reservatórios e instrutor da Universidade Petrobras.

1. Quem sou?

A Petrobras (2024) – Petróleo Brasileiro SA – é uma sociedade de economia mista, sob controle da União, que atua de forma integrada e especializada na indústria de petróleo, gás natural e energia. Tem expertise na Exploração e Produção (E&P) como resultado de décadas de desenvolvimento das bacias brasileiras, especialmente em águas profundas e ultra profundas, o que a tornou líder mundial neste segmento. Atualmente tem expandido seus investimentos em novas energias para uma transição energética justa.

Os dados gerados e registrados de forma geral na indústria de petróleo e gás aumentaram significativamente devido a melhoria nos dispositivos de aquisição de dados fornecendo grande quantidade de dados para serem processados e analisados [Mohammadpoor, 2020]. Contudo, com a crescente disponibilidade de dados de diferentes fontes e formas de representação, a qualidade dos dados surge como uma das questões mais importantes da gestão de dados [Côrte-Real, 2020].

2. Qual o nosso desafio?

Dado que a indústria de petróleo e gás lida com uma diversidade de dados e, conseqüentemente, com uma potencial diversidade de problemas de dados, era necessária uma metodologia de captura de problemas de qualidade de dados que contasse com técnicas e/ou mecanismos colaborativos para conciliar a visão dos diversos usuários de dados (funcionários) espalhados geograficamente desta empresa.

Levando em consideração que o processo de identificação e tratamento das questões de qualidade dos dados adquiridos e armazenados nesta indústria exige recursos - humanos, tecnológicos e financeiros - tanto na área de Tecnologia da Informação como nas próprias áreas de negócio e estes são finitos, o objetivo do trabalho foi definir mecanismos que direcionassem e priorizassem estes poucos recursos nos principais problemas de dados de determinada disciplina de negócio ou projeto contando principalmente com a perspectiva dos usuários de dados.

3. Qual a nossa solução?

Escolheu-se aplicar os conceitos do *design thinking* [Brown, 2008] por ser um processo de inovação centrado no ser humano e que enfatiza observação, colaboração, aprendizado rápido, visualização de ideias, prototipagem rápida de conceitos e análise de negócios simultâneos [Pham, 2022]. Um dos poucos trabalhos que aborda técnicas de *design thinking* com dados é o *data thinking* [Kronsbein, 2019], que fornece um protótipo de *canvas* a ser usado durante workshops de *design thinking* orientado a dados. Contudo este trabalho também não se aprofunda na captura de problemas de dados, mas numa visão do dado em seus produtos e serviços.

A metodologia proposta para captura e tratamento de problemas de qualidade de dados funciona em 4 etapas: (1) na etapa de **planejamento**, são selecionados os participantes que podem estar na Sede da empresa ou em uma das unidades espalhadas pelo país e os assuntos que serão tratados; em seguida, são utilizadas dinâmicas de grupo para (2) **mapeamento** dos dados utilizados em uma determinada disciplina funcional ou projeto estratégico; então, ainda em grupo, os usuários são convocados a (3) descrever os problemas (*data issues*) existentes nos dados levantados, sejam de qualidade, disponibilidade, acesso, etc. de forma a realizar um **diagnóstico** de *gaps* e problemas no ciclo de vida dos dados. Por fim, esses problemas são classificados e ranqueados, (4) gerando ações de alfabetização, responsabilização, estruturação, qualidade, integração, segurança/democratização dos dados, e de empoderamento dos usuários para utilizá-los, que compõem o **plano de ações** a serem tomadas para aumentar o grau de maturidade do ciclo de vida dos dados.

A etapa de planejamento define como a dinâmica será realizada (online, híbrida ou presencial); se será realizada para uma disciplina de negócios ou um projeto/case específico; há a seleção de participantes com papéis tanto de negócio como de tecnologia - como gestores e curadores de dados, arquitetos de informação, especialistas em governança de dados, usuários de dados, o próprio facilitador da dinâmica entre outros; são estabelecidos os *timeboxes* para a ideação junto ao *product owner* (caso seja um projeto) ou ao gestor/curador de dados (caso seja uma disciplina de negócio) e a preparação do quadro colaborativo. Este quadro é construído na plataforma de colaboração visual Miro (2023) contendo as fases realizadas na dinâmica, estas fases são representadas pelas suas letras iniciais na Figura 1 (M – mapeamento; D – diagnóstico;

PA – Plano de Ações). Ele é usado tanto nas interações online, híbridas como nas presenciais - neste caso como registro da interação em quadro físico.



Figura 1. Visão geral do quadro colaborativo utilizado na dinâmica

Foi decidido utilizar a etapa de Ideação para captura de problemas de dados – fase do *design thinking* em que profissionais geram e testam ideias na busca de soluções para problemas delimitados – mais precisamente utilizando a técnica do *double diamond* [British Council, 2023] (Figura 2).

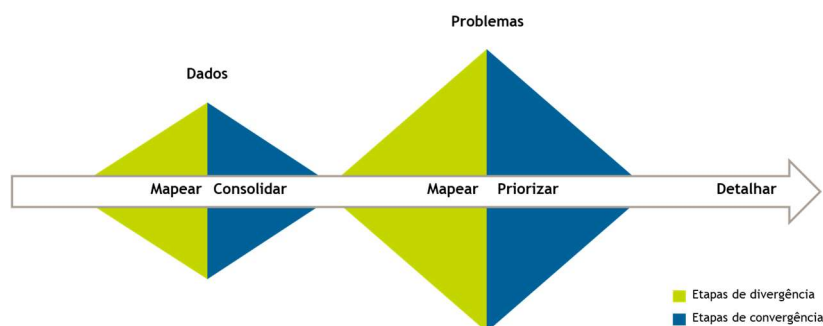


Figura 2. Double Diamond adaptado para captura de problemas de dados

No primeiro diamante, o foco é mapear os dados de interesse do grupo usando o quadro colaborativo online, levando em consideração a distribuição geográfica dos participantes. Na etapa de divergência, todos são incentivados a listar todos os dados relevantes para a disciplina ou caso de negócio em avaliação, sem limite de quantidade. Na segunda parte, os participantes são convidados a agrupar os dados de acordo com critérios propostos por eles ou pelo facilitador em tempo real. O objetivo é agrupar os dados em conjuntos com características semelhantes para serem utilizados nas próximas etapas; se mesmo assim os agrupamentos de dados forem acima de um *threshold* predefinido, pode haver uma priorização através de votação.

A etapa de diagnóstico segue uma estratégia semelhante à de mapeamento: inicia-se com um levantamento de questões ou problemas nos conjuntos de dados identificados na etapa de divergência, mais uma vez de modo irrestrito. Em seguida, os participantes votam para identificar quais os problemas de maior impacto nas suas atividades. O resultado dessa votação é utilizado para priorizar quais problemas serão detalhados usando uma versão modificada da técnica 5W2H (Figura 3) e esse grupo multidisciplinar posteriormente define em conjunto ações para melhoria e solução dos problemas, caracterizando um plano de ações - que já conta com uma base de conhecimento interna para resolução de algumas questões de dados conhecidas. O acompanhamento desse plano é realizado através de quadro *Kanban* na ferramenta de gestão ágil Jira [Atlassian, 2023].



Figura 3. Detalhamento dos problemas com uma versão do 5w2h

Como piloto, esta metodologia foi aplicada em uma disciplina de E&P. A rodada da técnica levou à avaliação e melhoria, restringindo o escopo para um case de negócio específico vinculado a um projeto estratégico, com resultados mais satisfatórios e um plano de ação mais preciso - que foi validado e é acompanhado em níveis mais estratégicos. Seguindo esta mesma diretriz, foi reaplicado em outros cases de distintas disciplinas de negócio deste mesmo segmento.

Este trabalho é uma das primeiras tentativas que combinam estas técnicas numa empresa de Petróleo e Gás e tem servido de base para as iniciativas orientadas a dados do segmento em que está sendo aplicada. A dinâmica se mostrou eficiente para mapear os dados e diagnosticar as dificuldades em disciplinas e cases de negócio, além de gerenciar a expectativa dos usuários de dados - uma vez que os poucos recursos de tecnologia e do próprio negócio existentes são alocados a partir desta priorização conjunta. Como desafios futuros pretende-se integrar esta metodologia com o método ágil em escala – SAFe (*Scaled Agile Framework*), equilibrar problemas de dados em diferentes níveis (estratégicos e pontuais), lidar com dados controlados por outras áreas ou fontes externas e construir métricas de evolução da qualidade de dados.

4. Referências

- Atlassian (2023) Jira Software. www.atlassian.com/jira-software
- British Council (2023) “The Double Diamond”. <https://www.designcouncil.org.uk/our-resources/the-double-diamond/>
- Brown, T. (2008). Design Thinking. Harvard Business Review, June, 84-92.
- Côrte-Real, N. Ruivo, P. Oliveira, T. (2020) “Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value?”, Information & Management, Volume 57, Issue 1
- Kronsbein, T and Mueller, R. (2019) “Data Thinking: A Canvas for Data-Driven Ideation Workshops” In: Proceedings of the 52nd Hawaii Intl Conference on System Sciences
- Miro (2023) Plataforma de colaboração visual Miro. <https://miro.com/pt>
- 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.
- Petrobras (2024). Quem somos. <https://www.petrobras.com.br/quem-somos/perfil>
- Pham, C. Magistretti, S. and Dell’Era, C. (2022) “The role of design thinking in Big Data innovations”, Innovation, 24:2, p. 290-314.