

Open VLF: Plataforma de Dados Abertos para *Very Low Frequency Data*

Daniel H. V. Kauffmann, Lucas S. Garcia, Rogério De Oliveira

Faculdade de Computação e Informática
Universidade Presbiteriana Mackenzie (UPM) – São Paulo, SP – Brasil

daniel.kauffmann@mackenzista.com.br,
lucassantiago.garcial@mackenzista.com.br,
rogerio.oliveira@mackenzie.br

Abstract. *This work implements an open data platform for VLF (Very Low Frequency) signals. The platform integrates multiple data repositories and offers functionalities for searching, visualizing, and downloading data collected from the SAVNET and AWESOME databases, including open-source tools and informative content. With a modular and scalable approach, the democratization of information for studies on space weather, solar flares, and atmospheric and electromagnetic phenomena of the ionosphere positively impacts research and teaching in the field. Furthermore, the structure of the platform allows for adaptation to other types of data and similar systems, reinforcing its expandable and flexible character within the practice of open science, promoting transparency and collaboration in the scientific community.*

Resumo. *Este trabalho implementa uma plataforma de dados abertos para sinais VLF (Very Low Frequency). A plataforma integra múltiplos repositórios de dados e oferece funcionalidades para busca, visualização e download de dados coletados das bases de dados SAVNET e AWESOME, incluindo ferramentas de código aberto e conteúdo informativo. Com uma abordagem modular e escalável, a democratização das informações para estudos sobre clima espacial, erupções solares e fenômenos atmosféricos e eletromagnéticos da ionosfera impactam positivamente a pesquisa e o ensino na área. Além disso, a estrutura da plataforma permite a adaptação para outros tipos de dados e sistemas similares, reforçando seu caráter expansível e flexível dentro prática da ciência aberta, promovendo a transparência e colaboração na comunidade científica.*

1. Introdução

No contexto atual em que dados e a ciência aberta assumem um papel central na pesquisa científica, a relevância de projetos que facilitam o acesso e a análise de dados complexos é indiscutível. A era digital trouxe um crescimento exponencial na quantidade de dados disponíveis, destacando a necessidade de plataformas eficientes para o gerenciamento e interpretação desses dados. Neste cenário, iniciativas que buscam democratizar o acesso à informação científica, permitindo a verificação e reprodução de resultados de pesquisa, se tornam cada vez mais fundamentais. Este projeto se alinha a essa tendência, visando contribuir para a comunidade científica através da facilitação do acesso e análise de dados científicos.

Este trabalho implementa uma plataforma integrada para o gerenciamento e consulta de dados VLF (*Very Low Frequency*). A escolha dos dados VLF não por acaso, mas sim reflexo de sua importância crescente na pesquisa científica, especialmente no estudo de fenômenos atmosféricos e ionosféricos. Os sinais VLF têm sido uma fonte valiosa de informações para a compreensão de diversos fenômenos naturais, desde padrões climáticos até atividades solares. A plataforma desenvolvida visa tornar esses dados mais acessíveis e manejáveis para pesquisadores, estudantes e entusiastas.

Os dados VLF são amplamente utilizados para a pesquisa em clima espacial e estudos ionosféricos. No entanto, o desafio reside no fato de que esses dados são muitas vezes complexos e volumosos, tornando seu acesso e análise uma tarefa desafiadora. O projeto enfrenta esse problema, buscando simplificar a maneira como esses dados são acessados e analisados, permitindo que pesquisadores e interessados possam se concentrar mais no estudo dos fenômenos do que nas dificuldades técnicas de manuseio dos dados.

Este projeto, portanto, fornece uma solução importante e relevante para o desafio de acessar e analisar dados VLF, e que tem potencial na democratização da maneira como a comunidade científica interage com esses dados. Ao tornar os dados mais acessíveis e manejáveis, a plataforma não só facilita a pesquisa científica, mas também apoia os princípios da ciência aberta, promovendo a transparência e colaboração dentro da comunidade acadêmica e científica.

2. Referencial Teórico

Para este estudo são particularmente importantes conceitos sobre ciência e dados abertos, que motivam este projeto, e sobre VLF, e suas aplicações no campo da pesquisa. Esses conceitos são brevemente desenvolvidos aqui.

2.1. Ciência e Dados Abertos

A ciência aberta visa tornar a pesquisa financiada publicamente mais acessível [OECD 2015], envolvendo iniciativas de publicação de trabalhos científicos, uso de dados abertos e infraestruturas de suporte [Unesco 2017]. Segundo a Comissão Europeia [EC 2021], a ciência aberta é uma abordagem cooperativa que utiliza tecnologias digitais e ferramentas colaborativas. Foster (n.d.) a define como uma prática que permite a colaboração e contribuição com acesso gratuito aos dados de pesquisa e processos. Vicente-Saez e

Martinez-Fuentes (2018) a veem como um conhecimento transparente e acessível desenvolvido através de uma rede colaborativa.

Dados abertos são essenciais para a ciência aberta [Fapesp, n.d.; Unesco 2021], sendo definidos pelo Open Data Handbook (n.d.) como dados livremente utilizáveis e redistribuíveis. Eles são importantes para vários setores, incluindo o cumprimento dos Objetivos de Desenvolvimento Sustentável (ODS) [Davies et al. 2019]. A Unesco (2021) os define como dados de pesquisa abertos disponíveis em formato legível e regidos pelos princípios FAIR. A OCDE (2007) recomenda que o acesso a dados abertos seja preferencialmente baseado na Internet.

Espera-se que a ciência e os dados abertos acelerem a pesquisa científica, aumentem a confiabilidade e qualidade das pesquisas e tornem a ciência mais sensível às necessidades públicas [OCDE 2016]. A disponibilização de dados abertos é uma tradição nas ciências geofísicas e astrofísicas, com o acesso aberto a dados científicos estabelecido já em 1957-1958 [NAP 2008].

Grupos de pesquisa, como Cohen e Golkowski [Cohen e Golkowski n.d.; Cohen 2020; Cohen et al. 2019], têm disponibilizado dados abertos de antenas VLF. Nosso projeto se assemelha ao projeto WALDO, mas se diferencia pela não vinculação dos dados a um repositório próprio, pela inclusão de dados do sistema SAVNET e pelo armazenamento temporário de visualizações de dados, proporcionando maior flexibilidade e economia de armazenamento.

3. Metodologia

Nesta seção, descreveremos a metodologia empregada no desenvolvimento da plataforma para consulta dos dados VLF gerados. O objetivo principal deste trabalho é permitir que pesquisadores e entusiastas de todo o mundo acessem e utilizem dados científicos VLF de forma acessível e eficiente. Para alcançar esse objetivo, empregamos uma abordagem de desenvolvimento flexível, escalável e agnóstica para qualquer serviço de nuvem, a fim de garantir a reutilização da infraestrutura em diversos ambientes de armazenamento e hospedagem de API.

3.1. Arquitetura

A arquitetura do sistema é composta por quatro agrupamentos de componentes interconectados que permitem o funcionamento da Plataforma *Open VLF* (Figura 1). O sistema é projetado para ser flexível, escalável e agnóstico, garantindo a adaptabilidade a diferentes ambientes de hospedagem e armazenamento. Veremos com mais detalhe cada um dos agrupamentos nas subseções seguintes.

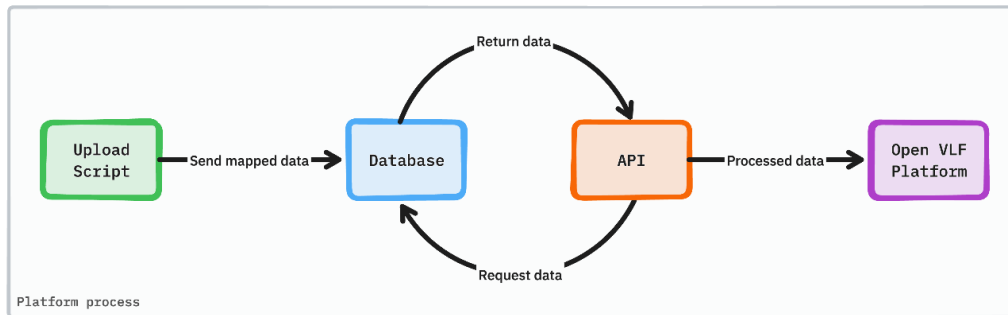


Figura 1. Arquitetura da plataforma em alto nível (componentes agrupados)

3.2. Frontend

O frontend da plataforma é um website construído utilizando o framework *VueJS* e hospedado no *Firebase Hosting* (Figura 2). Isso permite que os usuários acessem facilmente os dados VLF por meio de uma interface amigável e responsiva. A integração com a API e o armazenamento no *Firebase Hosting* garante que o website funcione de maneira eficaz e forneça uma experiência de usuário sem interrupções. Estes dois componentes constituem o *Open VLF Platform*.

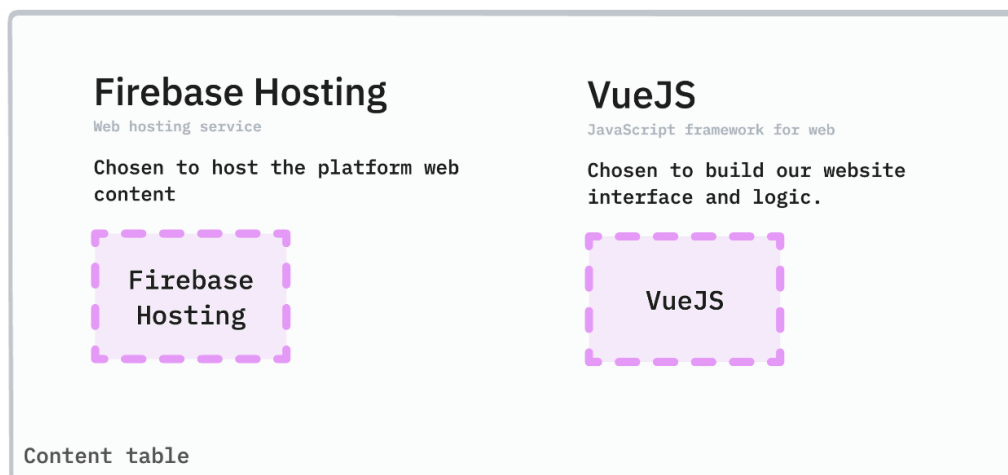


Figura 2. Componentes *Firebase Hosting* e *VueJS*

3.3. Aquisição de dados e backend

Um script Python é empregado para coleta e armazenamento dos dados (Figura 3). Esse script é projetado para ser executado em um computador que possui os arquivos VLF a serem processados. Sua principal função é mapear os arquivos e enviá-los para um repositório na *Amazon Web Services* (AWS), especificamente o *Amazon S3*. A modularidade do script permite que suas funções sejam facilmente adaptadas a mudanças na infraestrutura, garantindo sua longevidade e flexibilidade.

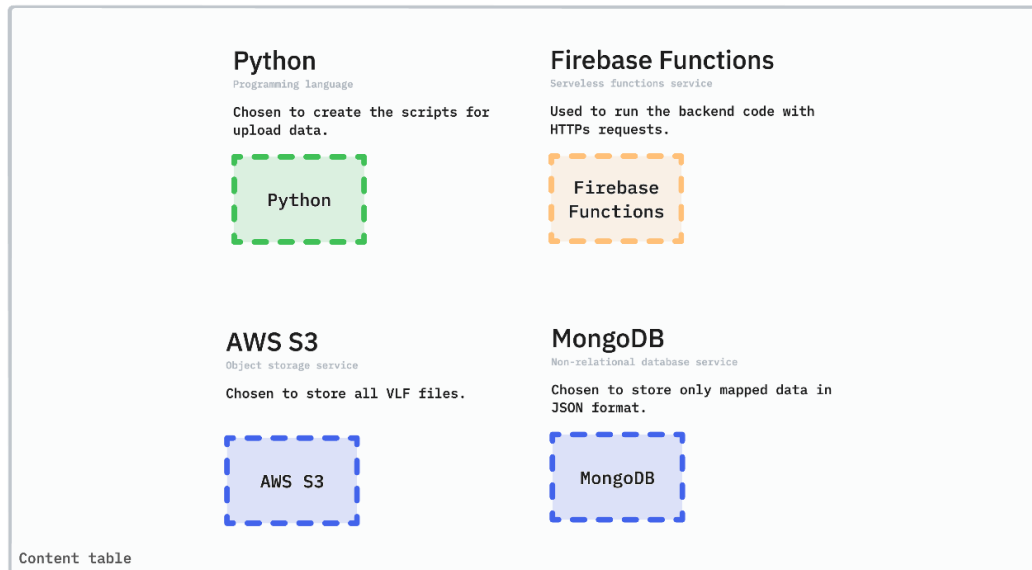


Figura 3. Componentes Python, Firebase Functions, AWS S3 e MongoDB

O script realiza duas tarefas principais: o armazenamento dos arquivos VLF no *Amazon S3* e a criação de um mapeamento em formato *JavaScript Object Notation* (JSON) contendo informações detalhadas sobre os arquivos, como a data de geração, a estação de origem, nome do arquivo, entre outros. Esse mapeamento é posteriormente armazenado no *MongoDB*, um banco de dados não relacional. Essa abordagem de separar o armazenamento de arquivos pesados no *S3* do armazenamento de metadados em um banco de dados mais leve otimiza a eficiência do sistema.

A API (*Application Programming Interface*) consome os dados do banco de dados de forma eficiente e oferece suporte ao nosso website, tornando a consulta e disponibilização dos dados VLF para os usuários rápidas e eficazes. A API também foi projetada para ser agnóstica, o que significa que a lógica dos *endpoints* da API podem ser facilmente utilizadas em diferentes ambientes e formatos de hospedagem, mantendo sua portabilidade e reusabilidade.

3.3.1. Formato dos Dados

Os dados VLF são coletados em dois formatos principais: arquivos MATLAB, oriundos do *Atmospheric Weather Educational System for Observation and Modeling of Electromagnetics* (AWESOME), e arquivos FITS, fornecidos pela *The South America VLF Network* (SAVNET). Os arquivos MATLAB são comuns nas engenharias e ciências, utilizados para armazenar informações como amplitudes e frequências de ondas eletromagnéticas de baixa frequência, oferecendo aos pesquisadores a flexibilidade de aplicar funções e algoritmos específicos para análises personalizadas. Esses arquivos são vantajosos para análises avançadas adaptadas às necessidades específicas da pesquisa em ondas VLF.

Em contrapartida, os arquivos FITS, amplamente usados na astronomia e astrofísica, também encontram aplicação nos estudos VLF. Eles são conhecidos por armazenarem dados detalhados, incluindo metadados importantes para documentar

observações científicas. Nos estudos VLF, esses arquivos podem incluir medidas de campo magnético e espectros de frequência. A estrutura hierárquica dos arquivos FITS, com cabeçalhos detalhados e blocos de dados, os torna ideais para registrar informações cruciais sobre a origem, unidades de medida e datas das observações, sendo valiosos para documentação detalhada dos dados.

4. Resultados

O resultado deste trabalho é uma plataforma online que permite a consulta dos dados VLF de forma acessível e eficaz. A interface foi modelada visando a acessibilidade para qualquer usuário interessado, de forma amigável e intuitiva para buscar, navegar e explorar dados VLF. A figura 4 mostra uma captura de tela da interface da plataforma, destacando as principais seções, filtros para busca e resultados.

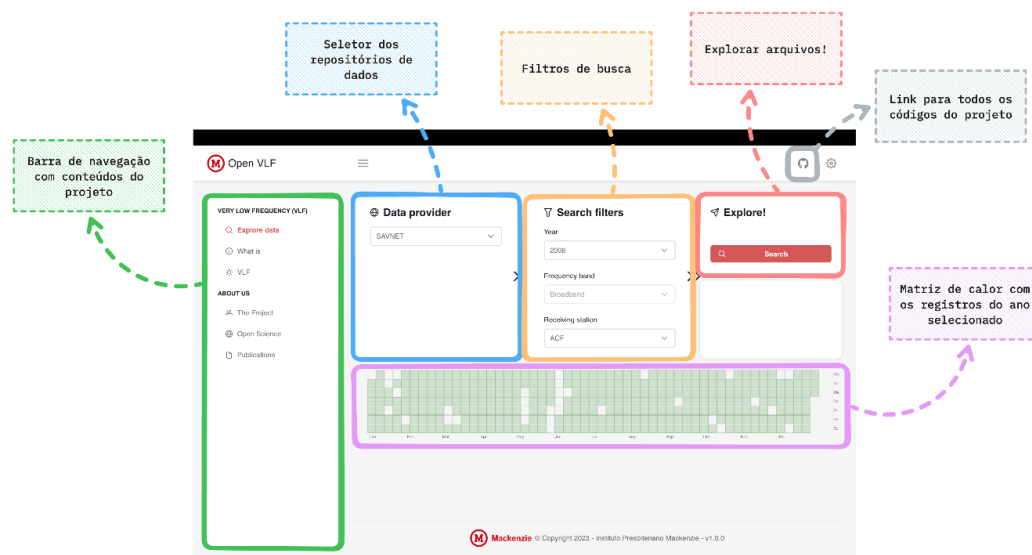


Figura 4. Interface da plataforma *Open VLF* (parte A)

Os usuários têm a capacidade de selecionar e baixar arquivos VLF de interesse diretamente pela plataforma (Figura 5). No entanto, como parte dos esforços para manter a integridade dos dados e identificar os usuários que fazem uso dessas informações, é solicitado que o nome e o endereço de e-mail sejam fornecidos ao realizar o download. Essa abordagem não apenas ajuda a manter um registro dos usuários, mas também estabelece um canal de comunicação que permite o fornecimento de referências aos dados baixados no futuro, reconhecendo e dando crédito àqueles que contribuíram com a geração dos dados.

Além das ferramentas citadas anteriormente que o *Firebase* oferece, foi utilizado também para gerar métricas de acessos e informações sobre o uso da plataforma o *Firebase Analytics*, podendo observar o tempo médio passado por um usuário na plataforma e visualizações da página.

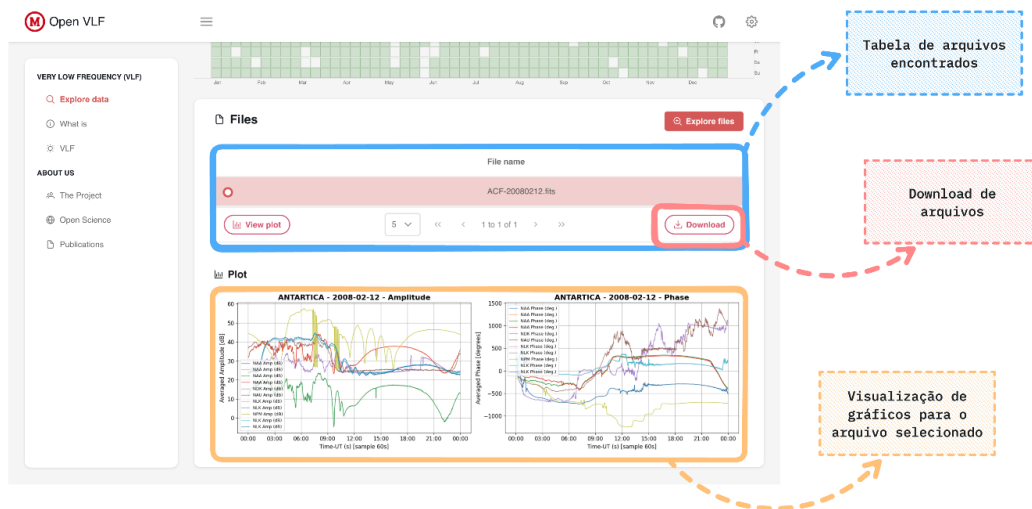


Figura 5. Interface da plataforma *Open VLF* (parte B)

4.1. Conteúdos

Além da exploração dos dados VLF, a plataforma *Open VLF* conta com seções de conteúdos sobre o projeto e explicações sobre este tipo de onda.

4.2. Métricas de armazenamento

Os dados de armazenamento do *bucket* no *AWS S3* possuem cerca de 176.969 arquivos armazenados, totalizando 73,6 GB. O gráfico possui uma linha do tempo, onde podemos observar as mudanças quando novos arquivos são enviados para a nuvem através do script citado na seção de metodologia. Na figura 6 e 7 podemos observar os formatos de armazenamento do S3 com detalhe a estrutura em JSON dos mapeamentos na plataforma MongoDB respectivamente criados durante a fase de upload.

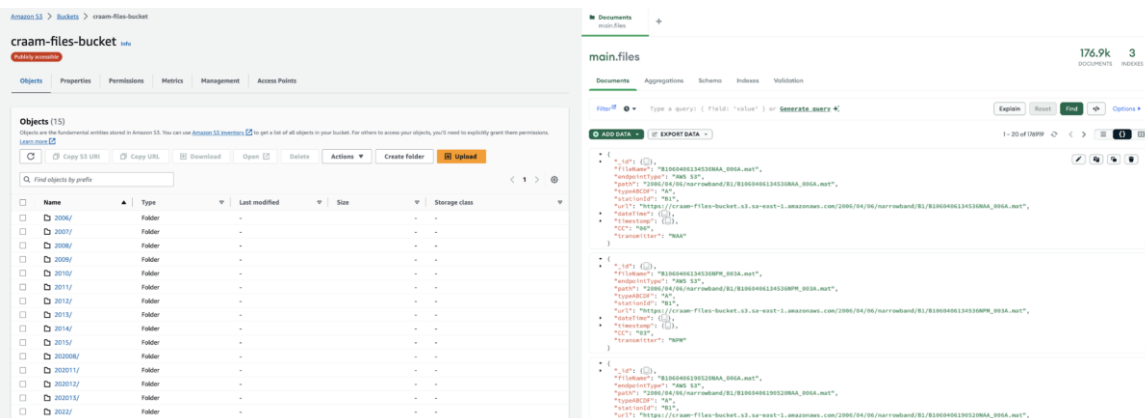


Figura 6. Arquivos VLF separados por data no *AWS S3* e em formato JSON no *MongoDB*

4.3. Métricas de desenvolvimento

Visando obter uma estimativa de tempo gasto para desenvolvimento de todos os componentes descritos na arquitetura anteriormente, juntamos informações de contribuição da plataforma *GitHub* onde se encontram todos os arquivos do projeto (Figura 8) e da ferramenta *WakaTime*, que monitora o tempo de código integrado as IDEs (*Integrated Development Environment*). Desta forma, conseguimos obter um cálculo de horas investidas no projeto, filtrados pelas tecnologias utilizadas pelos componentes, totalizando mais de 270 horas.

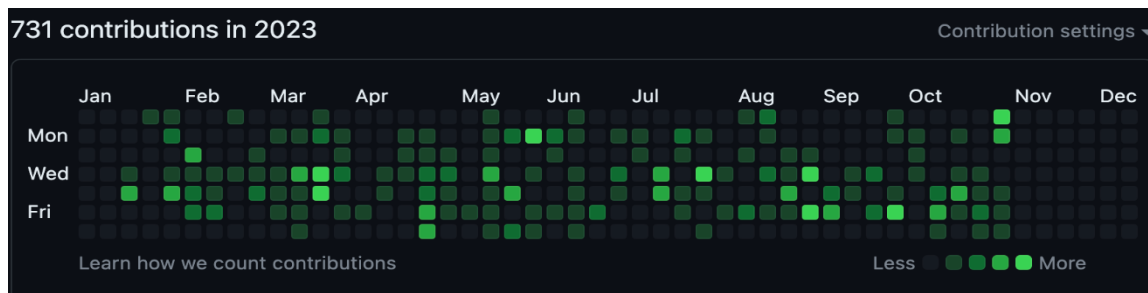


Figura 8. Contribuições feitas no projeto *Open VLF* no *GitHub* (2023)

5. Conclusão

Este trabalho desenvolve uma plataforma integrada para a consulta e análise de dados VLF (*Very Low Frequency*), com o objetivo de democratizar o acesso a esses dados e facilitar a análise científica em estudos atmosféricos e ionosféricos. A importância dessa iniciativa reside na crescente demanda por plataformas capazes de gerenciar e disponibilizar grandes volumes de dados científicos eficientemente. O projeto abordou esta necessidade, integrando dados das bases SAVNET e AWESOME e fornecendo uma interface intuitiva para pesquisa, visualização e download de dados.

Este projeto foi bem-sucedido ao fornecer fácil acesso aos dados VLF, contribuindo para a área de pesquisa em clima espacial. A arquitetura modular e escalável adotada permitiu não apenas a gestão eficiente dos dados existentes, mas também a flexibilidade para expandir e adaptar a plataforma a futuras necessidades e tipos de dados.

Contudo, há espaço para melhorias e expansões futuras. Primeiramente, a plataforma poderia ser aprimorada para abranger outros tipos de dados científicos além do VLF, ampliando seu escopo de aplicação. Isso não apenas aumentaria a utilidade da plataforma, mas também fomentaria uma colaboração mais ampla dentro da comunidade científica. Além disso, a inclusão de referências detalhadas nos dados baixados proporcionaria maior clareza sobre a origem e autoria dos dados, enriquecendo o valor acadêmico e a transparência da pesquisa. Por fim, a exploração mais aprofundada do *Firebase Analytics* poderia revelar insights valiosos sobre o uso da plataforma, permitindo otimizações contínuas e aprimoramento da experiência do usuário.

Em suma, este projeto representa um passo em direção à democratização do acesso a dados científicos e à promoção de práticas de ciência aberta, se enquadrando nos Grandes Desafios da Pesquisa em Sistemas de Informação no Brasil para a década de 2016-2026. As realizações atuais e as sugestões para melhorias futuras destacam o

potencial contínuo do projeto em contribuir para a pesquisa científica e para a comunidade acadêmica como um todo. Todo este projeto e seus artefatos podem ser acessados modo livre em <https://github.com/open-vlf> e os dados atuais em <https://s3.console.aws.amazon.com/s3/buckets/craam-files-bucket?region=sa-east-1&tab=objects>.

Agradecimentos

Obrigado à AWS Open Data [Open VLF 2023] pelo apoio ao armazenamento dos dados deste projeto, e aos professores Jean-Pierre Raulin e Emília Correa pelo apoio e orientações.

Referências

Amazon Web Services. (2023) AWS Open Data Program. URL: <https://aws.amazon.com/opendata>

AWS Craam Files. VLF Craam files (n.d.) <https://s3.console.aws.amazon.com/s3/buckets/craam-files-bucket?region=sa-east-1&tab=objects>

Cohen, M.; Golkowski, M. WALDO database (n.d.) <https://waldo.world>

Cohen, M. Returning Lightning Data to the Cloud (2020) <https://eos.org/science-updates/returning-lightning-data-to-the-cloud>

Cohen, M., Golkowski, M., Inan, U., & DeSilva, J. (2019). A Massive Public Repository of Global ELF/VLF Radio Data. In AGU Fall Meeting Abstracts (Vol. 2019, pp. SA33B-3156).

EC. European Commission, Directorate-General for Research and Innovation, Hessels, L., Koens, L., Diederens, P., Perspectives on the future of open science: effects of global variation in open science practices on the European research system, Publications Office of the European Union, (2021), <https://data.europa.eu/doi/10.2777/054281> (EN-PDF) ISBN 978-92-76-41133-8 doi: 10.2777/054281

FAPESP. Gestão de Dados (2017) <https://fapesp.br/gestaodedados>

FAPESP. Open Science (n.d.) <https://www.fapesp.br/openscience/>

Foster. What is Open Science? Introduction. (n.d.). <https://www.fosteropenscience.eu/content/what-open-science-introduction>

Kauffmann, DHV; Santiago, LS; Oliveira, R de. (2023) Open VLF Github Project. URL: <https://github.com/open-vlf>

Kauffmann, DHV; Santiago, LS; Raulin JP; Correia, E; Oliveira, R de. (2023) Open VLF: Scientific Open Data Initiative for CRAAM's SAVNET and AWESOME VLF Data. URL: <https://open-vlf.web.app/>

NAP, Committee on Scientific Accomplishments of Earth Observations from Space, National Research Council (2008). Earth Observations from Space: The First 50 Years of Scientific Achievements. The National Academies Press. p. 6.

doi:10.17226/11991. ISBN 978-0-309-11095-2. Acesso:

<https://nap.nationalacademies.org/read/11991/chapter/2>

OECD. Open Data in action Initiatives during the initial stage of the COVID-19 pandemic (2021). <https://www.oecd.org/gov/digital-government/open-data-in-action-initiatives-during-the-initial-stage-of-the-covid-19-pandemic.pdf>

OECD. Making Open Science a Reality, (2015) OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris, <https://doi.org/10.1787/5jrs2f963zs1-en>

Open Data Handbook. What is Open Data? (n.d.) <https://opendatahandbook.org/>

Raulin, J. P., Correia de Matos David, P., Hadano, R., Saraiva, A. C., Correia, E., & Kaufmann, P. (2009). The South America VLF NETwork (SAVNET). Earth, Moon, and Planets, 104(1), 247-261.

UNESCO. UNESCO Recommendation on Open Science (2021). SC-PCB-SPP/2021/OS/UROS. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=em>