

Catálogo de Metadados do Cubo de Dados do Brasil com o SpatioTemporal Asset Catalog

Matheus C. Zaglia¹, Lúbia Vinhas¹, Gilberto R. de Queiroz¹, Rolf Simoes¹

¹Divisão de Processamento de Imagens – Instituto Nacional de Pesquisas Espaciais (INPE)

Av. dos Astronautas, 1758 – 12227-010 – São José dos Campos – SP – Brazil

{matheus.zaglia, lubia.vinhas, gilberto.queiroz, rolf.simoes}@inpe.br

Abstract. *Technological advances in remote sensing in recent decades have resulted in the generation of large amounts of Earth observation data. The use of techniques that organize satellite images in the form of data cubes has been fundamental to enable the processing and analysis of big data. These data cubes have metadata that must be cataloged in order to discover and access your images. This paper aims to present the metadata cataloging system of Brazil Data Cube, developed by INPE, using the SpatioTemporal Asset Catalog standard.*

Resumo. *Os avanços tecnológicos na área de sensoriamento remoto nas últimas décadas têm resultado na geração de grandes volumes de dados de observação da Terra. A utilização de técnicas que organizam as imagens de satélites em forma de cubos de dados tem se mostrado fundamental para possibilitar o processamento e análise de grandes volumes de dados. Estes cubos de dados possuem metadados que devem ser catalogados de forma que seja possível descobrir e acessar suas imagens. Este trabalho tem como objetivo apresentar o sistema de catalogação de metadados do Cubo de dados do Brasil, em desenvolvimento pelo INPE, utilizando o padrão SpatioTemporal Asset Catalog.*

1. Introdução

Os impactos das atividades humanas no ambiente têm mobilizado esforços da comunidade de pesquisadores e de lideranças no mundo. O monitoramento da cobertura terrestre é uma das principais formas de apoiar políticas públicas para o cumprimento de metas e acordos de proteção ao meio ambiente. Hoje, centenas de sensores a bordo de satélites são capazes de capturar dados com cobertura espacial global, de forma consistente e com uma taxa de revisita cada vez mais alta. Essas imagens são utilizadas em diferentes tipos de aplicações tais como o monitoramento de desastres [Reis et al. 2011], o monitoramento de desmatamento [Valeriano et al. 2004] e classificação de uso e cobertura da terra [Costa et al. 2015].

Todo esse grande volume de dados tem demandado por inovações tecnológicas que permitam a organização e a análise sistemática desse material para os mais diversos fins. Mais recentemente, a comunidade científica tem voltado sua atenção para o conceito de Cubo de Dados de Observação da Terra (EODC) que traduz-se para tecnologias capazes de responder a essas demandas [Giuliani et al. 2017, Camara et al. 2018,

Appel and Pebesma 2019]. Uma das soluções de EODC mais conhecidas é o cubo de dados australiano [Lewis et al. 2017]. No Brasil, o Instituto Nacional de Pesquisas Espaciais iniciou um projeto para desenvolver o Cubo de Dados do Brasil.

Tradicionalmente, os acervos de imagens de sensoriamento remoto disponibilizam a seus usuários acesso a imagens individuais, as quais nem sempre são comparáveis entre si ao longo do tempo. Em um cubo de dados, todas as imagens de uma mesma cena são comparáveis no tempo, o que possibilita novas técnicas de monitoramento e análise da cobertura da Terra [Camara et al. 2018]. Uma das tecnologias centrais para que um cubo de dados funcione é o padrão de catalogação das imagens. Ao longo dos anos, diversos padrões de catalogação, que permitem a busca e a recuperação dos dados foram criados, entre eles o OpenSearch [Clinton 2018] e o Open Geospatial Consortium Catalogue Service (OGC CSW) [OGC 2016]. Mais recentemente, um grupo de pesquisadores propôs e tem desenvolvido um novo padrão de catalogação, considerado mais simples, o SpatioTemporal Asset Catalog (STAC) [Radiant Earth Foundation 2019].

Este trabalho tem como objetivo explorar o padrão STAC para a catalogação dos metadados do Cubo de Dados do Brasil (BDC) e apresentar uma implementação do serviço web que permite a busca, a recuperação e acesso aos metadados catalogados.

2. SpatioTemporal Asset Catalog (STAC)

O STAC é um padrão que especifica como metadados de recursos geoespaciais (por exemplo, imagens de satélite, arquivos de feições, dados de efemérides, *thumbnail*) são organizados, consultados e disponibilizados na *web*. Seu principal foco é a catalogação de metadados de imagens provenientes de observações da Terra por sensores orbitais.

Diferente dos padrões OpenSearch e OGC CSW, que adotam XML¹ para representação dos metadados, o STAC adota o formato JSON², em geral, mais simples para a programação de aplicações na *web* e integração com outras aplicações. O STAC é organizado em quatro componentes:

- **STAC Item:** Representado por uma *Feature* em GeoJSON com alguns campos adicionais, *links* para entidades relacionadas e recursos (imagens, *thumbnails*). Menor unidade que descreve o dado a ser descoberto.
- **STAC Catalog:** Estrutura de um grafo em formato JSON composta por *links* para itens, catálogos e coleções STAC.
- **STAC Collection:** Composta por uma coleção de STAC Items disponibilizados pelo provedor. Possui dados sobre extensão temporal e espacial, licença de uso, palavras-chaves, entre outros. Permite a descoberta em um nível mais alto do que itens individuais, pois descreve conjuntos de dados a partir de características comuns da coleção.
- **STAC API:** API RESTful que estende as capacidades da OGC API - Feature³ com duas rotas adicionais relacionadas ao STAC. Elas adicionam navegação aos catálogos e coleções, além de uma rota STAC REST para consulta, onde são retornados somente itens que correspondem a consulta do usuário.

¹<https://www.w3.org/XML/>

²<https://www.json.org/>

³https://github.com/engeospatial/WFS_FES

O acesso aos recursos representados pelos objetos STAC pode ser realizado através de catálogos estáticos, que são arquivos JSON navegáveis através de *links*, ou por catálogos dinâmicos, através de requisições a um serviço *web* que adote a especificação STAC API.

3. Cubo de Dados do Brasil (BDC)

Para explorar a utilização do padrão STAC no BDC foi desenvolvido um protótipo de um serviço *web* chamado BDC-STAC. Sua implementação utiliza o micro *framework* Flask⁴ do Python⁵ para o tratamento das requisições e respostas HTTP. A partir das requisições são realizadas consultas em um banco de dados MySQL⁶ onde estão contidos os metadados que descrevem o BDC (Figura 1).

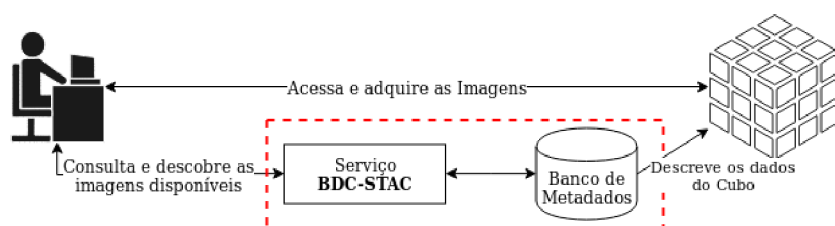


Figura 1. Arquitetura do serviço *web* BDC-STAC.

Para atender as necessidades do BDC foi criada uma extensão que define propriedades específicas (Tabela 1). Estas propriedades são utilizadas em STAC Collections e STAC Items. A identificação das propriedades para a extensão BDC é realizada através da utilização do prefixo "bdc:".

Tabela 1. Propriedades da extensão BDC

Propriedade	Objeto STAC	Valor	Descrição
bands	STAC Collection	[texto]	Bandas disponíveis no cubo
time_aggregations	STAC Collection	[texto]	Tipos de processamento para escolha do pixel
tiles	STAC Collection	[texto]	Quais <i>tiles</i> possuem imagens no cubo.
tschema	STAC Collection	texto	"A" para cubos anuais, "M" para cubos mensais, "S" para cubos sazonais.
tstep	STAC Collection	número	Dias entre as imagens
grs	STAC Collection	texto	Grade de referenciamento utilizado no cubo
time_aggregation	STAC Item	texto	Tipo de processamento para escolha do pixel da imagem
tile	STAC Item	texto	Qual <i>tile</i> a imagem pertence

Na Figura 2 são apresentadas as rotas, parâmetros de consulta e métodos HTTP implementados no serviço BDC-STAC para atender as especificações STAC API. Além dos parâmetros padrões, foram criados dois exclusivos do BDC: *time_aggregation* e *bands*.

⁴<https://flask.palletsprojects.com/>

⁵<https://python.org>

⁶<https://www.mysql.com/>

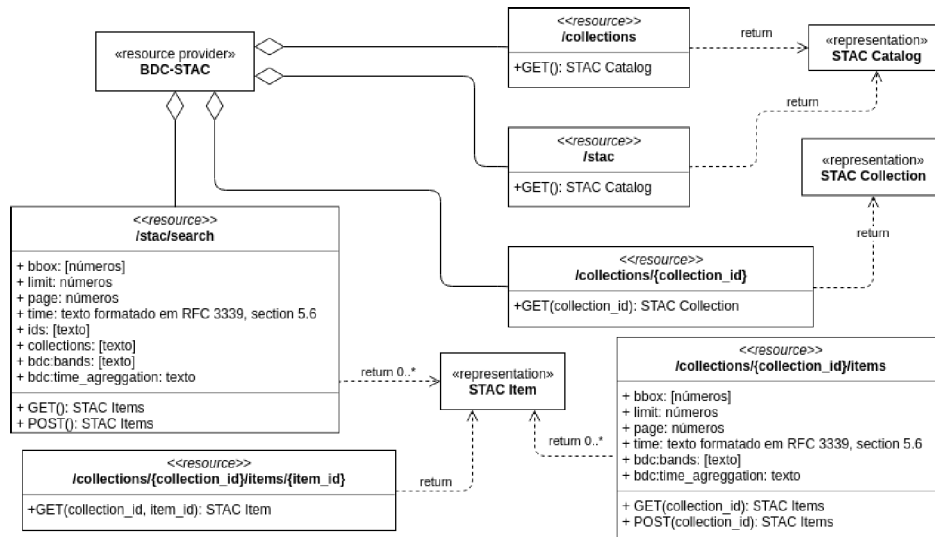


Figura 2. Mapeamento da STAC API implementada pelo serviço BDC-STAC.

4. Resultados

Nesta seção serão apresentados alguns resultados provenientes de acessos realizados as rotas do serviço BDC-STAC de acordo com a implementação da STAC API.

A Figura 3 mostra o retorno do acesso a rota `/stac`, onde é retornado um STAC Catalog que contém todos os cubos de dados disponíveis no BDC.

```

{
  "description": "Brazil Data Cubes Catalog",
  "id": "bdc",
  "stac_version": "0.7",
  "links": [
    { "href": "http://localhost:5050/collections", "rel": "self" },
    { "href": "http://localhost:5050/collections/C64m", "rel": "child",
      ↪ "title": "C64m" },
    { "href": "http://localhost:5050/collections/S10m", "rel": "child",
      ↪ "title": "S10m" },
    { "href": "http://localhost:5050/collections/S1016d", "rel": "child", "title":
      ↪ "S1016d" },
    { "href": "http://localhost:5050/collections/L3016d", "rel": "child", "title":
      ↪ "L3016d" },
    ...
  ]
}

```

Figura 3. STAC Catalog disponível na rota `/stac`

O acesso a rota `/collections/C64m` retorna um STAC Collection que descreve o cubo de dados C64m (Figura 4).

Na Figura 5 é possível ver o resultado de uma requisição a rota `/stac/search?collections=C64m&bdc.bands=ndvi,evi&bdc.time_aggregation=MEDIAN`. Nessa rota são utilizados filtros nas consultas para descobrir os dados somente do cubo C64m onde o método seleção de pixel seja MEDIAN e mostre somente as bandas ndvi e evi.

```
{
  "id": "C64m",
  "stac_version": "0.7",
  "description": "C64m datacube with products from CB4_AWFI(Sattelite/Sensor) with
  ↳ blue,green,red,nir,evi,ndvi,quality bands.",
  "license": null,
  "extent": {
    "spatial": [{"-19.98430824279785", "-54.7285270690918", "-41.86650085449219",
    ↳ "-12.186923027038574"}],
    "time": ["2016-09-01T00:00:00", "2019-05-31T00:00:00"]
  },
  "properties": {
    "bdc:time_aggregations": ["MEDIAN", "MERGED", "SCENE", "STACK"],
    "bdc:bands": ["blue", "green", "red", "nir", "evi", "ndvi", "quality"],
    "bdc:tiles": ["042050", "042051", "043051", "043052", "044049", "045049"],
    "bdc:tstep": "M",
    "bdc:tstep": 16,
    "bdc:grs": "aea_500k"
  },
  "links": [
    {"href": "http://localhost:5050/collections/C64m", "rel": "self"},
    {"href": "http://localhost:5050/collections/C64m/items", "rel": "items"},
    {"href": "http://localhost:5050/collections", "rel": "parent"},
    {"href": "http://localhost:5050/collections", "rel": "root"},
    {"href": "http://localhost:5050/stac", "rel": "root"}
  ]
}
```

Figura 4. STAC Collection disponível na rota /collections/C64m

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "id": "C64m_042050_2016-09-01_MEDIAN",
      "collection": "C64m",
      "bbox": [-54.6817, -16.1408, -51.4231, -14.1867],
      "geometry": { "type": "Polygon",
        "coordinates": [[[-54.6357, -14.1867], [-54.6817,-16.1084], [-51.4372,
        ↳ -16.1408], [-51.4231, -14.2188], [-54.6357, -14.1867]]]]},
      "properties": {
        "bdc:time_aggregation": "MEDIAN",
        "bdc:tile": "042050",
        "datetime": "2016-09-01T00:00:00"
      },
      "assets": {
        "thumbnail": {"href": "http://brazildatacube.dpi.inpe.br/Repository/Mosaic/C64m
        ↳ /042050/2016-09-01-2016-09-30/C64m_042050_2016-09-01_MEDIAN.png"},
        "evi": {"href": "http://brazildatacube.dpi.inpe.br/Repository/Mosaic/C64m/04205
        ↳ /0/2016-09-01-2016-09-30/C64m_042050_2016-09-01_evi_MEDIAN.tif"},
        "nvi": {"href": "http://brazildatacube.dpi.inpe.br/Repository/Mosaic/C64m/04205
        ↳ /0/2016-09-01-2016-09-30/C64m_042050_2016-09-01_ndvi_MEDIAN.tif"},
        ...
      },
      "links": [...]
    },
    ...
  ]
}
```

Figura 5. STAC Item disponível na rota /collections/C64m/items

5. Considerações finais

Este trabalho apresentou a utilização do padrão SpatioTemporal Asset Catalog (STAC) para catalogação dos metadados das imagens dos cubos de dados do projeto Cubo de Dados do Brasil (BDC). O STAC se mostrou uma maneira efetiva de catalogação, descoberta e recuperação dos metadados, sendo de fácil desenvolvimento. Além disso, é possível a criação de outros serviços mais específicos para consumir os dados catalogados pelo STAC.

O trabalho também apresentou uma extensão do STAC para atender demandas do BDC. O código fonte do serviço desenvolvido encontra-se disponível na forma de software livre em <https://github.com/brazil-data-cube/bdc-stac>.

Como trabalho futuro, será desenvolvido um serviço STAC para o BDC usando tecnologias de computação em nuvem, como arquiteturas *serverless*, para possibilitar maior escalabilidade do serviço.

Referências

- Appel, M. and Pebesma, E. (2019). On-Demand Processing of Data Cubes from Satellite Image Collections with the gdalcubes Library. *Data*, 4(3):92.
- Camara, G., Queiroz, G., Vinhas, L., et al. (2018). The e-sensing architecture for big earth observation data analysis.
- Clinton, D. (2018). Opensearch. <https://github.com/dewitt/opensearch>. Acesso em: 20/09/2019.
- Costa, W., Fonseca, L., and Korting, T. (2015). Classifying grasslands and cultivated pastures in the brazilian cerrado using support vector machines, multilayer perceptrons and autoencoders. *11th International Conference in Machine Learning, MLDM, Hamburg*.
- Giuliani, G., Chatenoux, B., De Bono, A., et al. (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.
- Lewis, A., Oliver, S., Lymburner, L., et al. (2017). The Australian Geoscience Data Cube — Foundations and lessons learned. *Remote Sensing of Environment*, 202:276–292.
- OGC (2016). OGC Catalogue Services 3.0 - General Model. <http://docs.opengeospatial.org/is/12-176r7/12-176r7.html>. Acesso em: 20/09/2019.
- Radiant Earth Foundation (2019). Spatiotemporal asset catalog.
- Reis, J. B. C., Santos, T. B., and Lopes, E. S. S. (2011). Monitoramento em tempo real de eventos extremos na região metropolitana de são paulo – uma aplicação com o sismaden. *14° SIMPÓSIO BRASILEIRO DE GEOGRAFIA FÍSICA APLICADA, 2011, Dourados, MS*.
- Valeriano, D. M., Mello, E. M., Moreira, J. C., et al. (2004). Monitoring tropical forest from space: the prodes digital project. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 35:272–274.