

# Análise de Ferramentas de Código Aberto para Gestão de Metadados: *OpenMetadata* e *Amundsen*

**Hudson A. B. da Silva<sup>1</sup>, João V. dos Santos<sup>1</sup>, Eduardo F. R. de Sousa<sup>1</sup>,  
José E. M. Jochem<sup>1</sup>, Ronaldo dos S. Mello<sup>1</sup>, Carina F. Dorneles<sup>1</sup>, Renato Fileto<sup>1</sup>**

<sup>1</sup>Departamento de Informática e Estatística - UFSC - Florianópolis - SC - Brasil

hudson.silva@ifpa.edu.br, santosjoao301@gmail.com,  
dudfuul@gmail.com, joseeduardomj@gmail.com,  
r.mello@ufsc.br, carina.dorneles@ufsc.br, r.fileto@ufsc.br

**Resumo.** Decidir a melhor ferramenta para gestão de metadados é um desafio para equipes de projetos de dados. Este artigo apresenta uma análise de duas ferramentas de código aberto para esta finalidade, *OpenMetadata* e *Amundsen*, com enfoque no projeto *Céos* e seu escopo em dados governamentais. A metodologia *OSSpal* foi utilizada para avaliação e comparação a fim de garantir decisões assertivas e aumento na confiança da escolha de uma ferramenta. Os prós e contras de cada ferramenta no escopo do projeto *Céos* foram obtidos, com destaque para a *OpenMetadata* com score de 4.55 contra 3.75 do *Amundsen*.

## 1. Introdução

Em um mundo mais conectado e cheio de informação, empresas e projetos precisam lidar com um número expressivo de dados, e a gestão dos mesmos pode se tornar uma tarefa difícil. Neste contexto, ferramentas de gestão de metadados podem ser uma boa solução para este problema, pois organizam, gerenciam e tornam acessíveis grandes volumes de dados, além de informar sobre a linhagem dos mesmos [Karkosková 2023]. Porém, com tantas opções disponíveis, como *DataHub*<sup>1</sup>, *Apache Atlas*<sup>2</sup> e *Google Data Catalog*<sup>3</sup>, além da falta de estudos relacionados, escolher a ferramenta certa pode não ser trivial.

Diante disso, este artigo visa a apresentação de 2 ferramentas de código aberto populares para gestão de metadados, *OpenMetadata* e *Amundsen*, bem como o uso da metodologia *OSSpal* para avaliar e comparar ambas com enfoque no domínio de dados governamentais, mais especificamente no escopo do projeto *Céos*<sup>4</sup>, um projeto de pesquisa executado pela UFSC com a colaboração e financiamento do Ministério Público de Santa Catarina (MPSC). Este projeto inclui o estudo, desenvolvimento e implementação de fluxos de trabalho (*workflows*) para coleta, integração, organização, processamento e análise de dados voltados à extração de conhecimento de forma automatizada ou semiautomatizada, visando apoiar a tomada de decisão em processos do MPSC. Em razão das restrições de espaço do artigo, não são fornecidos mais detalhes sobre o projeto *Céos*.

O restante do artigo é dividido em mais 5 seções. A seção 2 aborda os trabalhos relacionados. A seção 3 descreve as ferramentas utilizadas no estudo. A seção 4 apresenta

---

<sup>1</sup><https://datahubproject.io/>

<sup>2</sup><https://atlas.apache.org/>

<sup>3</sup><https://cloud.google.com/dataplex>

<sup>4</sup><https://ceos.ufsc.br/>

a metodologia *OSSpal*. A seção 5 aplica a metodologia e realiza uma análise comparativa das ferramentas e, por fim, a seção 6 apresenta a conclusão.

## 2. Trabalhos Relacionados

Alguns poucos trabalhos na literatura abordam e compararam ferramentas de código aberto. O estudo de [Zahoor et al. 2017] apresenta 5 métodos para avaliar a maturidade de ferramentas de código aberto: *C-OSMM*, *N-OSMM*, *QSOS*, *Open BRR* e o *E-OSS*. Ele elenca os pontos fortes de cada método e propõe uma comparação entre os mesmos. Já o trabalho de [Metelo et al. 2021] compara ferramentas para suporte a atividades de ciência de dados, no caso, *KNIME*, *Dataiku*, *Orange* e *Rapid Miner*. A metodologia *OSSpal* é utilizada para a avaliação dessas ferramentas.

Apenas um trabalho que compara ferramentas para gestão de metadados foi encontrado na literatura [Rodrigues et al. 2022]. Ele analisa as ferramentas *DataHub* e *Apache Atlas* também utilizando a *OSSpal*. Diferente deste trabalho, este artigo compara 2 outras ferramentas com o mesmo propósito, sendo complementar a este trabalho.

## 3. Descrição das Ferramentas

Esta seção apresenta as ferramentas avaliadas neste estudo: *OpenMetadata* e *Amundsen*.

### 3.1. OpenMetadata

A *OpenMetadata*<sup>5</sup> é uma plataforma aberta e unificada para a descoberta, observabilidade e governança de dados alimentada por um repositório central de metadados criada pela *Collate*<sup>6</sup> e pelos fundadores do *Apache Hadoop*, *Apache Atlas* e *Uber Databook*. Ela conta mais de 2000 implementações no escopo corporativo e mais de 8000 membros de código aberto. A plataforma se baseia em 5 pilares:

- *Descoberta*: Permite que equipes encontrem os principais ativos de dados, pois possui mais de 80 conectores para tal. Habilita a pesquisa e visualização dos mesmos e suporta a colaboração com outros profissionais da área.
- *Linhagem*: Permite conhecer a origem dos dados, pois possui visibilidade de ponta a ponta, permitindo compreender o fluxo de informações pessoais sensíveis para controle e política.
- *Observabilidade*: Permite visualizar as alterações nos dados rapidamente, pois conta com visibilidade da qualidade dos esquemas, da linhagem e da governança dos mesmos. Melhora a confiabilidade das informações por meio de testes de qualidade de dados e *KPIs*. *KPIs* são indicadores quantitativos que medem os principais processos internos de uma empresa<sup>7</sup>.
- *Colaboração*: Oferta *feed* de atividades, conversas e fluxos de trabalhos que habilitam a junção das equipes, promovendo a centralização de todo o conhecimento dos dados da organização.
- *Governança*: Gerencia riscos e conformidades coletivamente, reforçando a qualidade, controle de acesso e de auditoria. Preza por fluxos de trabalho automatizados para controlar os ativos e prioriza os dados confidenciais para prover uma governança aprimorada.

---

<sup>5</sup><https://open-metadata.org/>

<sup>6</sup><https://www.getcollate.io/>

<sup>7</sup><https://www.rdstation.com/blog/marketing/kpis/>

A OpenMetadata é baseada em APIs e padrões de metadados abertos visando a gestão eficiente de ativos de dados.

### 3.2. Amundsen

*Amundsen* é uma ferramenta desenvolvida pela empresa de transporte americana *Lyft*<sup>8</sup> com objetivo principal de facilitar a descoberta, compreensão e uso eficiente de dados dentro de uma organização para melhorar a produtividade de analistas, cientistas e engenheiros de dados<sup>9</sup>. Ela se baseia em um modelo descentralizado de metadados e prioriza a experiência do usuário na busca e exploração de ativos de dados.

Além de sua interface intuitiva, a Amundsen utiliza o mecanismo de busca *Elasticsearch*<sup>10</sup> para indexação eficiente e recuperação rápida de metadados, permitindo que os usuários encontrem informações relevantes de maneira ágil e também se integra a diversas fontes de dados, incluindo bancos de dados SQL e sistemas de processamento distribuído como *Apache Hive* e *Spark*. Ela conta também com soluções de armazenamento em nuvem.

As principais funcionalidades da Amundsen são:

- *Descoberta de dados*: Permite a pesquisa e descoberta de conjuntos de dados de forma intuitiva por meio de uma busca por texto através de um algoritmo inspirado no *PageRank*<sup>11</sup>, que recomenda resultados com base em nomes, descrições, tags e padrões de uso.
- *Metadados automatizados e curadoria*: Mostra detalhes de tabelas e colunas, e inclui a exibição de estatísticas opcionais, prévia das informações, usuários frequentes e data da última atualização. Tudo isso para ajudar os usuários a interpretar, entender e confiar nos dados disponíveis além da possibilidade da integração com sistemas externos de validação e monitoramento da qualidade, como *Monte Carlo* e *Soda*, que garante que os dados atendam padrões de integridade.
- *Integração com diversas fontes de dados*: Permite a utilização de diversos conectores para bancos de dados que forneça *DPAPI* (*Data Protection API*) ou *SQLAlchemy*. Possui também soluções de armazenamento em nuvem que colaboram com a centralização dos metadados.
- *Linhagem de dados*: Fornece visibilidade de ponta a ponta, permitindo visualizar a origem e o destino dos dados dentro de um pipeline. Possui integração com ferramentas de ETL para rastreamento e transformação dos dados, bem como acesso e atualização de informações de linhagem através de APIs REST.
- *Governança*: Adota uma abordagem descentralizada para governança que permite usuários colaborarem na atualização de metadados, adição de descrições, anotações, advertências e comentários sobre um conjunto de dados. Embora não gerencie permissões diretamente, pode ser integrado a sistemas de controle de acesso como o *LDAP* (*Lightweight Directory Access Protocol*) e *Auth* (*Autenticação e Autorização*) para restringir o acesso aos metadados.

---

<sup>8</sup><https://www.lyft.com/>

<sup>9</sup><https://www.amundsen.io/amundsen/>

<sup>10</sup><https://www.elastic.co/elasticsearch>

<sup>11</sup><https://pt.semrush.com/blog/o-que-e-pagerank/>

#### 4. Metodologia de Comparação: OSSpal

A metodologia *OSSpal* foi a escolhida para a avaliação das ferramentas neste trabalho por garantir melhores decisões e aumentar a confiança em softwares de código aberto. Além disso, é muito utilizada em estudos de comparação de ferramentas, como visto na Seção 2. *OSSpal* significa *Open Source Software pal*, sendo “*pal*” um tributo a Murugan Pal, um dos seus fundadores [Wasserman et al. 2017]. Ela possui 4 fases [Metelo et al. 2021]:

- *Fase 1*: Identifica as características a serem avaliadas;
- *Fase 2*: Atribui pesos às categorias de avaliação definidas na Fase 1;
- *Fase 3*: Atribui ponderação às categorias, de 1 (muito ruim) até 5 (muito bom);
- *Fase 4*: Calcula a pontuação final da avaliação da ferramenta.

A partir de documentos padrão de processos de avaliação, como ISO/IEC 9126 and ISO/IEC 25010:2011, a metodologia *OSSpal* também destaca 7 categorias importantes para o processo de avaliação de código aberto [Wasserman et al. 2017]:

- *Funcionalidade*: O quanto bem o software vai satisfazer os requisitos do usuário?
- *Características Operacionais de Software*: O software é seguro? O software performa bem? O software escala em grandes ambientes? A interface de usuário é boa? O uso do software para usuários finais é fácil? É fácil instalar, configurar, implementar e manter o software?
- *Suporte e Serviço*: O software tem um bom suporte? Existe suporte comercial e da comunidade? A organização provê serviços de formação e consultoria?
- *Documentação*: O software possui bons tutoriais e uma boa documentação?
- *Atributos Tecnológicos de Software*: A arquitetura do projeto é sólida? O software é modular, portátil, flexível, expansível e fácil de integrar? O design, código e testes do software são bons? O software é completo e livre de erros?
- *Comunidade e Adoção*: O software é adotado pela comunidade e indústria? O software tem uma comunidade de usuários ativa?
- *Processo de Desenvolvimento*: Qual o nível de profissionalismo do processo de desenvolvimento e da organização do projeto em geral?

A categoria *Funcionalidade* permite que o avaliador elenque funções que ele julgue importantes para uma ferramenta, no qual é dado um peso de 3 (muito importante) até 1 (pouco importante), ou caso a ferramenta não suporte a funcionalidade recebe pontuação 0. Após isso, deve-se classificar as avaliações em uma soma cumulativa que passa da escala de 1 até 3 para a escala de 1 até 5 [Metelo et al. 2021], ou seja:

- Menos de 65% = 1 (Inaceitável).
- Entre 65 e 80% = 2 (Pobre).
- Entre 80 e 90% = 3 (Aceitável).
- Entre 90 e 96% = 4 (Bom).
- Maior que 96% = 5 (Excelente) .

#### 5. Análise Comparativa

Esta seção apresenta a análise comparativa da *OpenMetadata* e da *Amundsen* com o uso da metodologia *OSSpal*. Para tal, as ferramentas foram avaliadas dentro do escopo da catalogação de metadados de fontes de dados de domínio governamental, mais especificamente relacionado ao projeto Céos.

Na metodologia adotada, cada categoria recebe um peso consoante sua importância na análise comparativa. A Tabela 1 mostra os valores definidos e julgados mais adequados ao escopo de uso das ferramentas. Esses valores são justificados a seguir.

**Tabela 1. Atribuição dos pesos às categorias — Fonte: Do autor.**

CATEGORIAS	PESO
Funcionalidades	35%
Documentação	20%
Características Operacionais do Software	15%
Suporte e Serviço	10%
Comunidade e Adoção	10%
Atributos Tecnológicos de Software	5%
Processo de Desenvolvimento	5%
<b>Total</b>	<b>100%</b>

Consideramos que as *funcionalidades* oferecidas pela ferramenta é o fator de mais relevância, com um peso de 35%, visto que o principal intuito de uso da ferramenta é suprir as necessidades dos profissionais em termos de gestão e acesso de/a metadados governamentais.

Em segundo lugar surge a *documentação*, com 20% do peso total. Esse valor se justifica pelo fato do projeto possuir muitas linhas de pesquisa e algumas delas utilizarem de forma ativa a ferramenta. Assim, se faz necessária uma documentação detalhada e bem organizada.

As *características operacionais do software*, com peso de 15%, se justifica por se tratar de um projeto de porte nacional que lida com um volume grande de dados e vários usuários. Na sequência, com 10% cada, temos *suporte e serviço* junto com *comunidade e adoção*, justificados pela necessidade de uma comunidade ativa e parceira para solucionar possíveis problemas e dúvidas ao longo do uso da ferramenta, assim como da vontade que o projeto esteja utilizando os mesmos padrões e ferramentas que outras empresas/projetos já usam. Por fim, temos *atributos tecnológicos do software* e *processo de desenvolvimento*, sendo as categorias consideradas menos importantes para o escopo do projeto e com peso de 5% cada uma.

Nas *funcionalidades* das ferramentas é aberto um leque de possibilidades para a escolha daquelas a serem levadas em consideração e comparadas. Com intuito de se aproximar o máximo possível de uma análise que mais satisfaz a necessidade de gestão de metadados no escopo do projeto, foram escolhidos alguns critérios que integram a parte da funcionalidade.

Primeiramente, o *acesso e pesquisa de conjuntos de dados* é imprescindível para o bom funcionamento das ferramentas independente do contexto que são aplicadas. Ao ir numa biblioteca com intuito de encontrar o livro que deseja, você usa um catálogo para descobrir se o livro se encontra, qual sua edição, uma breve descrição e onde está localizado. Ferramentas de gestão de metadados devem funcionar analogamente, porém em um ecossistema de dados. Um catálogo deve ter opções flexíveis de pesquisa e filtragem para

permitir que os usuários encontrem rapidamente conjuntos relevantes de metadados<sup>12</sup>.

[Papedesu et al. 2021] afirma que o acesso mais fácil aos dados é apenas uma parte do quebra-cabeça e que o ponto crucial que se deve ter em mente é que as informações somente são valiosas se forem confiáveis. Caso contrário haverá receio em tomar decisões levando-as em consideração. Logo, além de encontrar dados na organização é primordial para que os usuários consigam avaliar sua confiabilidade. Isso também inclui a capacidade descritiva no geral, como exibição de estatísticas, histórico de acessos e outras métricas. Assim sendo, a *avaliação e qualidade dos dados* foi também escolhida como uma funcionalidade necessária.

Outra escolha foi a *linhagem de dados*. A rastreabilidade, ou seja, de onde vêm os dados e como eles são transformados e processados, é essencial para a auditoria e confiabilidade. Uma boa plataforma de gestão e catalogação de metadados deve oferecer um sistema de linhagem que permita mapear e visualizar o fluxo dos dados e de suas informações, desde suas origens até o destino final.

Um ponto essencial a ser avaliado nas ferramentas de gestão de metadados é a quantidade de informações sobre os dados que é possível ser armazenada e descrita na aplicação. No escopo do projeto Céos foram mensurados 31 metadados de suma importância para a interoperabilidade, filtragem e caracterização dos dados, bem como a distinção dos mesmos em 2 níveis de granularidade [da Silva et al. 2024]. Com isso, a *capacidade descritiva* da ferramenta é também uma característica chave a ser avaliada.

A possibilidade de gerenciar o acesso e controlar as permissões dos usuários em relação aos ativos de dados também é um ponto relevante. Com as diversas linhas de pesquisa atuando no projeto, onde cada uma trabalha, muitas vezes, não com todos, mas com parte dos conjuntos de dados e considerando que alguns deles têm certo grau de sigilo, torna-se essencial que a ferramenta incorpore opções de *governança e controle de acesso* às informações embutidas em suas funcionalidades.

Ferramentas de gestão de metadados não devem apenas armazenar metadados, mas também monitorar mudanças e alertar os usuários sobre essas modificações. Assim sendo, o critério *monitoramento e observabilidade* também foi considerado.

A *gratuidade* também é um fator digno de atenção pois, apesar do projeto ser financiado pelo MPSC, ele é executado por membros da UFSC, sendo a pesquisa um dos pilares da instituição. Portanto, a partilha gratuita de resultados com o meio acadêmico advindos das pesquisas relacionadas ao projeto, bem como a possibilidade de que os estudos possam ser replicados por outras entidades de ensino, ou até mesmo servir de base científica para outras pesquisas, justifica a necessidade de avaliar esse tópico.

Por fim, é importante que as ferramentas também permitam que seus usuários interajam entre si, compartilhem conhecimento, adicionem anotações e documentem informações sobre os ativos de dados. Com isso, alguns recursos como a possibilidade de anexar *tags* às informações, *feeds* de atividade e integração com sistemas de mensagens podem aumentar a eficácia da ferramenta. Deste modo, a *comunicação interna* é também uma característica considerada.

A Tabela 2 pontua cada uma das características em 3 (muito importante) até 1

---

<sup>12</sup><https://www.oracle.com/ca-en/big-data/data-catalog/what-is-a-data-catalog/>

(pouco importante) e também apresenta a avaliação de cada ferramenta<sup>13</sup> dentro do escopo das funcionalidades com o uso da metodologia OSSpal. Já a Tabela 3 apresenta as pontuações de cada ferramenta, levando em conta o uso das métricas utilizadas pela OSSpal, bem como as funcionalidades identificadas como importantes para ferramentas de gestão de metadados no escopo do projeto Céos.

**Tabela 2. Pontuação e avaliação das funcionalidades — Fonte: Do autor.**

FUNCIONALIDADES	PONTUAÇÃO	OMD	AMUNDSEN
Acesso e Pesquisa de Conjuntos de Dados	3	3	3
Avaliação e Qualidade dos Dados	3	3	3
Linhagem de Dados	3	3	2
Capacidade Descritiva	3	3	2
Governança e Controle de Acesso	2	2	2
Monitoramento e Observabilidade	2	2	1
Gratuidade	2	2	2
Comunicação	1	1	1
<b>Soma Cumulativa</b>	<b>19</b>	<b>19</b>	<b>17</b>
<b>Pontuação</b>	<b>100%</b>	<b>100%</b>	<b>90%</b>
	<b>5</b>		<b>4</b>

**Tabela 3. Avaliação geral das categorias — Fonte: Do autor.**

CATEGORIAS	OMD	AMUNDSEN
Funcionalidades	5	4
Documentação	4	5
Características Operacionais do Software	3	3
Suporte e Serviço	5	2
Comunidade e Adoção	4	3
Atributos Tecnológicos de Software	4	3
Processo de Desenvolvimento	5	2

Na sequência, uma discussão sobre a avaliação de cada ferramenta é apresentada.

### 5.1. OpenMetadata

As funcionalidades do OpenMetadata receberam nota 5, conforme visto na Tabela 2. A capacidade descritiva é um diferencial da ferramenta, que conta com um sistema de *tags* e glossários, além de possuir um sistema de governança e controle de acesso embutidos, onde todas essas informações podem ser usadas como filtro na sua área de pesquisa. Ela possui uma curadoria automática dos dados, além de ser possível adicionar novos outros processos de qualidade. No quesito linhagem de dados a ferramenta conta com 8 conectores de *pipeline*: *Airbyte*, *Airflow*, *Dagster*, *DBT*, *Fivetran*, *Flink*, *NiFi* e *Spline*. A

<sup>13</sup>A ferramenta OpenMetadata é identificada por OMD.

ferramenta possui um *feed* que mostra todas as mudanças feitas nos ativos de dados da organização e conta com um sistema de comunicação completo, com a possibilidade de atribuir tarefas aos usuários.

Com relação às demais categorias, temos primeiramente a documentação. Em relação a documentação, a OpenMetadata é bastante completa dado o número de funcionalidades que a ferramenta proporciona, com tutoriais de instalação íntegros e documentação para cada conector de dados suportado. Porém, apesar de contar com uma barra de pesquisa, ela peca um pouco na forma como é organizada, dificultando a procura por algum tema específico, o que gerou uma nota 4 na sua avaliação.

O software da ferramenta pode ser implementado usando *Docker*, o que facilita sua instalação e implementação. Ele se mostra seguro, pois possui um esquema de *hash* em seu banco de dados interno, e também conta com uma bonita e fluida interface com o usuário. Porém, se submetido a um volume grande de processamento de dados, ele apresenta lentidão, talvez por possuir uma curadoria de dados automática, fato esse que prejudica o seu desempenho, e justifica a nota 3 para as características operacionais de software.

Supporte e serviço se mostra um dos pontos fortes da ferramenta, o que lhe fez receber nota 5. A mesma conta com uma vasta comunidade com mais de 6000 contribuidores no *Github*, além de promover o acesso à plataforma *Slack* para dúvidas e problemas. Ela conta com um *blog* que publica conteúdos relevantes e um canal no *YouTube* com tutoriais, gravação dos encontros mensais onlines com os membros de comunidade e *Webinars*.

A OpenMetadada possui um canal no *Slack* com mais de 8000 membros, com constantes contribuições dos membros, tanto da comunidade quanto do suporte da ferramenta. Também conta com uma boa adesão industrial, com empresas como Loggi, in-Drive, PayU Finance, Moove, dentre outras, que a adotam. Desta forma, ela foi avaliada com a nota 4 na categoria Comunidade e Adoção.

Conforme salientado anteriormente, sua implementação pode ser feita através do *Docker*, sendo um ponto positivo. A mesma utiliza o *Elasticsearch* como mecanismo de busca e permite a escolha tanto do PostgreSQL quanto MySQL como banco de dados interno. Ela possui seu próprio *Airflow*, que executa fluxos de trabalho de ingestão que foram implantados por meio da interface de usuário, e fornece suporte à API REST para obter dados armazenados na aplicação. Considerou-se, assim, a ferramenta como sendo robusta e foi dada uma nota 4 no quesito atributos tecnológicos de software.

Por fim, a ferramenta se mostra sólida, com um *site* bem organizado e uma preocupação com o seu profissionalismo. Ela apresenta um *roadmap* atualizado, constantes atualizações e uma aba com as mudanças e problemas resolvidos, o que justificou a nota 5 em processo de desenvolvimento.

## 5.2. Amundsen

O Amundsen no quesito funcionalidades teve a nota 4 atribuída em função dos resultados da Tabela 2, pecando um pouco na parte de capacidade descritiva e monitoramento dos dados. Já a documentação é bem completa, dividida em guia do usuário e guia do desenvolvedor, com tutoriais de instalação e configuração. Além disso, também explica como funciona cada componente da ferramenta, por isso recebeu a nota 5.

Apesar do suporte ao sistema de gerenciamento de permissões, a ferramenta não possui um próprio, exigindo que a organização faça essa gerência. Ainda, um ponto a ser analisado em relação às características operacionais do software é a escalabilidade. Para grandes volumes de dados, ela não se oferece um sistema de balanceamento de carga ou particionamento nativo, o que pode exigir ajustes para otimizar o desempenho em ambientes distribuídos. Sua instalação só é facilitada pela possibilidade de usar *contêineres*, reduzindo a complexidade da configuração manual, mas ainda é preciso configurar múltiplos serviços para o seu bom funcionamento, justificando assim a nota 3 na categoria.

Em suporte e serviço, percebe-se que a falta de uma empresa por trás do desenvolvimento é negativo no longo prazo. Apesar de ter sido criada pela Lyft, não há uma empresa dedicada ao suporte contínuo e, embora tenha um Slack para um suporte comunitário, os usuários são poucos engajados e ficam sem trocas de mensagens por meses. Ainda, a ferramenta não recebe uma atualização significativa desde 2021. As pequenas atualizações que são feitas no repositório oficial do GitHub demoram meses entre uma e outra. Em função disso, a pontuação da categoria suporte e serviço ficou em 2.

Várias empresas adotaram o Amundsen, entre elas PicPay, Snapchat e PUBGf. No entanto, o que deixa a nota de comunidade e adoção em 3 é a falta de uma comunidade maior e mais interativa, pois a atividade nos canais de comunicação é baixa.

Quanto aos atributos tecnológicos de software, verificamos que ela tem uma arquitetura modular que traz certa facilidade para seu uso. O atributo mais importante é a facilidade na integração, pois é possível se conectar com outras tecnologias sem a necessidade de reescrita no código base. Entretanto, a ferramenta apresenta uma flexibilidade limitada, já que várias integrações exigem alterações no *Databuilder*. Logo, a nota atribuída a essa categoria foi 3 pelo código-fonte não ser tão modular e expansível.

Em relação ao processo de desenvolvimento, a plataforma é mantida por contribuições da comunidade e não possui uma organização oficial para seu desenvolvimento, o que acarreta em um processo lento de atualização. Até mesmo o *roadmap* presente no *site* oficial está desatualizado desde 2021. Assim sendo, a categoria foi avaliada com nota 2.

Com isso, os resultados finais de cada ferramenta, assim como os cálculos com base em pontuação e peso de cada categoria, estão sumarizados na Tabela 4.

**Tabela 4. Pontuação final das ferramentas — Fonte: Do autor.**

CATEGORIAS	OMD	AMUNDSEN
Funcionalidades	$5 \times 0.35 = 1.75$	$4 \times 0.35 = 1.4$
Documentação	$4 \times 0.20 = 0.8$	$5 \times 0.20 = 1.0$
Características Operacionais do Software	$3 \times 0.15 = 0.45$	$3 \times 0.15 = 0.45$
Suporte e Serviço	$5 \times 0.10 = 0.5$	$2 \times 0.10 = 0.2$
Comunidade e Adoção	$4 \times 0.10 = 0.4$	$3 \times 0.10 = 0.3$
Atributos Tecnológicos de Software	$4 \times 0.10 = 0.4$	$3 \times 0.10 = 0.3$
Processo de Desenvolvimento	$5 \times 0.05 = 0.25$	$2 \times 0.05 = 0.1$
<b>Pontuação Final</b>	<b>4.55</b>	<b>3.75</b>

## 6. Conclusão

Este trabalho apresenta uma avaliação de 2 ferramentas de gestão de metadados, Open-Metadata e Amundsen, com o uso da metodologia OSSpal e com foco na utilização das mesmas no escopo do projeto Céos. Com base na metodologia, a OpenMetadata obteve um score final de 4.55 e a ferramenta Amundsen um score final de 3.75.

Os resultados mostram que a OpenMetadata destacou-se em aspectos como suporte, serviço, comunidade e processo de desenvolvimento. O que contribui para isso é o fato de não ter uma equipe oficial destinada ao desenvolvimento e manutenção da Amundsen. As empresas que adotaram a plataforma possuem equipes internas para mitigar os problemas e aprimorá-lo para se adequar ao quadro da companhia. Em contrapartida, a Amundsen mostrou ter uma melhor documentação se comparado com o OpenMetadata.

Como trabalhos futuros, sugere-se que sejam realizados testes práticos em ambientes de produção para uma avaliação de experiência e desempenho real de cada ferramenta e como elas se comportam não só no escopo do projeto Céos, mas também em outros ecossistemas de dados.

## Referências

- da Silva, H. A. B., Santos, J. V. d., Jochem, J. E. M., Fleck, A., Mello, R. d. S., Dorneles, C. F., and Fileto, R. (2024). Uma Proposta Baseada no Dublin Core para Catalogação de Metadados de Fontes de Dados Governamentais. In *XXXIX Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 876–882. Sociedade Brasileira de Computação.
- Karkosková, S. (2023). Architectural Design of Metadata Management Tool. In *IX Int. Conf. on Control, Decision and Information Technologies (CoDIT)*, pages 65–70. IEEE.
- Metelo, M., Bernardino, J., and Pedrosa, I. (2021). Avaliação de Ferramentas Open Source para Data Science usando a Metodologia OSSpal. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E46):588–606.
- Papudesu, C., of Product Management, Catalog, and Collibra, L. a. (2021). How a Data Catalog Can Help Your Business Reach New Heights. <https://www.dbta.com/Editorial/Trends-and-Applications/How-a-Data-Catalog-Can-Help-Your-Business-Reach-New-Heights-144574.aspx>. Accessed: 2025-3-9.
- Rodrigues, D., Almeida, M., Guimarães, P., and Santos, M. Y. (2022). DataHub and Apache Atlas: A Comparative Analysis of Data Catalog Tools. In *XXII Conferência da Associação Portuguesa de Sistemas de Informação (CAPSI)*, pages 90–109. Associação Portuguesa de Sistemas de Informação.
- Wasserman, A. I., Guo, X., McMillian, B., Qian, K., Wei, M.-Y., and Xu, Q. (2017). OSSpal: Finding and Evaluating Open Source Software. In *Open Source Systems: Towards Robust Practices: 13th IFIP WG 2.13 International Conference, OSS, Proceedings 13*, pages 193–203. Springer International Publishing.
- Zahoor, A., Mehboob, K., Natha, S., et al. (2017). Comparison of Open Source Maturity Models. *Procedia Computer Science*, 111:348–354.