

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

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 é um movimento para tornar a pesquisa financiada publicamente mais amplamente acessível à comunidade científica, às empresas e à sociedade em geral [OCDE 2015]. Ela envolve iniciativas que vão da publicação de trabalhos científicos ao uso de dados abertos, passando por infraestruturas de suporte, código aberto, etc. [Unesco 2017]. Há diferentes definições de Ciência Aberta. A Comissão Europeia [EC 2021], por exemplo, define a ciência aberta como "uma nova abordagem do processo científico baseada no trabalho cooperativo e em novas formas de difundir o conhecimento através da utilização de tecnologias digitais e de novas ferramentas colaborativas". Foster (n.d.) define a ciência aberta como "...a prática da ciência de tal forma que outros possam colaborar e contribuir, onde os dados de pesquisa, notas de laboratório e outros processos de pesquisa estão disponíveis gratuitamente, sob termos que permitem a reutilização, redistribuição e reprodução da pesquisa e seus dados e métodos subjacentes", e apresenta ainda uma taxonomia completa de termos relacionados como acesso, dados, políticas, métricas, padrões e serviços abertos. Por fim, Vicente-Saez e Martinez-Fuentes (2018) fazem uma revisão sistemática para chegar a uma definição comum entre diversos trabalhos e chegam a definir a ciência aberta como "um conhecimento transparente e acessível que é compartilhado e desenvolvido através de rede colaborativa".

Dados abertos são um dos elementos chave da ciência aberta [Fapesp, n.d. Unesco 2021] e, também, encontra várias definições. Para o Open Data Handbook [Open Data Handbook n.d.] “*dados abertos são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa - sujeitos apenas, no máximo, ao requisito de atribuição e compartilhamento*”. Apresenta ainda uma definição mais completa e precisa em que enfatiza os aspectos de a) disponibilidade e acesso; b) reuso e redistribuição; e, c) participação universal (qualquer um podendo usar, reusar e distribuir); e conclui indicando a interoperabilidade, isto é, a capacidade de diversos sistemas, organizações e com diferentes tipos de dados trabalharem juntos (interoperar). Nosso foco aqui é o de dados científicos, mas dados abertos também se referem a dados de empresas, governos e associações com poucos setores onde os dados abertos não apresentam algum papel, com destaque para os dados relacionados aos Objetivos de Desenvolvimento Sustentável (ODS) e da Carta de Dados Abertos do G8 [Davies, et al. 2019]. Especificamente para dados de pesquisa a Unesco traz a seguinte definição [Unesco 2021]: “dados de pesquisa abertos que incluem, entre outros, dados digitais e analógicos, brutos e processados, e os metadados que os acompanham, bem como pontuações numéricas, registros textuais, imagens e sons, protocolos, código de análise e fluxos de trabalho que podem ser usados abertamente, reutilizados, retidos e redistribuídos por qualquer pessoa, sujeitos a reconhecimento. Os dados de pesquisa abertos estão disponíveis a qualquer momento e de modo amigável, em formato legível às pessoas e por máquina, de acordo com os princípios de boa governança e gestão de dados, notadamente os princípios FAIR (*Findable, Accessible, Interoperable e Reusable*), apoiados por curadoria e manutenção regulares”. Este estudo empregará esta definição para orientar o seu direcionamento. A OCDE [OCDE 2007] ainda recomenda que o acesso dos dados abertos seja preferencialmente baseado na Internet, o que também será levado em conta aqui.

Independente de diferentes definições espera-se da ciência e dos dados abertos uma série de benefícios para a ciência e sociedade como [OCDE 2016]:

1. Acelerar a pesquisa científica e torná-la mais eficiente. Por exemplo, reduzindo a duplicação, tornando mais fácil e amplo o acesso.
2. Aumentar a confiabilidade e a qualidade das pesquisas e seus resultados. Por exemplo, possibilitando maior reprodutibilidade e avaliação transparentes.
3. Tornar a ciência mais sensível às necessidades públicas e desafios da sociedade. Por exemplo, tornando-a acessível a pessoas e grupos fora da comunidade científica e com o engajamento público.

Exemplos claros desses benefícios para a ciência e sociedade incluem o projeto do genoma humano e mais recentemente a resposta da ciência à pandemia de COVID-19 acelerando a obtenção de soluções para um desafio global [OCDE 2021].

Disponibilizar dados de forma aberta é uma tradição na área de dados das ciências geofísicas e astrofísicas. O conceito de acesso aberto a dados científicos foi estabelecido já em 1957-1958 [NAP 2008] como uma forma de minimizar riscos de perda de dados e maximizar a acessibilidade para grande quantidade de dados que eram produzidos por satélites em escala global, e hoje encontram-se inúmeros sites de compartilhamento de

dados, incluindo dados e imagens de satélites e de radiotelescópios, com informações do sol e seus fenômenos (SOLARMONITOR), do clima espacial (NOAA), do sistema global de navegação por satélite (CDDIS NASA) e da rede de radiotelescópios ALMA (*Atacama Large Millimeter Array*) para citar apenas alguns.

Grupos de pesquisa também disponibilizam dados abertos de antenas VLF. Destaca-se aqui a iniciativa de Cohen e Golkowski [Cohen e Golkowski n.d.; Cohen 2020; Cohen et. al 2019] que iniciaram em 2019 a disponibilização pública de dados VLF, gerados desde os anos 70 até agora, de dezenas de estações gerenciadas pelas universidades de Stanford, Georgia Tech e Colorado-Denver. O *Sudden Ionospheric Disturbances Monitoring Station A118* website [SID n.d.] é outro exemplo que fornece também dados abertos de VLF.

Nosso projeto tem muito em comum com o projeto WALDO, mas há ao menos três diferenças entre essa iniciativa e o nosso projeto: a principal é de que o nosso projeto não vincula a disponibilização dos dados a um repositório próprio do projeto, o que traz uma redução do custo, uma maior flexibilidade e agilidade para disponibilização dos dados (um dos fatores que levam o projeto WALDO ainda estar, passados 3 anos realizando a transferência de dados); segundo, o projeto WALDO inclui somente dados gerados por sistemas de antena AWESOME. Nosso projeto inclui dados gerados pelo sistema SAVNET, de diferente formato, e está desenhado para suportar ainda outros formatos de dados; terceiro, a visualização dos dados é desenhada para gerar gráficos apenas sobre demanda. Gráficos gerados são armazenados apenas temporariamente, como um cache, caso sejam novamente consultados. Isso permite uma economia de armazenamento (e lembramos que nosso sistema não tem um armazenamento próprio dedicado) comparado ao sistema WALDO, que armazena visualizações de todos os dados, e produzidas previamente, em formato PNG (*Portable Network Graphics*).

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.

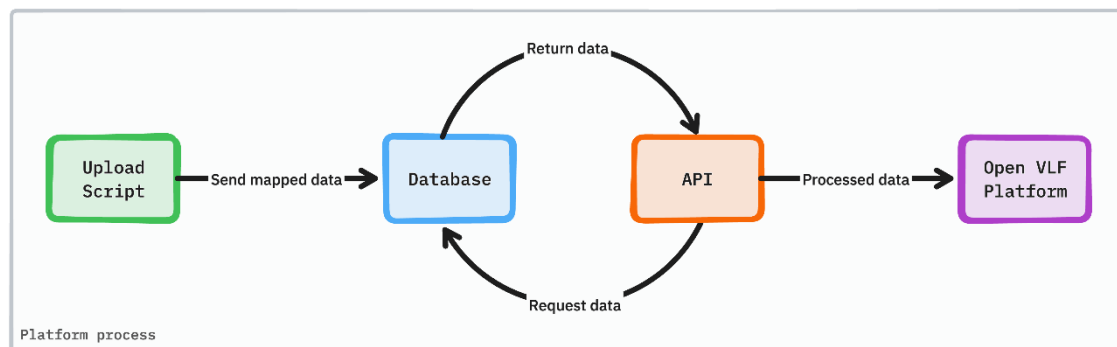


Figura 1. Arquitetura da plataforma em alto nível (componentes agrupados)

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 (Figura 3).

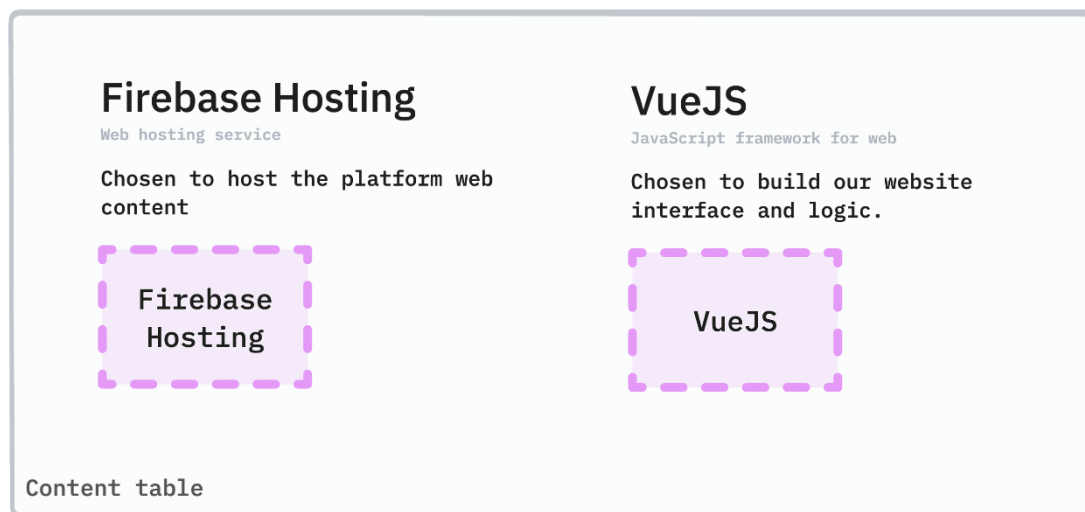


Figura 2. Tabela de explicação dos componentes *Firebase Hosting* e *VueJS*

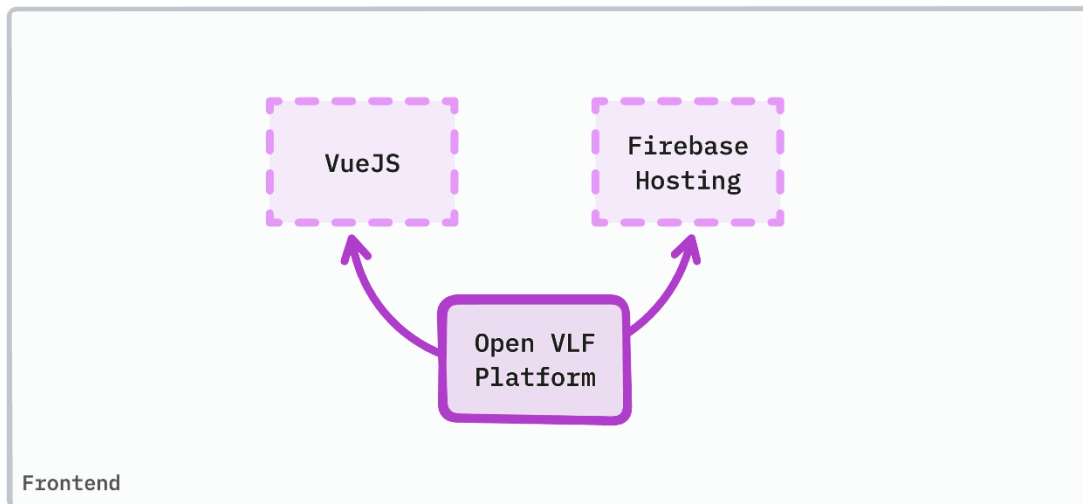


Figura 3. Resumo do agrupamento *Open VLF Platform*

3.3. Aquisição de dados e backend

Um script Python é empregado para coleta e armazenamento dos dados (Figura 4). 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.

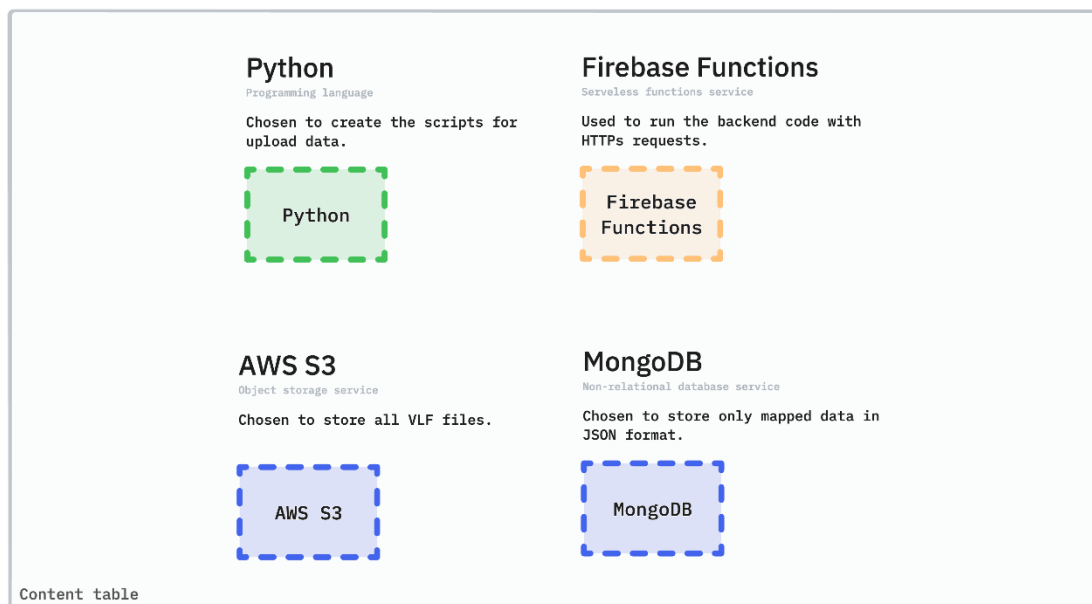


Figura 4. Tabela de explicação dos componentes *Python*, *Firebase Functions*, *AWS S3* e *MongoDB*

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 (Figura 5). 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.

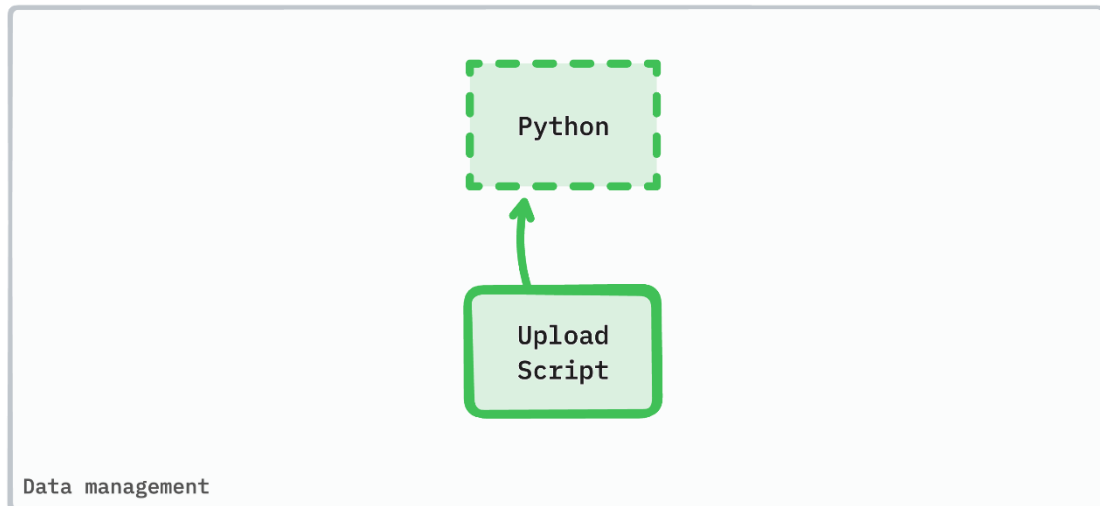


Figura 5. Resumo do agrupamento *Data Management*

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 (Figura 6).

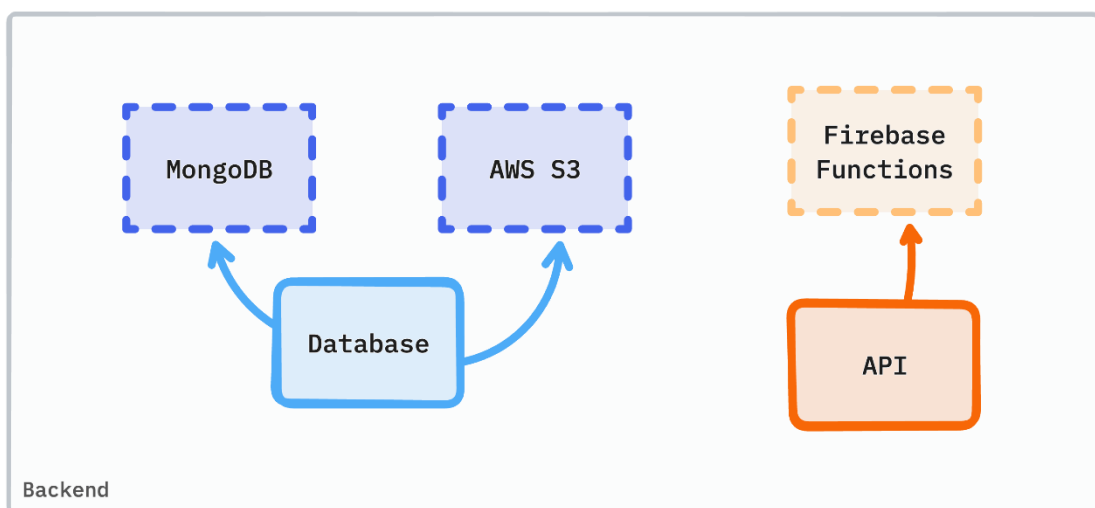


Figura 6. Resumo do agrupamento *Backend*

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 7 mostra uma captura de tela da interface da plataforma, destacando a seção de download de um arquivo encontrado.

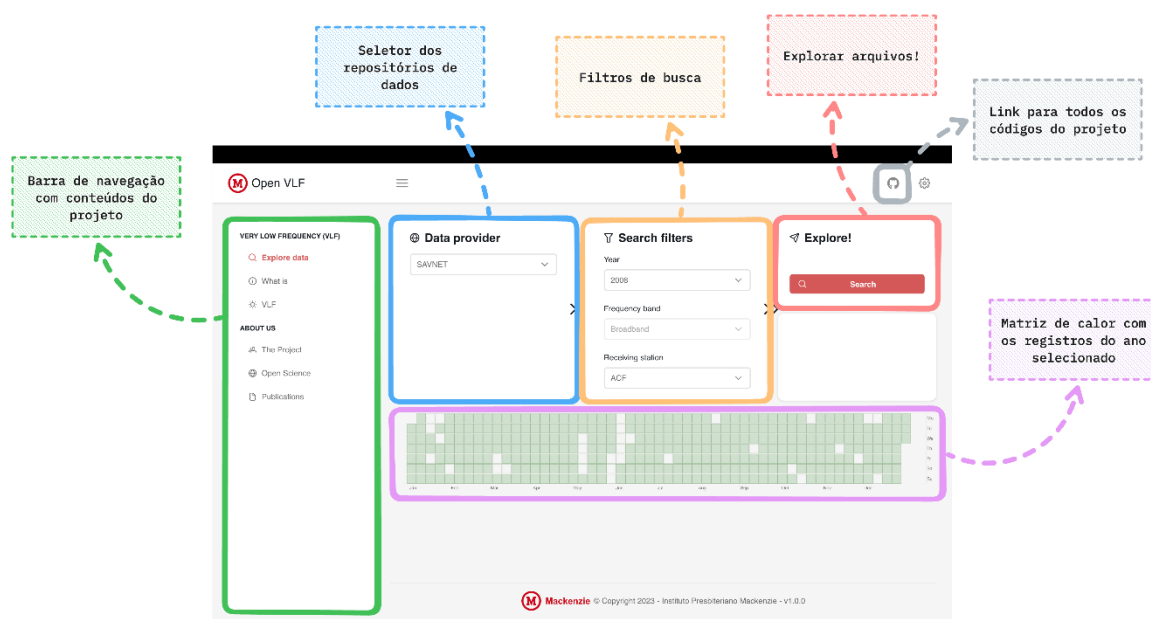


Figura 7. Detalhamento 1 da interface da plataforma *Open VLF*

Os usuários têm a capacidade de selecionar e baixar arquivos VLF de interesse diretamente pela plataforma (Figura 8). 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 (Figura 9).

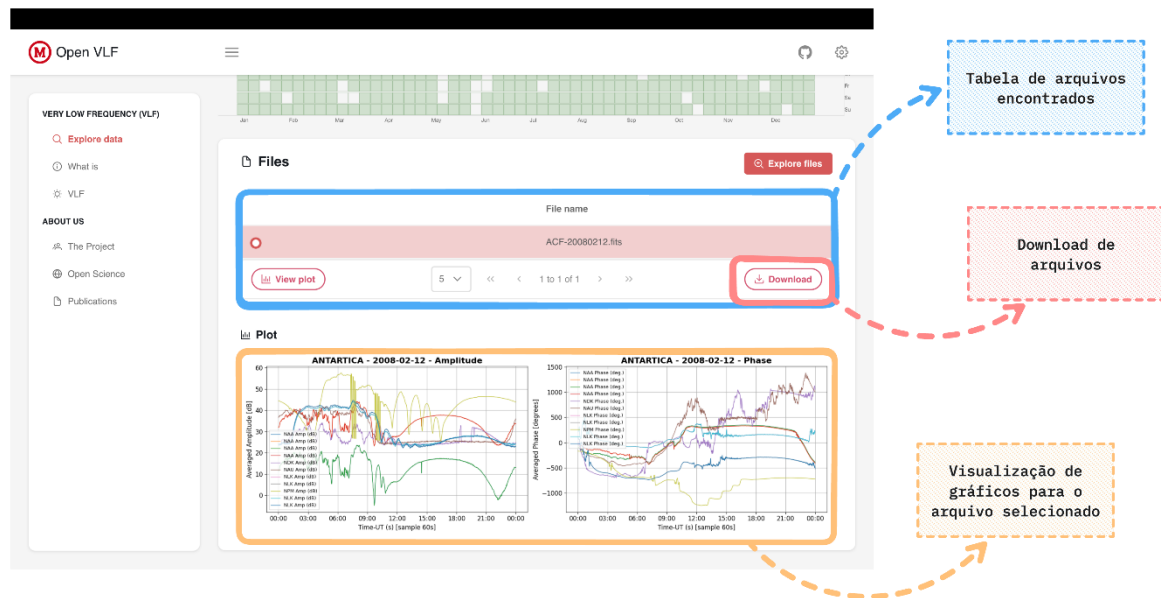


Figura 8. Detalhamento 2 da interface da plataforma *Open VLF*

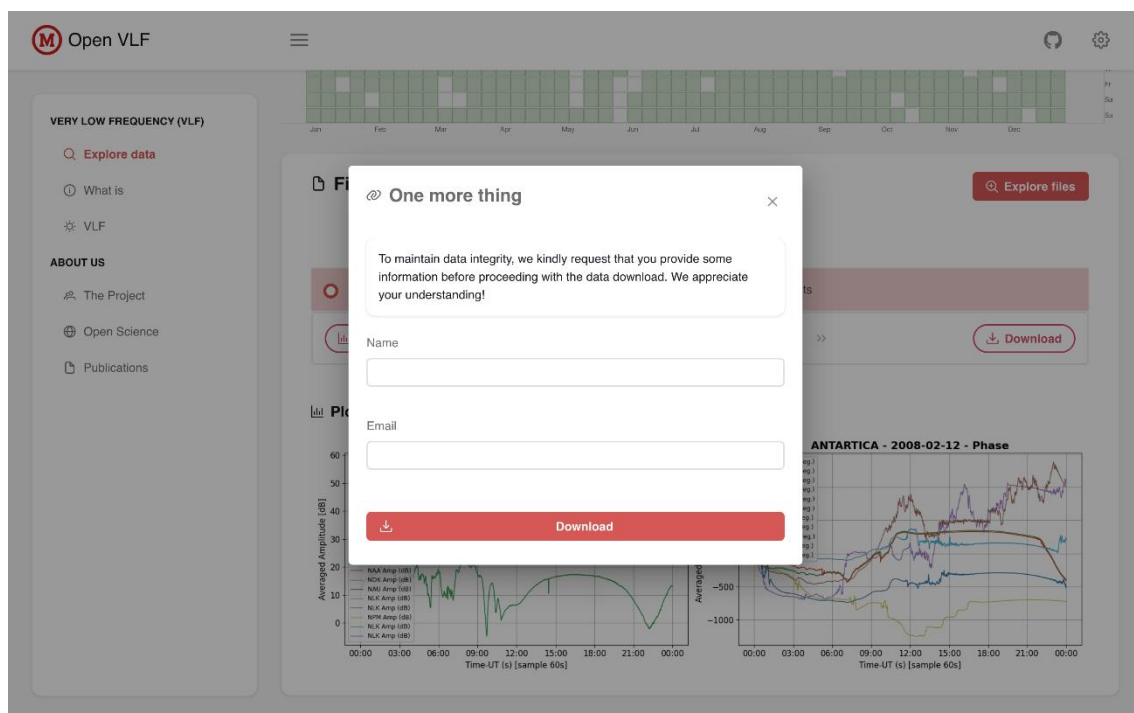


Figura 9. Modal de download

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* (Figura 10). Como mostra a figura acima, podemos observar quanto tempo médio é passado por um usuário na plataforma e visualizações da página.

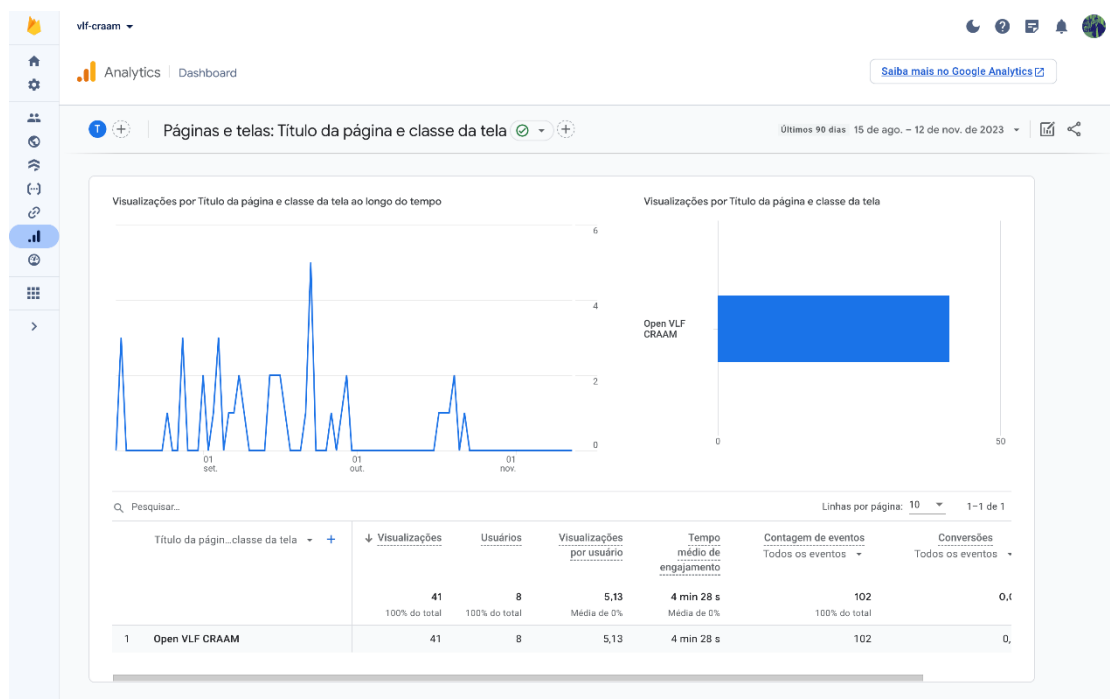


Figura 10. Página de dados gerado no *Firebase Analytics*

4.1. Conteúdos

Além da exploração dos dados VLF, a plataforma *Open VLF* conta com seções de conteúdos sobre o projeto (Figura 11) e explicações sobre este tipo de onda (Figura 12).

The screenshot shows the 'Open VLF' website. The header includes the 'Open VLF' logo and a menu icon. The main content area is titled 'What is' and contains the following text:

This site is maintained by **CRAAM** (Mackenzie Radio Astronomy and Astrophysics Center) a research center operated by the **UPM** (Mackenzie Presbyterian University) and **INPE** (National Institute for Space Research), to provide public and free access for researchers, students and the interested public access to VLF (Very Low Frequency) data from CRAAM's Antennas systems. All data stored is supported by **Amazon AWS** supported by **AWS Open Data Program**.

Very Low Frequency (VLF) signals can be used for navigation services, communication with submarines and are a powerful tool to study the low altitude Earth's ionosphere, atmospheric and geophysics phenomena, space weather, magnetic field and solar flares. Here, we will provide historical and updated VLF data from two CRAAM's system antennas, the **SAVNET** and **AWESOME** systems.

The **SAVNET** consists of a network of receiving/tracking 11 VLF stations, located in Latin America and Antarctica. Five stations are in Brazil, 3 in Peru, 1 in Argentina, 1 in Mexico and 1 in the Brazilian Antarctic Research Station Comandante Ferraz. The **CRAAM's AWESOME** systems include two antennas (Brazil and Antarctica). Data since 2002 from both system are available with some absence of data during periods when collections are interrupted.

The footer features logos for 'Universidade Presbiteriana Mackenzie', 'INPE', and 'Open Data on AWS Data Exchange'. The copyright notice at the bottom reads: 'Mackenzie © Copyright 2023 - Instituto Presbiteriano Mackenzie - v1.0.0'.

Figura 11. Seção sobre o projeto na plataforma *Open VLF*

Open VLF

VERY LOW FREQUENCY (VLF)

- Explore data
- What is
- VLF**

ABOUT US

- The Project
- Open Science
- Publications

VLF

Very Low Frequency (VLF) signals typically occupy the frequency range from 3 kHz (kilohertz) to 30 kHz signals and are known to propagate over long distances within the waveguide bounded by the Earth's surface and the lower ionosphere (D-region at daytime and E-region at nighttime). The longer wavelengths from VLF compared to higher-frequency radio waves, which gives them unique propagation characteristics. They can be affected by factors like the Earth's curvature, ground conductivity, and the ionosphere.

These characteristics made VLF suitable for many research and applications:

- Research ionosphere phenomena's: VLF can be used to capture many phenomena's that affect ionosphere, such as solar flares, space weather and the magnetosphere.

Figura 12. Seção sobre ondas VLF na plataforma *Open VLF*

Buscando cultivar mais a comunidade acadêmica e científica no âmbito do VLF, adicionamos uma seção com publicações relacionadas a VLF, SAVNET e ao CRAAM (Figura 13). Também, explicando mais sobre a Ciência Aberta e como citar este projeto.

Open VLF

VERY LOW FREQUENCY (VLF)

- Explore data
- What is
- VLF

ABOUT US

- The Project
- Open Science
- Publications**

Publications

Publications related to VLF, SAVNET and CRAAM group.

- JORGE SAMANES; JEAN-PIERRE RAULIN; CAO JINBIN. Estimation of the nighttime height of the lower ionosphere using VLF waves propagation. 2016 URSI Asia-Pacific Radio Science Conference, August 21-25, 2016/Seoul, Korea.
- ANTONIO A. S. MAGALHAES; JEAN-PIERRE RAULIN; JORGE SAMANES. Lower ionosphere nighttime height estimates using VLF waves propagation. Session ST09-URSI, MOMAG2016, July 25-29, 2016, Porto Alegre, Brazil.
- SAMANES, J. E.; RAULIN, J.-P. ; MACOTELA, E. ; GUEVARA-DAY, W. Estimating the modal interference distance using the South America VLF Network (SAVNET). Radio Sci. Vol.30,02, p. 122-129, February 2015.
- BORGAZZI, A. ; LARA, A. ; PAZ, G. ; RAULIN, J.-P. The Ionosphere and the Latin America VLF Network Mexico (LAVNet-Mex) Station. Advances in Space Research, v. 54, p. 536-545, 2014.
- RAULIN, JEAN-PIERRE; TROTTET, GÉRARD; GIMÉNEZ DECASTRO, C. GUILLERMO ; CORREIA, EMILIA; MACOTELA, E. LILIANA . Nighttime Sensitivity of Ionospheric VLF Measurements to X-ray Bursts From a Remote Cosmic Source. Journal of Geophysical Research: Space Physics, v. 119, p. n/a-n/a, 2014.
- SAMANES, JORGE E. ; RAULIN, JEAN-PIERRE ; MACOTELA, EDITH L. ; GUEVARA-DAY, WALTER R. . Estimating the VLF Modal Interference Distance using the South America VLF Network (SAVNET). Radio Science, v. 50, p. 122-129, 2014.
- JEAN-PIERRE RAULIN, GÉRARD TROTTET, MATHIEU KRETZSCHMAR, EDITH L. MACOTELA, ALESSANDRA PACINI, FERNANDO C. P. BERTONI, INGOLF DAMMASCH. Response of the low ionosphere to X-ray and Lyman- α solar flare emissions. J. of Geoph. Res: Sp. Phys, vol. 118, p. 570-575, 2013.
- CORREIA, E. ; RAULIN, J.-P. ; KAUFMANN, P. ; BERTONI, F. ; QUEVEDO, M.T. Inter-hemispheric analysis of daytime low ionosphere behavior from 2007 to 2011. Journal of Atmospheric and Solar-Terrestrial Physics, v. 92, p. 51-58, 2013.
- BERTONI, FERNANDO CELSO PERIN ; RAULIN, JEAN-PIERRE ; RIVERO GAVILÁN, HERNÁN ; KAUFMANN, PIERRE ; RODRIGUEZ, RODOLFO ; CLILVERD, MARK ; CARDENAS, JORGE SAMANES ; FERNANDEZ, GERMÁN . Lower ionosphere monitoring by the South America VLF Network (SAVNET): C-region occurrence and atmospheric temperature variability. Journal of Geophysical Research: Space Physics, v. 118, p. n/a-n/a, 2013.
- HAYAKAWA, M. ; RAULIN, J.-P. ; KASAHARA, Y. ; BERTONI, F. C. P. ; HOBARA, Y. ; GUEVARA-DAY, W. . Ionospheric perturbations in possible

Figura 13. Seção sobre as publicações relacionadas na plataforma *Open VLF*

4.2. Métricas de armazenamento

Na figura 14, podemos observar os dados de armazenamento do *bucket* no *AWS S3*, onde mostra um total 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.

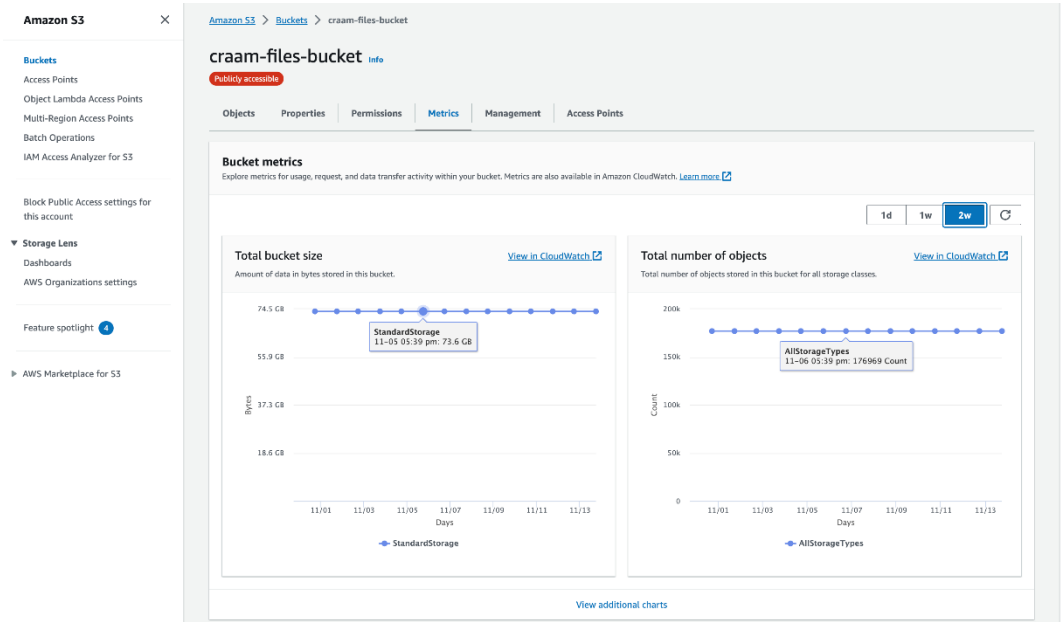


Figura 14. Métricas do AWS S3 com os arquivos VLF

The screenshot displays the 'Objects' tab for the bucket 'craam-files-bucket'. It shows a list of 15 objects, all of which are folders named by year (2006/ to 2022/). The table includes columns for Name, Type, Last modified, Size, and Storage class. The interface includes a sidebar with navigation options like Buckets, Access Points, and Storage Lens, and a top navigation bar with tabs for Objects, Properties, Permissions, Metrics, Management, and Access Points.

Name	Type	Last modified	Size	Storage class
2006/	Folder	-	-	-
2007/	Folder	-	-	-
2008/	Folder	-	-	-
2009/	Folder	-	-	-
2010/	Folder	-	-	-
2011/	Folder	-	-	-
2012/	Folder	-	-	-
2013/	Folder	-	-	-
2014/	Folder	-	-	-
2015/	Folder	-	-	-
202008/	Folder	-	-	-
202011/	Folder	-	-	-
202012/	Folder	-	-	-
202013/	Folder	-	-	-
2022/	Folder	-	-	-

Figura 15. Arquivos VLF separados por data no AWS S3

Além da organização dos arquivos de registros VLF no *AWS S3*, na figura 16 podemos observar com detalhe a estrutura em JSON dos mapeamentos na plataforma *MongoDB* criados durante a fase de *upload*.

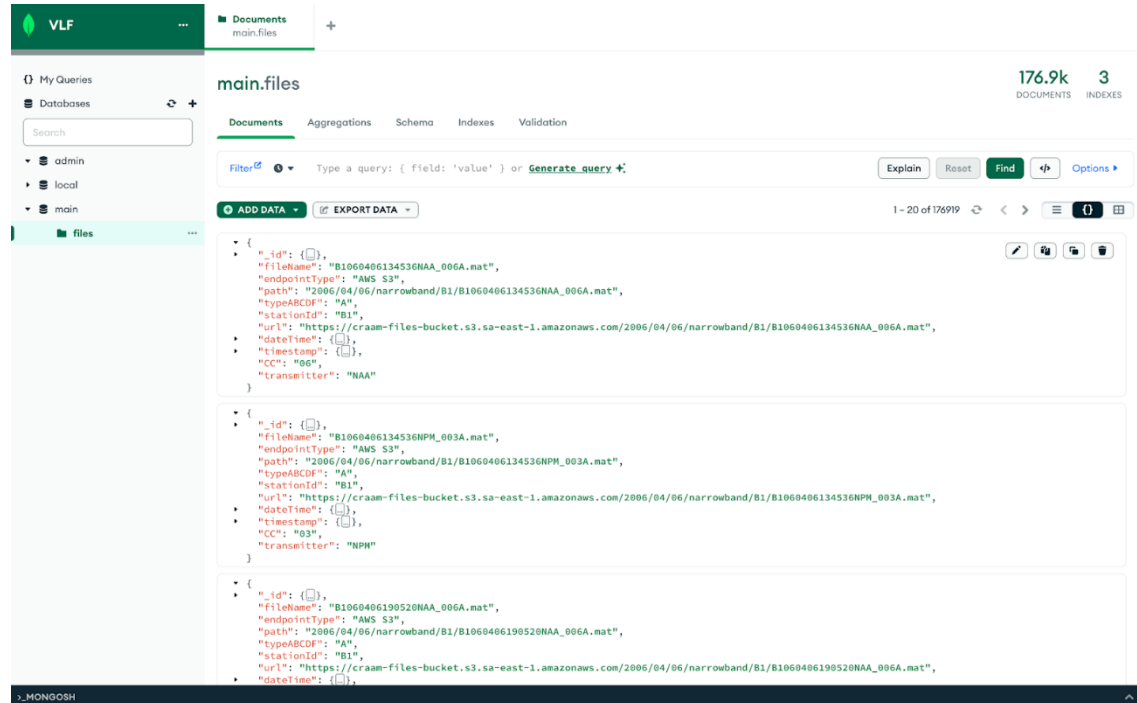


Figura 16. Mapeamento dos arquivos VLF em formato JSON no *MongoDB*

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 17) e da ferramenta *WakaTime* (Figura 18), 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.

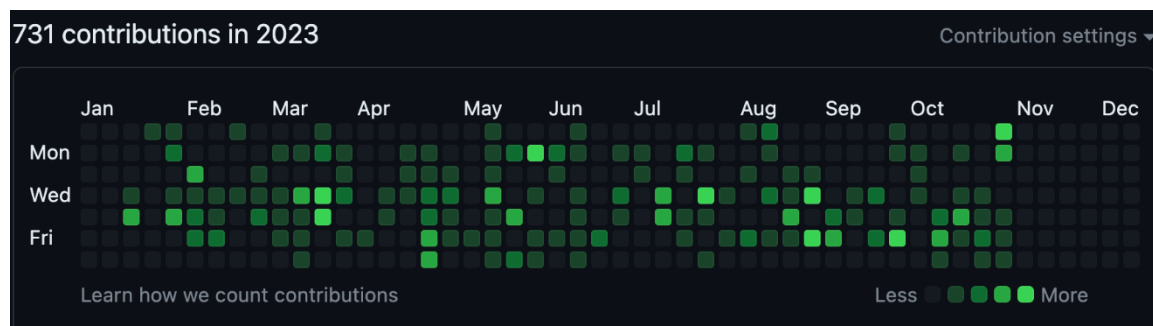


Figura 17. Matriz de calor das contribuições feitas no projeto *Open VLF* no *Github* durante 2023

2023 (Open VLF Project) **273 hrs 14 mins**



LANGUAGES ?

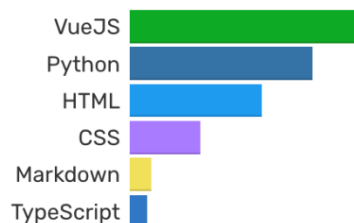


Figura 18. Gráfico de horas de desenvolvimento nos componentes do projeto

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).

Davies, T., Walker, S. B., Rubinstein, M., & Perini, F. The state of open data: Histories and horizons. (2019). African Minds. <https://stateofopendata.od4d.net/>

EC. European Commission, Directorate-General for Research and Innovation, Hessels, L., Koens, L., Diederer, 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>

Google LLC (2023) Firebase Documentation. URL: <https://firebase.google.com/docs>

Hamlett, Alan (2023) WakaTime. URL: <https://wakatime.com>

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/>

MongoDB Inc. (2023) Mongo Developer Center. URL: <https://www.mongodb.com/developer/>

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>

Nina, A. (2020). Application of Databases Collected in Ionospheric Observations by VLF/LF Radio Signals. In Knowledge Discovery in Big Data from Astronomy and Earth Observation (pp. 419-434, chapter 23). Elsevier.
<https://doi.org/10.1016/B978-0-12-819154-5.00035-7>

OCDE. 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>

OECD. OECD Principles and Guidelines for Access to Research Data from Public Funding (2007). <https://www.oecd.org/sti/inno/38500813.pdf>

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.

SID. Sudden Ionospheric Disturbances Monitoring Station A118 (n.d.).
<https://sidstation.loudet.org/home-en.xhtml>

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

Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. Journal of Business Research, 88, 428–436. doi:10.1016/j.jbusres.2017.12.043