

Utilizando o padrão SpatioTemporal Asset Catalog para visualização de dados

Thais de Medeiros¹, Bruno dos Santos¹, Gilberto Oliveira¹, Thales Körting²,
Gilberto de Queiroz²

¹Programa de Pós Graduação em Sensoriamento Remoto - Instituto Nacional de Pesquisas Espaciais (INPE) - São José dos Campos, SP - Brasil

²Divisão de Observação da Terra e Geoinformática - Instituto Nacional de Pesquisas Espaciais (INPE) - São José dos Campos, SP - Brasil

thaispmedeiros97@gmail.com, bruno.santos@inpe.br, gilbertoedi@gmail.com, thales.korting@inpe.br, gilberto.queiroz@inpe.br

Abstract. *The Brazil Data Cube project (BDC) is an initiative aiming at the production of the multidimensional data cube, analysis-ready data, through images obtained by satellites. In addition to data products, the BDC also has systems created to facilitate access to its data collection. Among the services offered, the SpatioTemporal Asset Catalog (STAC) stands out, focused on cataloging image metadata from sensors. In this sense, this work produced data visualization features from the STAC.py library. For viewing the BDC collections, a new class was created to render information in HTML format. They were used the Matplotlib, Earthpy, and Holoviews libraries, both for visualization of metadata summarizations, as well as for visualization of images and histograms.*

Resumo. *O projeto Brazil Data Cube (BDC) é uma iniciativa de produção de cubos de dados multidimensionais, prontos para análise, a partir de imagens de satélites. Além dos produtos de dados, o BDC também tem criado sistemas para facilitar o acesso ao seu acervo de dados. Dentre os serviços oferecidos, destaca-se o SpatioTemporal Asset Catalog (STAC), focado na catalogação de metadados de imagens provenientes dos sensores. Neste sentido, este trabalho produziu funcionalidades de visualização dos dados a partir da biblioteca STAC.py. Para visualização das coleções do BDC, foi criada uma nova classe para renderizar as informações em formato HTML. Foram utilizadas as bibliotecas Matplotlib, Earthpy e Holoviews, tanto para visualização de sumarizações dos metadados, quanto para visualização das imagens e histogramas.*

1. Introdução

A análise de imagens de Sensoriamento Remoto tem se mostrado uma abordagem eficiente para aquisição de dados atualizados da superfície terrestre (GOMEZ, et al., 2016). Atualmente, a comunidade científica tem acesso livre a um amplo catálogo de imagens disponíveis em diferentes resoluções espaciais, temporais e espectrais (FERREIRA et al., 2020). Com o intuito de facilitar as análises de séries temporais de imagens advindas de satélites, os cientistas têm utilizado cubos de dados

prontos para análise (*Analysis-Ready Data* – ARD). ARD pode ser definido como “dados de satélite processados e organizados em uma forma que permite a análise imediata, com mínimo de esforço ao usuário, e com uma interoperabilidade ao longo do tempo” (SIQUEIRA et al., 2019). Além disso, o termo “cubo de dados” (em inglês, *data cube*) refere-se a um conjunto de imagens temporais que apresentam seus pixels alinhados espacialmente (APPEL; PEBESMA, 2019). O processamento de cubos de dados ARD envolve desde a calibração radiométrica até a conversão dos dados para reflectância de superfície (GIULIANI et al., 2017).

Neste sentido, o projeto *Brazil Data Cube* (BDC), desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (INPE), é uma iniciativa criada com o objetivo de produzir uma sequência de cubos de dados multidimensionais, prontos para análise, a partir de imagens de satélites de observação da Terra, focados em médias resoluções espaciais. Além disso, tem como intuito a geração de informações acerca do uso e cobertura do solo a partir de tais cubos de dados, usando *machine learning* e análise de séries temporais. O BDC, atualmente, trabalha com a modelagem dos seguintes tipos de dados advindos de diversos satélites e sensores (Tabela 1).

Tabela 1: Descrição das características dos satélites e sensores usados para geração dos cubos de dados.

Satélite	Sensor	Resoluções
SENTINEL-2	<i>MultiSpectral Image</i> (MSI)	Resolução espacial de 10 m, com 22 bandas, variando de 0,442 μm a 2,202 μm .
LANDSAT-8	<i>Operational Land Imager</i> (OLI)	Resolução espacial de 30 m nas bandas multiespectrais (costal, azul, verde e infravermelho) e 15 m na banda pancromática.
CBERS-4	Câmera de Campo Largo (WFI)	Resolução espacial de 64 m, com 4 bandas espectrais, variando de 0,44 a 0,89 μm .
CBERS-4	Câmera multiespectral regular (MUX)	Resolução espacial de 20 m, com 4 bandas espectrais, também, variando de 0,44 a 0,89 μm .
AQUA/TERRA	<i>Moderate Resolution Imaging Spectroradiometer</i> (MODIS)	Apresenta 36 bandas espectrais, onde 2 delas operam com 250 m de resolução espacial, 5 operam com 500 m e, o restante 1 km.

Fonte: Embrapa Territorial, 2020.

Ademais, o projeto fornece um conjunto de aplicações e serviços que possibilitam que pesquisadores e usuários tenham acesso às imagens a partir de uma infraestrutura computacional. Dentre eles destacam-se: *Web Time Series Services* (WTSS, <https://github.com/brazil-data-cube/wtss.py>), *Web Land Trajectory Service* (WLTS, <https://github.com/brazil-data-cube/wlts>) e *SpatioTemporal Asset Catalog* (STAC, <https://github.com/brazil-data-cube/stac.py>), entre outros. O STAC, produto a ser utilizado no presente trabalho, é um serviço que especifica como os metadados dos recursos geoespaciais são organizados, consultados e disponibilizados dentro da *web*, onde seu principal foco está na catalogação de metadados de imagens provenientes dos sensores orbitais (ZAGLIA, et al., 2019). O STAC é dividido em quatro componentes, conforme a Tabela 2.

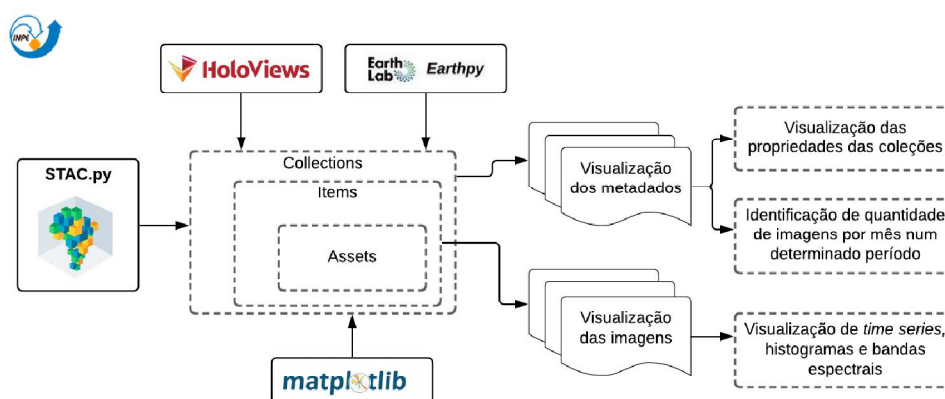
Tabela 2: Descrição dos componentes do serviço STAC, do Brazil Data Cube.

Componente	Descrição
<i>Catalog</i>	Fornecer uma estrutura em formato JSON que permite vincular e acessar coleções e itens presentes dentro do STAC.
<i>Collection</i>	É uma especialização do <i>Catalog</i> que permite o acesso de informações adicionais sobre uma coleção espaço-temporal de dados.
<i>Item</i>	É representado por uma estrutura GeoJSON que fornece os metadados de campos adicionais, ou seja, links para entidades relacionadas e recursos (imagens, thumbnails). Refere-se à menor unidade que descreve o dado.
<i>Asset</i>	Refere-se a um “ativo” espaço-temporal que representa informações sobre a terra, capturadas em um determinado espaço e tempo.

Fonte: ZAGLIA, et al., 2019.

Diante do exposto o objetivo geral do trabalho é produzir novas funcionalidades de visualização dos dados da biblioteca STAC, de modo a facilitar a interação com o usuário, sendo divididas em: [1] Visualização dos metadados e [2] Visualização das imagens, conforme exposto na Figura 1.

Figura 1: Descrição dos procedimentos metodológicos aplicados para visualização dos dados da biblioteca STAC.py.



2. Visualização dos metadados

2.1. Visualização das propriedades das coleções

No BDC, uma coleção é um objeto pertencente à classe *stac.collection*. Os conceitos de classes e de objetos fazem parte de um paradigma de programação conhecido como Programação Orientada a Objeto (POO) (BOOCH; RUMBAUGH; JACOBSON, 2005). Em *Python*, linguagem utilizada no desenvolvimento da biblioteca STAC.py, a criação dos modelos é feita por meio das classes¹.

¹ Uma classe é um conjunto de características e comportamentos que definem o conjunto de objetos pertencentes à essa classe. A classe em si é um conceito abstrato, funcionando como um molde, que se torna concreto a partir da criação de um objeto - as instâncias das classes.

Em relação à biblioteca STAC.py, praticamente todas as classes possuem um método - uma função - que transforma a informação do tipo dicionário (json, dict ou objetos baseados em dicionários) num template HTML. Como resultado, os dados que estão em formato chave:valor podem ser visualizados de forma estruturada quando são chamados em ambientes de computação web, como por exemplo Google Colab ou Jupyter Notebook.

No entanto, essa renderização para HTML não ocorre para visualização de todas as coleções de maneira unificada. Atualmente, para visualização das informações contidas numa coleção, é necessário instanciar um objeto passando como parâmetro a identificação de uma coleção em específico. Sendo assim, para visualizar as informações de forma estruturada de todas as coleções presentes no catálogo do BDC, seria necessário criar um objeto para cada coleção existente.

A falta de uma visualização estruturada e unificada para todas as coleções, motivou a criação de uma classe do tipo *allcollections*. Antes, estas informações eram expressas por meio de dicionários, dificultando o entendimento e utilização de seus atributos. A Figura 2 apresenta como ocorre a instanciação da classe *stac.allcollections* para o objeto chamado *service*.

Figura 2: (a) Dados Originais e (b) Renderização em HTML para visualização.

a)

```
In [5]: service.collections
Out[5]: [{"id": "S2_L1C-1",
  "stac_version": "0.9.0",
  "stac_extensions": ["commons", "datacube", "version"],
  "title": "Sentinel-2 - MSI - Level-1A",
  "version": 1,
  "deprecated": false,
  "description": "This Image Collection contains the Images from the collection S2_L1C Images processed to Surface Reflectance through Sen2Cor v2.0.0.",
  "license": "",
  "properties": {"origin": "EOS",
  "orbitbands": [{"name": "MSI",
  "common_name": "coastal",
  "description": "",
  "srs": "0-0",
  "max": 10000.0,
  "nodata": "0-0",
  "scale": 0.0001,
  "center_wavelength": 0.4437,
  "full_width_half_max": 0.092,
  "data_type": "int16"}]}
```

b)

```
In [3]: service = stac.STAC('https://brasildatacube.dpi.inpe.br/stac/', access_token='change-me')
In [7]: type(service.collections)
Out[7]: stac.allcollections.AllCollections
In [8]: service.collections
Out[8]:
```

S2_L1C-1	+
S2_MSI_L2_SR_LASRC-1	+
LC8SR-1	+
LC8_DN-1	+
MOD13Q1-6	+

2.2. Visualização da contagem de imagens de uma coleção

As imagens presentes no *Brazil Data Cube* se encontram dentro de coleções. Estas são separadas por produtos dos satélites. Para selecionar imagens de determinado local é preciso primeiro realizar a seleção de uma coleção e filtrar por características desejadas. Após selecionadas é possível verificar a quantidade de cenas disponíveis. Isto

é possibilitado pelas informações disponíveis em forma de dicionário nas propriedades da classe *Items*. Nele estão presentes informações sobre a cena, incluindo o *tile* do qual as imagens pertencem, nomes das bandas disponíveis e sua data de aquisição.

Após a definição das imagens desejadas, é realizado um *loop* onde são armazenadas as datas das imagens de acordo com seu mês e ano de aquisição. As informações podem ser representadas por outras bibliotecas e em intervalos de tempo determinado.

3. Visualização das imagens

Para a visualização das séries temporais de imagens, seus histogramas e bandas espectrais foram utilizadas três bibliotecas: *Matplotlib*, *Earthpy* (Figura 3) e *Holoviews* (Figura 4).

Figura 3: Exemplo de visualização das bandas em composição colorida, usando *Earthpy* (à esquerda) e *Matplotlib* (à direita).

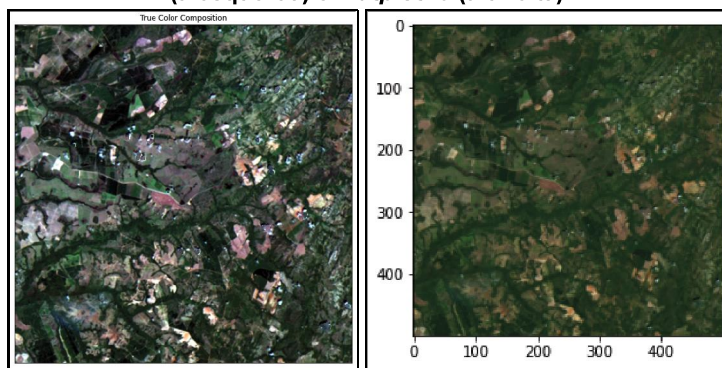
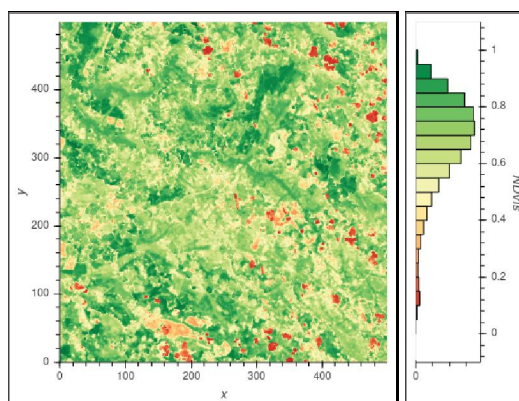


Figura 4: Exemplo de visualização dos dados de NDVI e seus histogramas, usando *Holoviews*.



A *Matplotlib* é uma biblioteca abrangente, destinada para a criação de visualizações estáticas, animadas e interativas em *Python* (<https://matplotlib.org/>). Já o *Holoviews* é uma biblioteca em *Python* de código aberto projetada para tornar a análise e visualização de dados de modo simples. Com o *Holoviews*, é possível expressar, em poucas linhas de código, o que o usuário deseja fazer, permitindo assim uma

concentração maior naquilo que deseja-se explorar e transmitir (<https://holoviews.org/>). Por fim, a *EarthPy* é um pacote *Python* que torna mais fácil trabalhar com dados raster espaciais e vetoriais usando ferramentas de código aberto, no qual seu objetivo é tornar o trabalho com dados geoespaciais mais fácil e intuitivo para os cientistas (<https://earthpy.readthedocs.io/>).

4. Conclusão

O presente trabalho teve como objetivo criar um protótipo de extensão da biblioteca STAC.py e permitir diferentes maneiras de visualização de dados. Foram testadas classes e bibliotecas para contribuir ao projeto. A nova classe do tipo *allcollections* criada permite uma diferente visualização das informações das coleções, facilitando a identificação de metadados para determinado recorte espaço-temporal. A biblioteca *Holoviews* permitiu representar a imagem e seu histograma, auxiliando análises visuais, além de permitir visualizações multitemporais. O *Earthpy* facilitou a visualização de dados matriciais. Por fim, as extensões criadas auxiliaram no acesso de determinadas informações presentes no BDC. Futuros trabalhos podem ser feitos nesta linha, para complementar e auxiliar o projeto. Neste sentido, destaca-se a importância da visualização de dados no contexto da Era do *Big Data*, a qual proporciona uma apresentação gráfica da informação e uma compreensão qualitativa dos conteúdos informativos, possibilitando o reconhecimento de padrões, tendências e relações que existem entre os grupos de dados. Além disso, em virtude do grande volume de produtos gerados a cada dia, a identificação e organização de metadados se mostra relevante, pois permite a simplificação na descrição dos recursos e o entendimento das várias facetas que existem em um ativo de informação.

Referências bibliográficas

- Appel, M.; Pebesma, E. On-Demand Processing of Data Cubes from Satellite Image Collections with the Gdalcubes Library. *Data* 2019, 4, 92.
- Ferreira, K. R.; Queiroz, G. R.; Vinhas, L.; Marujo, R. F. B.; Simoes, R. E. O.; Picoli, M. C. A.; Camara, G.; Cartaxo, R.; Gomes, V. C. E.; Santos, L. A.; Sanchez, A. H.; Arcanjo, J. S.; Fronza, J. G.; Noronha, C. A.; Costa, R. W.; Zaglia, M. C.; Ziotti, F.; Korting, T. S.; Soares, A. R.; Chaves, M. E. D.; Fonseca, L. M. G. Earth Observation Data Cubes for Brazil: Requirements, Methodology and Products. *Remote Sensing*, 2020, 12, 4033, 1-19.
- Giuliani, G.; Chatenoux, B.; De Bono, A.; Rodila, D.; Richard, J.P.; Allenbach, K.; Dao, H.; Peduzzi, P. Building an Earth Observations Data Cube: Lessons Learned from the Swiss Data Cube (SDC) on Generating Analysis Ready Data (ARD). *Big Earth Data* 2017, 1, 100–117.
- Siqueira, A.; Tadono, T.; Rosenqvist, A.; Lacey, J.; Lewis, A.; Thankappan, M.; Szantoi, Z.; Goryl, P.; Labahn, S.; Ross, J.; et al. CEOS Analysis Ready Data For Land—An Overview on the Current and Future Work. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 5536–5537.
- Zaglia, M. C.; Vinhas, L.; Queiroz, G. R.; Simoes, R. Catalogação de Metadados do Cubo de Dados do Brasil com o SpatioTemporal Asset Catalog. In Proceedings of the XX GEOINFO, São José dos Campos, SP, Brasil, 11-13 November, 2019, p. 280-285.