

Núcleo de Fusão de Dados de um Gêmeo Digital da indústria de petróleo e gás

Jaqueline B. Correia, Mara Abel, Karin Becker

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91501-970 – Porto Alegre – RS – Brasil

{jbcorreia, marabel, karin.becker}@inf.ufrgs.br

Abstract. *A competitividade na indústria de óleo e gás tem exigido altos investimentos tecnológicos para decisões centradas em dados. Uma das tendências são os Gêmeos Digitais, que se valem de espaços virtuais e de serviços analíticos avançados para monitorar e aprimorar os espaços físicos. Um Núcleo de Fusão de Dados (NFD) inter-relaciona estes sistemas. A plataforma de dados OSDU é uma iniciativa de vários parceiros para eliminar silos de dados no ecossistema de petróleo e alavancar inovações através de uma abordagem orientada a dados. Neste trabalho, analisamos em que medida a plataforma de dados OSDU pode atender às necessidades de implementação de um NFD, com foco em interoperabilidade, integração, e linhagem de dados.*

1. Introdução

Para minimizar custos e aumentar a produtividade, a indústria de petróleo e gás (Oil&Gas - O&G) tem se apoiado em técnicas de inteligência artificial (IA) avançadas para a modelagem preditiva, e em Gêmeos Digitais (GD) para criar um ciclo entre o espaço físico e espaço virtual, onde diferentes cenários podem ser investigados e explorados, em um processo constante de retroalimentação. Essas tecnologias visam possibilitar decisões centradas em dados, que agreguem valor e no tempo correto (*right time*).

Um GD permite identificar oportunidades dentro de um campo petrolífero, como a redução de custos de processos de produção, aumento da produtividade de petróleo e gás, monitoramento da vida útil de equipamentos a fim de aumentar sua eficiência, realização de diferentes simulações, etc. São muitos os ganhos a partir da implementação de um GD, mas boa parte dos trabalhos publicados abordam conceitos teóricos ao invés de implementações reais, sugerindo que a implementação destes na indústria de O&G ainda encontra-se em estágios iniciais [Wanasinghe et al. 2020]. Além disso, essa indústria está ainda muito centrada em aplicações que criam e gerenciam seus dados para propósitos próprios, com escopo limitado. O resultado são silos que dificultam a troca de dados entre sistemas, sua integração e enriquecimento para investigação de cenários usando técnicas avançadas, bem como o controle da linhagem de dados que permita rastrear o caminho entre fontes de dados e decisões.

Dentre as arquiteturas propostas para GDs [Wanasinghe et al. 2020], é adotada neste trabalho a de cinco componentes explicitada na Figura 1, que evidencia a importância de um Núcleo de Fusão de Dados (NFD). Os subsistemas *Espaço Físico*, sua contrapartida *Espaço Virtual* e *Serviços* produzem dados, os quais possuem características próprias em termos de estrutura, semântica e sistemas utilizados para gerenciar seu ciclo de vida. O NFD serve como ponto de ingestão dos dados originais e de retorno em tempo correto, para dirigir o processo de otimização interativa resultante da interação entre os

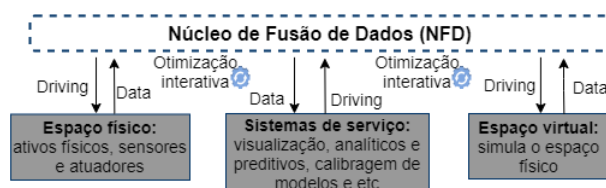


Figura 1. Arquitetura de um GD. (Adaptado de [Wanasinghe et al. 2020]).

subsistemas através de componentes de *Conexão*. Conceitos como interoperabilidade, integração, linhagem de dados e governança são fundamentais ao papel do NFD.

A questão de interoperabilidade e integração de dados têm sido abordada na indústria de O&G por diferentes consórcios, tais como a Professional Petroleum Data Management¹ (PPDM) e Energistics². Essas contribuem com padrões e melhores práticas para a gestão de dados no domínio (e.g. listas de referência, modelos de dados), ou com esquemas padronizados para troca de dados entre sistemas. Esses padrões são soluções de escopo limitado para os desafios atuais de gestão de dados, propostos de forma independente, e de difícil integração.

Como resposta a essas limitações, o consórcio The Open Group, a indústria de O&G, provedores de serviços em nuvem e provedores de sistemas de domínio vêm colaborando na construção da plataforma de dados OSDU³ (Open Subsurface Data Universe). Focada atualmente no escopo de subsuperfície *upstream* de petróleo e gás, a plataforma OSDU visa quebrar os silos de dados, evitar transferências de dados entre sistemas e viabilizar inovações (e.g. modelos preditivos), a fim de construir sistemas centrados em dados que sejam a fonte autorizada de dados (System of Record - SoR). Para tal, oferece uma arquitetura para gestão de dados nativa em nuvem, baseada em APIs padronizadas e arquitetura de microsserviços, escalável e agnóstica de tecnologia. De forma análoga a Lagos de Dados [Sawadogo and Darmont 2021], mantém os dados no seu formato nativo, oferecendo recursos para sua contextualização através de metadados, bem como recursos para governança, organização física e lógica dos dados, armazenamento e processamento de grandes volumes de dados de forma escalável. Com crescente adoção da indústria, a plataforma OSDU tornou-se o padrão *de facto* para gestão de dados neste domínio.

Neste artigo são avaliados os recursos oferecidos pela plataforma de dados OSDU como base para o NFD de um GD focado na indústria de O&G, considerando a ótica de integração, interoperabilidade, governança e linhagem de dados, e com isso, são propostos recursos complementares que permitam explorar melhor o potencial da plataforma.

O restante do texto está estruturado como segue. A Seção 2 apresenta um cenário ilustrativo no domínio e a Seção 3 sumariza a plataforma de dados OSDU. A Seção 4 discute o potencial da plataforma de dados OSDU para o NFD de um GD voltado à O&G, sugerindo recursos adicionais. A Seção 5 apresenta conclusões e trabalhos futuros.

2. Cenário Ilustrativo

Consideremos o cenário ilustrativo em que um engenheiro de reservatório necessita simular diferentes cenários de previsões para entender o comportamento do reservatório, e, ao longo do tempo, realizar atualizações para otimizar economicamente a produção. Este

¹<http://ppdm.org/>

²<https://www.energistics.org/energistics-university/>

³<https://osduforum.org/>

ajuste de histórico, envolve a exploração de dados históricos de produção (e.g. variáveis tais como pressão, temperatura e vazão de petróleo e gás) e parâmetros do modelo de reservatório (e.g. volume, fluido, permeabilidade, porosidade ou saturação). Outras informações também podem ser consideradas, como o número e localização de poços do reservatório, equipamentos de superfície, bem como métodos de recuperação de petróleo.

Simular tais cenários pode envolver técnicas complexas, diferentes serviços e sistemas (e.g. técnicas de IA, simuladores do domínio, visualização) e diferentes atores. Mas têm em comum a necessidade de explorar dados, os quais precisam ser localizados, avaliados, integrados, transformados para sua análise e exploração, possivelmente gerando novos dados. Muitos desses dados são gerenciados por aplicações específicas, no exercício de diferentes disciplinas envolvidas na indústria (e.g. geologia, produção, etc). Além disso, existem dificuldades inerentes à captação e representação dos sistemas próprios que os gerenciam. Finalmente, encontrado o ajuste adequado no histórico de produção, é necessário implantá-lo no escopo físico, acompanhando seus resultados.

Observa-se portanto, que um ajuste de histórico de produção é um processo complexo, trabalhoso e suscetível a erros, e que o escopo e disponibilidade dos dados, sua qualidade, a habilidade de integração e interoperabilidade de dados heterogêneos e de fontes distintas, bem como a habilidade de controlar a linhagem dos dados ao longo de todo o processo, são fatores que impactam diretamente nos resultados.

3. Plataforma de Dados OSDU

A plataforma de dados OSDU visa prover um núcleo básico de serviços orientados à construção de SoR orientados a dados para a indústria de O&G. Os serviços são nativos em nuvem, agnósticos quanto ao provedor específico, e devem compor aplicações segundo protocolos padrão e arquiteturas de microsserviços. Provedores de serviço em nuvem (i.e. Amazon AWS, Google), implementam esses serviços através de APIs específicas, usando seus recursos próprios para armazenamento e processamento de grandes volumes de dados de forma escalável, alto desempenho e segurança. Já os provedores de solução do domínio competem construindo soluções sobre esta plataforma, considerando arquiteturas de microsserviços. Os princípios de dados que nortearam o projeto da plataforma definem que todos os dados têm valor, e portanto devem identificáveis, mantidos em sua forma bruta e ingeridos pelo sistema com mínimo atrito possível. Além disso, devem ser sujeitos a um processo de governança mínima viável, e continuamente monitorado para controle da qualidade.

A atual *release* R3 oferece serviços na forma de APIs voltados à ingestão e armazenamento de dados, definição de esquemas de metadados, indexação, busca, controle de acesso e segurança, conformidade, notificações de eventos, e serviços utilitários (referências espaciais e conversões de unidade) [The Open Group 2020]. Na Figura 2 é resumida a visão da plataforma OSDU em relação ao ciclo de vida dos dados, e os serviços oferecidos pelo núcleo básico. O consórcio tem trabalhado na extensão destes serviços, tais como APIs voltadas à ingestão de dados específicos (e.g. dados sísmicos, *streaming*).

Os dados ingeridos são contextualizados por registros de metadados definidos por esquemas predefinidos, que seguem um conjunto restrito de tipos: *Master Data* (entidades físicas, como poços), *Reference Data* (tipos de dados ou conceitos, como tipo de documento ou unidade de medida), *Dataset* (detalhes computacionais dos datasets), *Work Product Component* e *Work Product* (contexto de negócio dos *datasets*). As organizações também podem definir seus próprios esquemas. Os esquemas/registros de metadados são

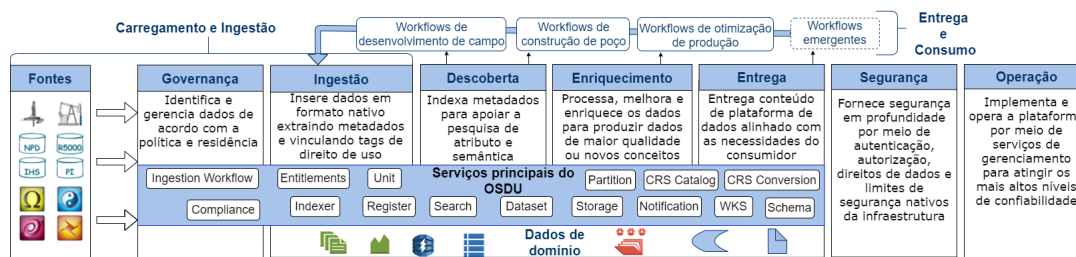


Figura 2. Plataforma de dados OSDU: Ciclo de vida dos dados e Núcleo Básico de Serviços (Adaptado de [The Open Group 2020]).

descritos em JSON, mas a proposta não contempla mecanismos explícitos para extensibilidade de tipos. Uma vez ingeridos, os registros de metadados são indexados, e podem ser localizados através de um mecanismo de busca baseado em Elastic Search.

4. Núcleo de Fusão de Dados de um GD para indústria de O&G: Discussões

4.1. Plataforma de dados OSDU: Vantagens e Limitações

Considerando o exposto na Seção 3, conclui-se que esta plataforma oferece recursos importantes para um NFD focado na indústria de O&G, mas não é uma solução completa. Considerando o cenário ilustrado na Seção 2, a plataforma OSDU proporciona um SoR que é referência de autoridade de dados para fazer o ajuste de histórico, auxiliando em romper os silos de dados deste ecossistema. Os dados gerados/consumidos pelos sensores e sistemas do espaço físico, simuladores do espaço virtual e diferentes serviços utilizados na análise, são centralizados nesta plataforma. Os serviços do núcleo cobrem o ciclo de vida da gestão de dados, da ingestão de dados heterogêneos ao consumo, e os serviços de registro/notificação permitem monitorar eventos relacionados a focos de interesse.

Em termos de metadados, a plataforma oferece um sistema de tipos simplificado, permitindo muitos graus de liberdade na definição de esquemas. Embora seja uma tendência agregar novos esquemas ao longo do tempo, é necessário hoje criá-los de forma *ad hoc*, ou esperar pelo resultado de grupos de trabalho na padronização. Considerando o cenário ilustrativo, este hoje é o caso de séries temporais para os dados de produção, pois a plataforma lida com esse tipo de dado como um arquivo qualquer, sem um esquema apropriado, apesar de sua prevalência neste contexto. Os muitos graus de liberdade e falta de padronização podem dificultar a busca de dados no futuro, e sua avaliação. Ainda, observa-se a ausência de recursos que permitam caracterizar os dados de forma semântica.

Quando os dados são ingeridos, seus metadados são indexados. Para encontrar dados de interesse, o serviço de busca permite pesquisar texto completo em campos *string*, intervalos de datas, numéricos e geoespaciais. A ausência de enriquecimento semântico, portanto, limita as buscas ao nível sintático. No cenário ilustrado, a busca de dados de interesse para as simulações e análises sobre o histórico de produção poderia ser facilitada pela disponibilização de metadados descrevendo o contexto semântico dos dados.

A plataforma está limitada a processos mínimos de governança, centrados em associar sintaticamente aos dados *Legal tags* que definem questões legais de acesso aos dados e controle de acesso. O gerenciamento e uso destas *Legal tags* é livre, e não há um padrão definido permitindo que cada usuário da plataforma decida da sua maneira a quantidade de informação que deseja associar ao dado. Para o cenário ilustrado, a presença de *Legal tags* pode não ser o suficiente para que o engenheiro compreenda as restrições envolvidas no uso e análise de dados.

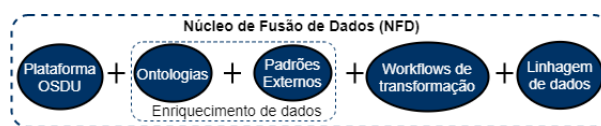


Figura 3. Núcleo de Fusão de Dados para um GD de O&G. Fonte: autoria própria.

Finalmente, não se observam serviços que explicitamente lidem com as operações de avaliação de qualidade, transformação e integração de dados, necessárias ao consumo dos dados da plataforma por recursos analíticos ou de aprendizado de máquina, e tampouco mecanismos de linhagem de dados que mantenham o registro de todo o ciclo de vida do dado, desde sua origem, transformações, até seus usos na tomada de decisões.

4.2. Recursos Complementares para o NFD

Na Figura 3 são sugeridos alguns recursos complementares voltados à gestão de dados em um NFD. Como a plataforma de dados OSDU possui pontos de extensibilidade através de suas APIs, esses recursos poderiam ser construídos sobre as mesmas, tanto na forma de camadas adicionais de serviços externos à plataforma, ou como serviços adicionais do núcleo básico OSDU na forma de *plugins*, a fim de constituir um NFD com as conexões explicitadas na Figura 1.

Enriquecimento de Dados. Enriquecer dados significa utilizar recursos para torná-los ricos em contexto e semântica visando um entendimento comum. Uma ontologia é a especificação formal e explícita de uma conceituação compartilhada [Studer et al. 1998]. Nesse sentido, ontologias existentes descrevendo diferentes aspectos do domínio de O&G podem ser usadas para agregar valor semântico aos metadados sintáticos incorporados na plataforma de dados OSDU. Padrões externos utilizados pela indústria de O&G (e.g. PPDM, Energistics, CFIHOS⁴) também podem ser aplicados para agregar valor às metadescrições. Várias funcionalidades podem ser baseadas no enriquecimento semântico, entre elas [Xiao et al. 2018, Schneider and Šimkus 2020]: a) expansão do serviço de busca a características semântica (e.g. por similaridade entre conceitos, por propriedades relacionadas); b) enriquecer o processo de transformação e linhagem de dados com metadados semânticos; c) inferências de domínio, baseadas em conhecimento prévio; etc. No cenário ilustrado na Seção 2, o engenheiro poderia reusar dados de outros ativos (e.g. em poços semelhantes); permitir correlações entre poços semelhantes a partir de alguma característica; avaliar experimentos/simulações feitos com propósitos semelhantes; entre outros.

Workflows de Transformação. As tarefas de transformação de dados são atualmente realizadas por programas ou *scripts* de transformação de dados de propósito genérico ou específico. Serviços de transformação de dados iterativa permitem a criação de um fluxo de trabalho de transformações, onde um usuário pode, a partir de um conjunto predefinido de operações, aplicar transformações sobre uma coleção de dados, assim como pode também criar diferentes conjuntos de operações de transformação de dados [Omitola et al. 2012]. O uso de serviços ou ontologias de transformação iterativa de dados trazem escalabilidade para a transformação de dados. No cenário ilustrado na Seção 2, esse recurso facilitaria o trabalho do engenheiro de reservatório para, por exemplo, selecionar *subsets* de interesse, executar automaticamente a conversão de unidades, ajustar a granularidade temporal ou integrar dados de diferentes *datasets*. Devem ser acompa-

⁴<https://www.jip36-cfihos.org/>

nhados de recursos que associem metadados ao histórico de transformação a partir de suas fontes originais, através de mecanismos de proveniência [Simmhan et al. 2005].

Linhagem de Dados. Muitas vezes usada como sinônimo de proveniência, os recursos de linhagem de dados incluem também o contexto do negócio em termos de origens e decisão. É essencial para a transparência não só da origem dos dados, mas também de todo o fluxo de transformações aos quais foram submetidos, e resultados/decisões que geraram. Através da linhagem deve ser possível acompanhar a qualidade e confiabilidade dos dados, auditar rastros dos dados, permitir a replicação de procedimentos, atribuir propriedades ou responsabilidades (e.g. *copyright*, erro), ou prover contexto informacional que pode ser consultado e analisado [Simmhan et al. 2005]. Em um sistema como o GD, em que as decisões que o retroalimentam e ajustam os diferentes espaços, físico e virtual, devem também ser objeto da linhagem[Eirinakis et al. 2020]. No cenário ilustrado na Seção 2, o engenheiro precisa manter o rastro de todas as simulações/experimentos feitos, sua relação dos dados explorados, e seus efeitos nos espaços explorados.

5. Conclusão e trabalho futuro

Neste artigo foi apresentada uma visão de um NFD de um GD focado na indústria de O&G, que considerou as vantagens e limitações de usar a plataforma de dados OSDU como base para o mesmo. Na sequência, serão propostas funcionalidades que incorporem esses recursos, e desenvolveremos uma prova de conceito que integre os mesmos aos serviços da plataforma de dados OSDU. A definição de metadados que incorporem estrutura e semântica para dados brutos externos, suas transformações e linhagem de dados, deve embasar esta proposta.

Agradecimentos: Pesquisa parcialmente apoiada por CAPES e Projeto PETWIN (financiamento FINEP e Consórcio de LIBRA)

Referências

- Eirinakis et al., P. (2020). Enhancing cognition for digital twins. In *2020 IEEE Intl. Conf. on Engineering, Technology and Innovation (ICE/ITMC)*.
- Omitola et al., T. (2012). Capturing interactive data transformation operations using provenance workflows. In *Extended Semantic Web Conference*, pages 29–42. Springer.
- Sawadogo, P. and Darmont, J. (2021). On data lake architectures and metadata management. *J. Intell. Inf. Syst.*, 56(1):97–120.
- Schneider, T. and Šimkus, M. (2020). Ontologies and Data Management: A Brief Survey. *KI - Künstliche Intelligenz*, 34(3):329–353.
- Simmhan, Y. L., Plale, B., and Gannon, D. (2005). A survey of data provenance in e-science. *SIGMOD Rec.*, 34(3):31–36.
- Studer, R., Benjamins, V. R., and Fensel, D. (1998). Knowledge engineering: principles and methods. *Data & knowledge engineering*, 25(1-2):161–197.
- The Open Group (2020). The open group guide osdu™ system concept. Technical report.
- Wanasinghe et al., T. (2020). Digital twin for the oil and gas industry: Overview, research trends, opportunities, and challenges. *IEEE Access*, 8:104175–104197.
- Xiao et al., G. (2018). Ontology-based data access: A survey. In *Proc. of the Twenty-Seventh Intl. Conf. on Artificial Intelligence, IJCAI-18*, pages 5511–5519.