

# Medindo o Brasil a partir do espaço: tecnologias de *big data* e inteligência artificial

## *Measuring Brazil from Space: big data and artificial intelligence technologies*

Karine Reis Ferreira<sup>1</sup>, Gilberto Ribeiro de Queiroz<sup>2</sup>

<sup>1</sup> Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP, Brasil. ORCID: <https://orcid.org/0000-0003-2656-5504>

<sup>2</sup> Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP, Brasil. ORCID: <https://orcid.org/0000-0001-7534-0219>

Autor para correspondência/Mail to: Karine Reis Ferreira, [karine.ferreira@inpe.br](mailto:karine.ferreira@inpe.br)



**Karine Reis Ferreira** trabalha no Instituto Nacional de Pesquisas Espaciais (INPE). É coordenadora do projeto Brazil Data Cube e do Laboratório GeoBIDS (Geoinformatics and Big Data Science Lab).



**Gilberto Ribeiro de Queiroz** trabalha no Instituto Nacional de Pesquisas Espaciais (INPE). É coordenador do projeto Brazil Data Cube e do Laboratório GeoBIDS (Geoinformatics and Big Data Science Lab).



Copyright © 2021 Ferreira & Queiroz. Todo o conteúdo da Revista (incluindo-se instruções, política editorial e modelos) está sob uma licença Creative Commons Atribuição 4.0 Internacional. Ao serem publicados por esta Revista, os artigos são de livre uso em ambientes educacionais, de pesquisa e não comerciais, com atribuição de autoria obrigatória. Mais informações em <http://revistas.ufpr.br/atoz/about/submissions#copyrightNotice>.

### Resumo

A Dra. Karine Reis Ferreira e o Dr. Gilberto Ribeiro de Queiroz respondem questões sobre o projeto de pesquisa Brazil Data Cube, sua finalidade, impacto e relação com inteligência artificial e *big data*.

**Palavras-chave:** Big data; Inteligência artificial; Inovação tecnológica; Data cube; Brazil Data Cube.

### Abstract

Dr. Karine Reis Ferreira and Dr. Gilberto Ribeiro de Queiroz from Instituto Nacional de Pesquisas Espaciais (INPE) answers questions about big data technology and artificial intelligence, specially in the context of contexto of Brazil Data Cube project.

**Keywords:** Big data; Artificial intelligence; Technological innovation; Data cube; Brazil Data Cube.

## 1. Profissionalmente, quem são Karine e Gilberto?

Karine é doutora em Computação Aplicada e trabalha no Instituto Nacional de Pesquisas Espaciais (INPE) há 20 anos com pesquisa e desenvolvimento (P&D) em GeoInformática, participando e coordenando projetos inovadores nessa área como TerraLib, TerraBrasilis e Brazil Data Cube. É docente do curso de pós-graduação em Computação Aplicada do INPE. Sua principal linha de pesquisa é representação e processamento de dados espaço-temporais, análise de séries temporais de imagens e grandes bases de dados de observação da Terra.

Gilberto é tecnologista sênior do Instituto Nacional de Pesquisas Espaciais (INPE). Possui mestrado e doutorado em Computação Aplicada. Trabalha com projetos de pesquisa e desenvolvimento relacionados às geotecnologias que apoiam as atividades do Programa de Monitoramento da Amazônia e Demais Biomas Brasileiros. Desde de 2016 é docente permanente da linha de pesquisa em Geoinformática e Ciência de Dados Geoespaciais no curso de pós-graduação em Computação Aplicada do INPE.

## 2. O que é o Brazil Data Cube?

Brazil Data Cube (BDC) é um projeto de pesquisa, desenvolvimento e inovação tecnológica do Instituto Nacional de Pesquisas Espaciais (INPE). Esse projeto está produzindo dados a partir de grandes volumes de imagens de sensoriamento remoto de média resolução para todo o território nacional e desenvolvendo uma plataforma computacional para processar e analisar esses dados usando inteligência artificial, aprendizado de máquina e análise de séries temporais de imagens (Ferreira et al., 2020a). Essa iniciativa começou em 2019 como um subprojeto do projeto Monitoramento Ambiental dos Biomas Brasileiros, financiado com recursos do Fundo Amazônia, por meio da colaboração financeira BNDES e FUNCATE.

Os dados produzidos no projeto BDC incluem dados prontos para análise (*Analysis-Ready Data* - ARD), cubos de dados multidimensionais e mosaicos a partir de imagens dos satélites CBERS-4/4A, Sentinel-2 e Landsat-8. A plataforma computacional é composta por serviços web, aplicativos de *software* e ambientes de computação iterativa. Usando inteligência artificial, aprendizado de máquina e análise de séries temporais de imagens, mapas de uso e cobertura da Terra estão sendo produzidos a partir desses cubos de dados (Picoli et al., 2020; Simoes et al., 2021). Mais informações sobre o projeto BDC podem ser encontradas em <http://brazildatacube.org>.

### 3. O que é *Big Data* no contexto do Brazil Data Cube? Qual é a necessidade de armazenamento para os dados deste projeto?

Atualmente, temos disponível um grande volume de dados abertos de imagens de satélites de observação da Terra, com diferentes resoluções espaciais e temporais. Somente em 2019, os satélites Landsat-7/8, MODIS Terra/Aqua e Sentinel-1/2/3 geraram 5 petabytes de imagens (Soille et al., 2018). Para gerenciar, processar e analisar esse grande volume de dados, novas plataformas computacionais foram desenvolvidas, como Google Earth Engine (GEE), Open Data Cube (ODC) e SEPAL (Gomes, Queiroz, & Ferreira, 2020).

No projeto BDC, estamos armazenando, processando e analisando coleções de dados ARD e cubos de dados de imagens dos satélites Sentinel-2, Landsat-8, CBERS-4/4A e Terra/Aqua (MODIS). O volume estimado desse acervo é de 3 petabytes. Para manipular de maneira eficiente todo esse acervo, alguns serviços *web* e sistemas da plataforma BDC foram desenvolvidos para rodar em ambientes de computação em nuvem, como a Amazon Web Service (AWS) (Ferreira et al., 2020b). A AWS, através de seu programa Earth on AWS (<https://aws.amazon.com/earth/>), mantém um repositório aberto e sem custos de imagens de satélites de observação da Terra. O sistema Data Cube Builder desenvolvido no projeto BDC é capaz de rodar na AWS usando seus serviços de computação distribuída e gerar cubos de dados a partir de coleções de imagens do Earth on AWS, de maneira eficiente e sem a necessidade de fazer *download* dessas coleções. Além disso, estamos desenvolvendo uma interface com a plataforma ODC (Gomes, Carlos, Queiroz, Ferreira, & Santos, 2021) e pacotes de *software* capazes de processar os dados em partes e de modo paralelo (Simoes et al., 2021).

### 4. Qual o papel da Inteligência Artificial no projeto?

No projeto BDC, estamos usando análise de séries temporais de imagens e métodos de *machine learning* e *deep learning* (como *Random Forest*, *Support Vector Machine*, *Recurrent Neural Networks* (CNN) e *TempCNN*) para extrair informação de uso e cobertura da Terra a partir de cubos de dados multidimensionais de imagens de sensoriamento remoto (Ferreira et al., 2020a; Picoli et al., 2020; Simoes et al., 2021). Além disso, estamos desenvolvendo métodos para avaliar e melhorar a qualidade de amostras de uso e cobertura da Terra usadas para treinar esses classificadores, baseados em redes neurais e agrupamentos de séries temporais de imagens (Santos, Ferreira, Camara, Picoli, & Simoes, 2021; Santos, Ferreira, Picoli, et al., 2021).

Um dos produtos de *software* desenvolvidos no projeto é um pacote na linguagem de programação R, que implementa boa parte do suporte computacional necessário à geração desses mapas de uso e cobertura da Terra, apoiando todo o *workflow* envolvido. Este pacote, denominado *sits* fornece funcionalidades que vão desde a extração das séries temporais associadas às amostras de treinamento, algoritmos de análise de agrupamentos (*clustering*) das amostras, abstração dos modelos de aprendizado de máquina, pós-processamento, validação e visualização de dados (Simoes et al., 2021). Assim como todos os sistemas desenvolvidos pelo projeto, o *sits* é *software* livre e seu código fonte encontra-se disponível no GitHub (<https://github.com/e-sensing/sits>).

### 5. Como vocês veem a demanda para capacidade de armazenamento e processamento para o projeto Brazil Data Cube em 10 anos?

Atualmente, grupos de pesquisa, nacionais e internacionais, e instituições de Estado têm identificado o potencial de uso da plataforma BDC para atenderem suas demandas e missões. Neste sentido, além do volume crescente de dados previamente definidos no início do projeto, estão surgindo demandas para que a plataforma inclua novos conjuntos de dados. Portanto, o volume de dados será sempre um dos grandes desafios do BDC, tanto em relação ao armazenamento e gerenciamento de dados da plataforma quanto para a extração de informações. A nossa ideia é continuar avançando nossa infraestrutura e tecnologias para suportar imagens mais recentes assim como novos dados que serão incorporados por demanda.

### 6. Qual é o impacto e quais são os benefícios do projeto para a sociedade brasileira?

O projeto BDC está gerando dados abertos e produtos de *software* livre de licença que são úteis para a sociedade brasileira. Estamos colaborando para maximizar o valor e impacto do dado de observação da Terra, que é uma importante fonte de dados sobre nosso planeta. Dessa forma, o BDC está possibilitando o desenvolvimento de novas aplicações e a melhoria de sistemas já existentes. Usando esses produtos de dados e *software*, em uma única plataforma, qualquer pessoa pode acessar, visualizar e processar dados para gerar diferentes tipos de informações, como por exemplo, mapas de uso e cobertura da Terra.

Além disso, o BDC está desenvolvendo inovação tecnológica para o Programa de Monitoramento da Amazônia e outros Biomas do INPE e seus projetos associados, como o DETER, PRODES e TerraClass; que identificam áreas

desmatadas e classificam o que essas áreas se tornaram (ex.: pastagem ou agricultura) depois do desmatamento. Além de desenvolver tecnologias para acessar e visualizar grandes volumes de imagens de sensoriamento remoto, o BDC está desenvolvendo algoritmos usando aprendizado de máquina e análise de séries temporais de imagens para extrair informações de uso e cobertura da Terra a partir dos cubos de dados.

No cenário mundial, o projeto BDC coloca o Brasil na vanguarda em pesquisa e desenvolvimento de cubos de dados de observação da Terra e de tecnologias associadas. Atualmente, existem diferentes projetos e iniciativas para gerar cubos de dados de observação da Terra para países ou regiões específicas, como por exemplo, Australian Data Cube (Lewis et al., 2021), Swiss Data Cube (Giuliani et al., 2017) e Africa Regional Data Cube (Killough, 2019). O projeto BDC é a iniciativa do Brasil nessa área, desenvolvendo no âmbito nacional e internacional liderança científica e tecnológica neste tema.

### **7. O que mais animam (ou motivam) vocês em relação ao Brazil Data Cube?**

Nossa maior motivação nesse projeto é a capacidade de aliar pesquisa e desenvolvimento de alto nível de inovação com benefícios concretos para a sociedade e para o Estado Brasileiro. No BDC, estamos conseguindo balancear a produção científica e acadêmica com a produção de tecnologias e produtos de dados que possam chegar à sociedade e atendê-la da melhor forma possível. Além de colocar nosso país na vanguarda desse tema, estamos capacitamos profissionais e alunos na área de *big data*, cubos de dados de observação da Terra, aprendizado de máquina, análise de séries temporais e geoinformática.

## REFERÊNCIAS

- Ferreira, K. R., Queiroz, G. R., Vinhas, L., Marujo, R. F. B., Simoes, R. E. O., Picoli, M. C. A., ... Fonseca, L. M. G. (2020a). Earth observation data cubes for Brazil: Requirements, methodology and products. *Remote Sensing*, *12*(24). doi: 10.3390/rs12244033
- Ferreira, K. R., Queiroz, G. R., Vinhas, L., Marujo, R. F. B., Simoes, R. E. O., Picoli, M. C. A., ... Zaglia, M. C. (2020b). Using remote sensing images and cloud services on AWS to improve land use and cover monitoring. In *2020 Latin American GRSS & ISPRS Remote Sensing Conference*. Santiago, Chile: LAGIRS.
- Giuliani, G., Chatenoux, B., De Bono, A., Rodila, D., Richard, J. P., Allenbach, K., ... Peduzzi, P. (2017). Building an earth observations data cube: Lessons learned from the Swiss data cube (SDC) on generating analysis ready data (ARD). *Big Earth Data*, *1*(1-2), 100-117. doi: 10.1080/20964471.2017.1398903
- Gomes, V. C. F., Carlos, F. M., Queiroz, G. R., Ferreira, K. R., & Santos, R. (2021). Accessing and processing Brazilian earth observation data cubes with the open data cube platform. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (v. 4, p. 153-159). doi: 10.5194/isprs-annals-V-4-2021-153-2021
- Gomes, V. C. F., Queiroz, G. R., & Ferreira, K. R. (2020). An overview of platforms for big earth observation data management and analysis. In *Remote Sensing* (v. 8). doi: 10.3390/rs12081253
- Killough, B. (2019). The impact of analysis ready data in the Africa regional data cube. In *Proceedings of the IGARSS 2019: IEEE International Geoscience and Remote Sensing Symposium* (p. 5646-5649). Yokohama: IEEE. doi: 10.1109/IGARSS.2019.8898321
- Lewis, L., Olivera, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., ... Wanga, L.-W. (2021). The Australian geoscience data cube: Foundations and lessons learned. *Remote Sensing Journal*, *202*, 276-292. doi: 10.1016/j.rse.2017.03.015
- Picoli, M. C. A., Simoes, R., Chaves, M., Santos, L. A., Sanchez, A., Soares, A., ... Queiroz, G. R. (2020). CBERS data cube: A powerful technology for mapping and monitoring Brazilian biomes. In *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* (p. 533-539). doi: 10.5194/isprs-annals-V-3-2020-533-2020
- Santos, L. A., Ferreira, K. R., Camara, G., Picoli, M. C. A., & Simoes, R. E. (2021). Quality control and class noise reduction of satellite image time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, *177*, 75-88. doi: 10.1016/j.isprsjprs.2021.04.014
- Santos, L. A., Ferreira, K. R., Picoli, M., Camara, G., Zurita-Milla, R., & Augustijn, E.-W. (2021). Identifying spatiotemporal patterns in land use and cover samples from satellite image time series. *Remote Sensing Journal*, *13*(5), 974. doi: 10.3390/rs13050974
- Simoes, R., Camara, G., Queiroz, G., Souza, F., Andrade, P. R., Santos, L., ... Ferreira, K. (2021). Satellite image time series analysis for big earth observation data. *Remote Sensing*, *13*(13). doi: 10.3390/rs13132428
- Soille, P., Burger, A., De Marchi, D., Kempeneers, P., Rodriguez, D., Syrris, V., & Vasilev, V. (2018). A versatile data-intensive computing platform for information retrieval from big geospatial data. *Future Generation Computer Systems*, *81*, 30-34. doi: 10.1016/j.future.2017.11.007
- Como citar esta entrevista (APA):  
Ferreira, K. R. & Queiroz, G. R. de (2021). Medindo o Brasil a partir do espaço: tecnologias de big data e inteligência artificial. *AtoZ: novas práticas em informação e conhecimento*, *10*(3), 1 - 4. Recuperado de: <http://dx.doi.org/10.5380/atoz.v10i3.81966>