



EDC FORUM 2017

Big Data Analytics & GIS – September 21-22, Münster

**Please do not unplug any cables, the desktop
computers are alarmed!**

HANDS-ON WORKSHOP: PROCESSING EARTH OBSERVATION TIME SERIES WITH SCIDB AND R

Marius Appel

marius.appel@uni-muenster.de



ifgi

Institute for Geoinformatics
University of Münster



WWU
MÜNSTER

OBJECTIVES

- provide an overview of SciDB in practical applications and
 - how it can be used to analyze Earth observation data
 - where it has advantages / limitations
- see how accessible it is without previous experiences with the technology
- We will look at SciDB mostly from the perspective of a „data scientist“

WHO ARE YOU?

OUTLINE (MORNING)

1. Sample dataset
2. Introduction and demonstration of basic concepts in SciDB
 - Architecture
 - Arrays in Detail
 - AFL query language
3. Introduction to the SciDB R client
 - The R package scidb
 - The R package scidbst

OUTLINE (MORNING)

4. Example time series analyses

- Time series extraction
- Image averaging
- Greenest image
- Water classification
- Seasonality assessment

5. Ingestion of EO data

6. Hands-on

ORGANIZATION

- Material online at <https://github.com/appelmar/edcforum2017>
- Please
 - do not remove any data in SciDB except for arrays that you have created during the workshop
 - work with very small subsets if you try out things, all participants share the same SciDB installation!
 - ask questions, contribute ideas!

OUTLINE (AFTERNOON)

Hackthon with open topics:

- More time to experiment with SciDB and the provided Sentinel dataset
- Installation of SciDB and its extensions using Docker
- Implementation of further methods, e.g.: compute difference images from two given dates
- Ingest and analyze other data and implement (e.g., MODIS, SRTM)
- Try out SciDB-Py, e.g. by re-implementing the R examples
- your own ideas!

SENTINEL 2

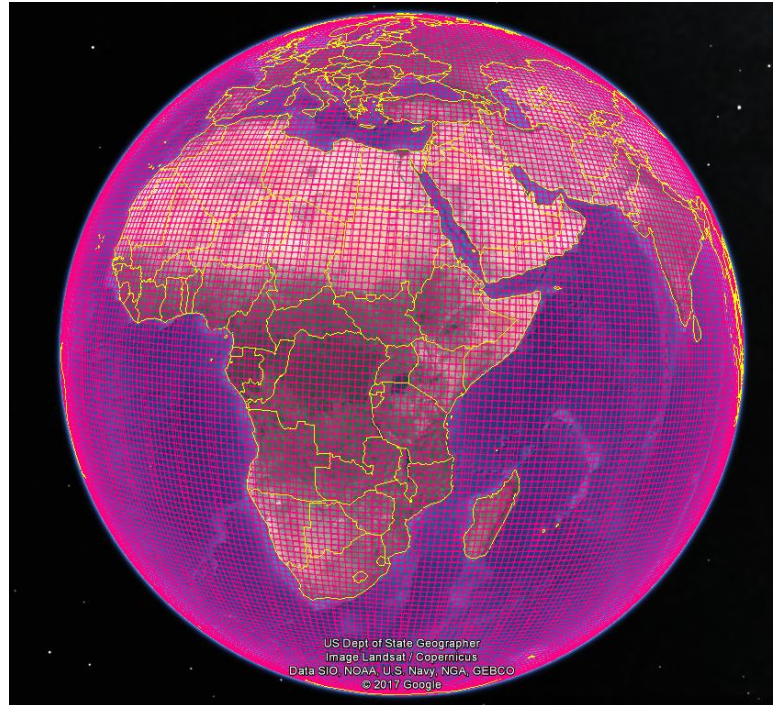
- Two satellites:
 - Sentinel-2A (launched 23 June 2015)
 - Sentinel-2B (launched 07 March 2017)
- complements Landsat and SPOT missions
- part of EU Copernicus program
- 13 Spectral Bands at (10m, 20m, or 60m) spatial resolution
- covers latitudes from 56° south to 83° north
- revisit time: 5 days (with two satellites)

SENTINEL 2

- Applications:
 - Water monitoring
 - Forest / vegetation monitoring and change detection
 - Spatial planning
 - Crop monitoring
 - Emergency Management
- free and open access!

SENTINEL 2

Tiling of orthorectified products (Level 1C / 2A)



https://sentinel.esa.int/documents/247904/1955685/S2A_OPER_GIP_TILPAR_MPC__20151209T095117_V20150622T00000_21000101T000000_B00.kml

SENTINEL 2 – DATA ORGANIZATION

- S2A_OPER_PRD_MSIL1C_PDMC_20160514T213146_R121_V20150810T084516_201508
 - S2A_OPER_PRD_MSIL1C_PDMC_20160514T213146_R121_V20150810T084516_201508
 - AUX_DATA
 - DATASTRIP
 - S2A_OPER_MSI_L1C_DS_EPA_20160514T170317_S20150810T084516_N02.02
 - QI_DATA
 - GRANULE
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KEB_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KEC_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KED_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFB_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_N02.02
 - AUX_DATA
 - IMG_DATA
 - QI_DATA
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFD_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KGB_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KGC_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KGD_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KHB_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KHC_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KHD_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T35KKS_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T35KKT_N02.02
 - S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T35KKU_N02.02
 - AUX_DATA
 - IMG_DATA
 - QI_DATA
 - HTML
 - rep_info

Name	Änderungs...	Typ	Größe
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B01.jp2	14.05.2016 ...	JP2-Datei	3.663 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B02.jp2	14.05.2016 ...	JP2-Datei	119.141 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B03.jp2	14.05.2016 ...	JP2-Datei	123.402 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B04.jp2	14.05.2016 ...	JP2-Datei	129.554 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B05.jp2	14.05.2016 ...	JP2-Datei	32.858 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B06.jp2	14.05.2016 ...	JP2-Datei	32.773 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B07.jp2	14.05.2016 ...	JP2-Datei	32.792 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B08.jp2	14.05.2016 ...	JP2-Datei	129.694 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B8A.jp2	14.05.2016 ...	JP2-Datei	32.999 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B09.jp2	14.05.2016 ...	JP2-Datei	3.662 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B10.jp2	14.05.2016 ...	JP2-Datei	2.281 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B11.jp2	14.05.2016 ...	JP2-Datei	32.995 KB
S2A_OPER_MSI_L1C_TL_EPA_20160514T170317_A000690_T34KFC_B12.jp2	14.05.2016 ...	JP2-Datei	32.927 KB

SENTINEL 2 – DATA ACCESS

Copernicus Open Access Hub

The screenshot displays the Copernicus Open Access Hub interface. The top navigation bar includes the ESA and Copernicus logos, the site name, and user icons. Below the navigation bar is a search bar with the text "Insert search criteria...". The main content area shows a list of search results for Sentinel-2 data. The results are displayed in a table with columns for product ID, download URL, mission, instrument, sensing date, and size. The first result is for product S2A_MSIL2A_20170917T103021_N0205_R108_T32UMC_20170917T103018. The second result is for product S2A_MSIL2A_20170917T103021_N0205_R108_T31UGT_20170917T103018. The third result is for product S2A_MSIL2A_20170917T103021_N0205_R108_T32ULC_20170917T103018. The fourth result is for product S2A_MSIL1C_20170917T103021_N0205_R108_T31UGT_20170917T103018. The fifth result is for product S2A_MSIL1C_20170917T103021_N0205_R108_T32ULC_20170917T103018. The background of the page shows a map of Central Europe with a green rectangular area of interest overlaid, covering parts of Germany and the Netherlands. The map includes labels for various cities and regions, such as Zwolle, Overijssel, Enschede, Osnabrück, Bielefeld, Paderborn, Hamm, Dortmund, Essen, Moers, Bottrop, Düsseldorf, Wuppertal, Remscheid, Köln, Bonn, Aachen, Siegen, Kassel, and Hannover.

Request Done: (footprint: "Intersects(POLYGON((7.515318491836685
51.91618255192779,7.747333946850549 51.91618255192779,7.747333946850549
52.031622892233344,7.515318491836685 52.031622892233344,7.515318491836685
S2A_MSIL2A_20170917T103021_N0205_R108_T32UMC_20170917T103018
Download URL: <https://scihub.copernicus.eu/dhus/odata/v1/Products/'0f22eda7-597b-4e00-9207-f>
Mission: Sentinel-2; Instrument: MSI; Sensing Date: 2017-09-17T10:30:21.026Z; Size: 1.09 GB

S2A_MSIL2A_20170917T103021_N0205_R108_T31UGT_20170917T103018
Download URL: <https://scihub.copernicus.eu/dhus/odata/v1/Products/'011ae730-bdbe-4171-9790->
Mission: Sentinel-2; Instrument: MSI; Sensing Date: 2017-09-17T10:30:21.026Z; Size: 280.83 MB

S2A_MSIL2A_20170917T103021_N0205_R108_T32ULC_20170917T103018
Download URL: <https://scihub.copernicus.eu/dhus/odata/v1/Products/'8f529be1-6b20-4ddc-98c6-t>
Mission: Sentinel-2; Instrument: MSI; Sensing Date: 2017-09-17T10:30:21.026Z; Size: 406.69 MB

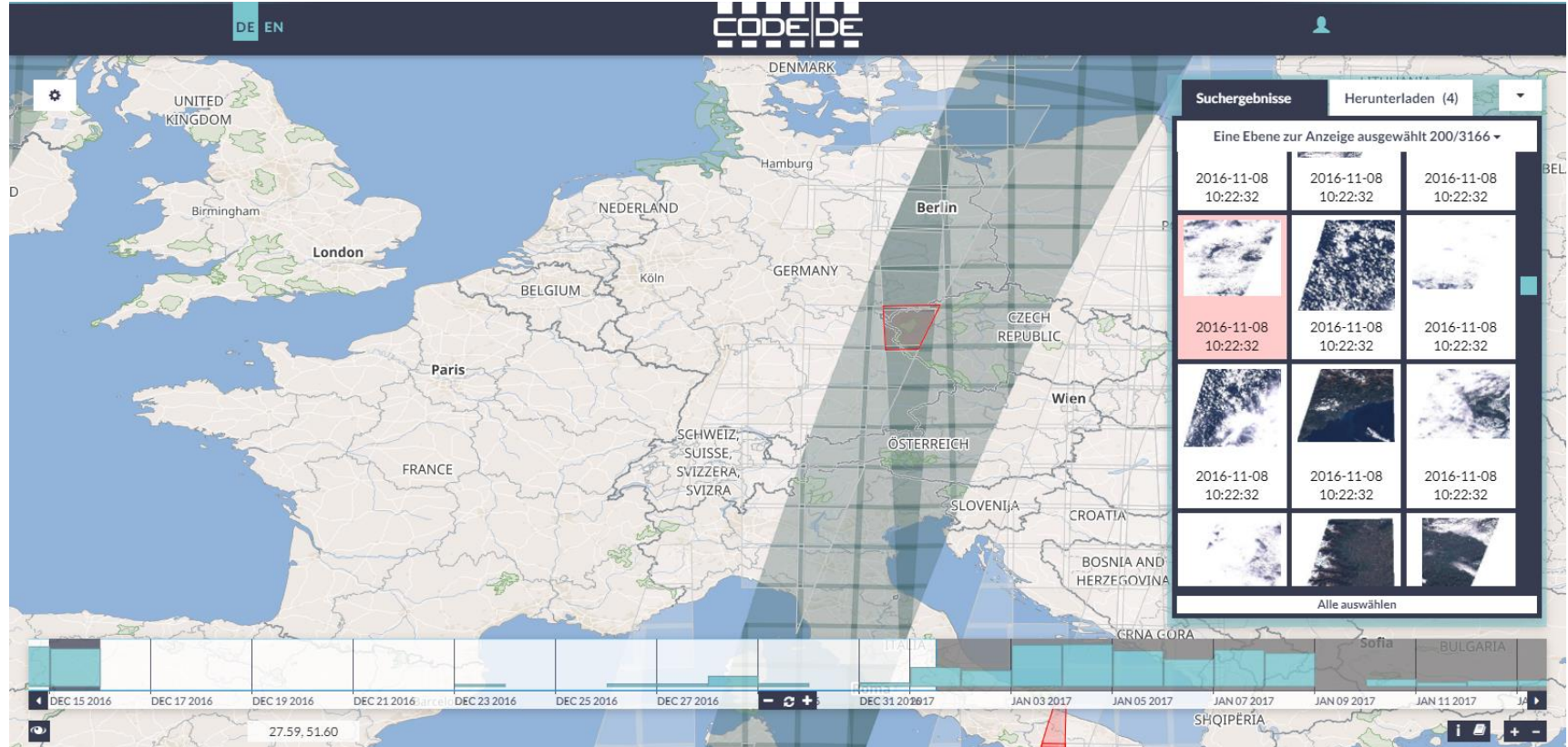
S2A_MSIL1C_20170917T103021_N0205_R108_T31UGT_20170917T103018
Download URL: <https://scihub.copernicus.eu/dhus/odata/v1/Products/'5d4307cb-f356-4577-bea7-t>
Mission: Sentinel-2; Instrument: MSI; Sensing Date: 2017-09-17T10:30:21.026Z; Size: 214.02 MB

S2A_MSIL1C_20170917T103021_N0205_R108_T32ULC_20170917T103018

Products per page: 25 << page: 1 of 16 >> CLOSE

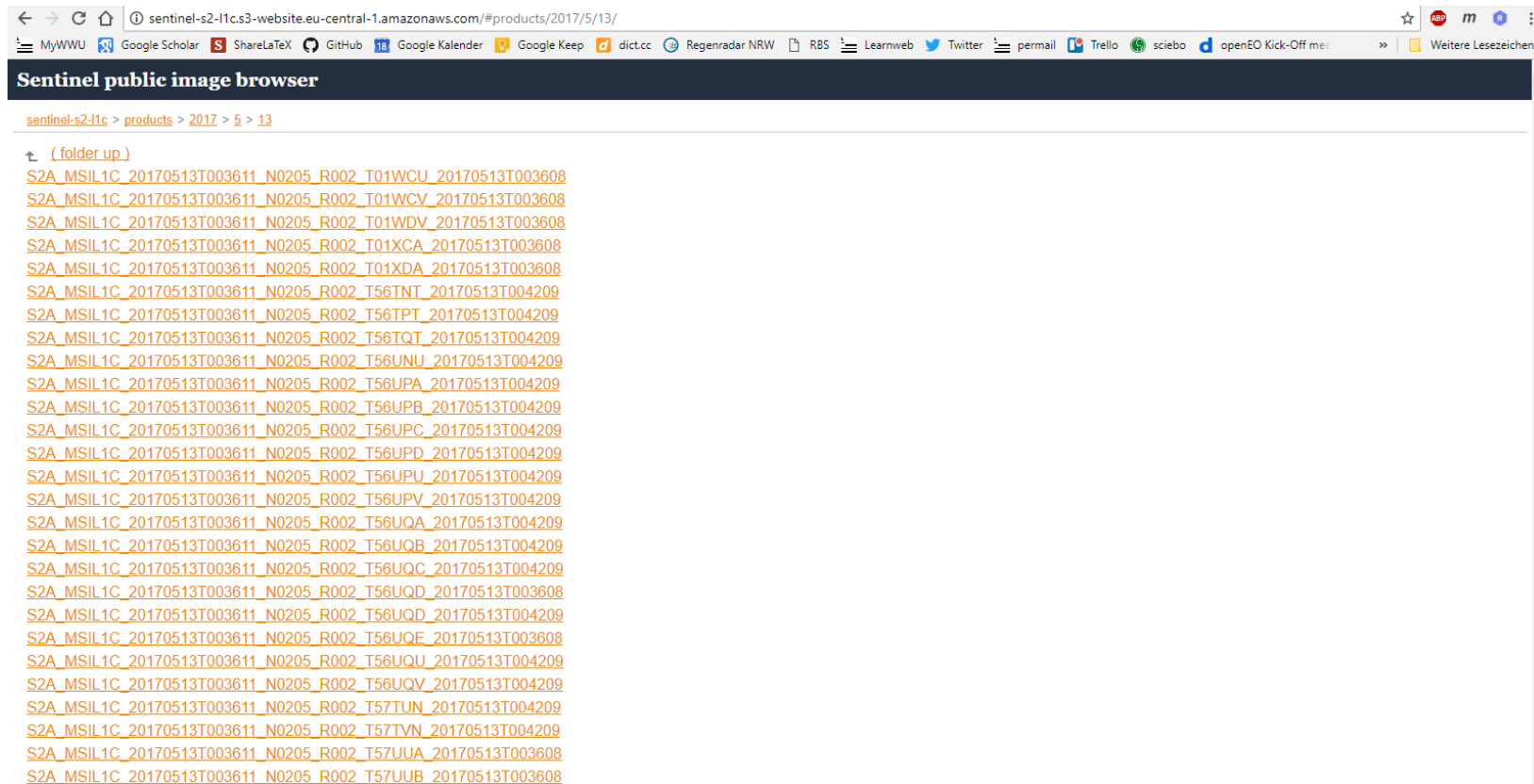
SENTINEL 2 – DATA ACCESS

Copernicus Data and Exploitation Platform – Deutschland (CODE-DE)



SENTINEL 2 – DATA ACCESS

Amazon Web Services



The screenshot shows a web browser window with the URL `sentinel-s2-l1c.s3-website-eu-central-1.amazonaws.com/#products/2017/5/13/`. The browser's address bar and tabs are visible at the top. Below the browser window, a dark blue header bar contains the text "Sentinel public image browser". The main content area displays a breadcrumb trail: `sentinel-s2-l1c > products > 2017 > 5 > 13`. Below this, a link `(folder up)` is shown. The majority of the page is filled with a list of 36 Sentinel-2 image IDs, each on a new line and underlined. These IDs follow a consistent naming convention: `S2A_MSIL1C_20170513T003611_N0205_R002_T01WCU_20170513T003608`, where the last part of the ID (the timestamp) varies for each entry. The list ends with `S2A_MSIL1C_20170513T003611_N0205_R002_T57UUB_20170513T003608`.

sentinel-s2-l1c > products > 2017 > 5 > 13

(folder up)

[S2A_MSIL1C_20170513T003611_N0205_R002_T01WCU_20170513T003608](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T01WCV_20170513T003608](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T01WDV_20170513T003608](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T01XCA_20170513T003608](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T01XDA_20170513T003608](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56TNT_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56TPT_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56TQT_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UNU_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UPA_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UPB_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UPC_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UPD_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UPU_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UPV_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UQA_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UQB_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UQC_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UQD_20170513T003608](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UQD_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UQE_20170513T003608](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UQU_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T56UQV_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T57TUN_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T57TVN_20170513T004209](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T57UUA_20170513T003608](#)

[S2A_MSIL1C_20170513T003611_N0205_R002_T57UUB_20170513T003608](#)

SENTINEL 2 – DATA ACCESS

Google Earth Engine

The screenshot displays the Google Earth Engine web interface. The browser address bar shows the URL <https://code.earthengine.google.com/#>. The search bar contains the text "sentinel". The left sidebar lists various data access functions under the "Scripts" tab, including `ee.data.getDownloadId(params, callback)`, `ee.data.getInfo(id, callback)`, `ee.data.getList(params, callback)`, `ee.data.getMapId(params, callback)`, `ee.data.getTableDownloadId(params, callback)`, `ee.data.getThumbId(params, callback)`, `ee.data.getTileBaseUrl()`, `ee.data.getTileUrl(mapId, x, y, z)`, `ee.data.getValue(params, callback)`, and `ee.data.getXsrfToken()`.

The main editor area shows a script titled "02 - Hello Image *". The script content is as follows:

```
Imports (1 entry)
var s2: ImageCollection "Sentinel-2: MultiSpectral Instrument (MSI), Level-1C"

1
2
3 var s2f = s2.filterDate("2017-07-01", "2017-07-12").select(["B4", "B3", "B2"]);
4
5 Map.addLayer(s2f, {min: 0, max: 2000 });
6
7
```

The right sidebar contains the "Inspector", "Console", and "Tasks" tabs. The "Console" tab is active, displaying the instruction "Use print(...) to write to this console.".

The bottom section of the interface shows a satellite map of a coastal area. The map is overlaid with a grid of colored rectangles, representing the spatial extent of the Sentinel-2 data being accessed. The map includes a "Layers" panel, a "Karte" (Map) button, and a "Satellit" (Satellite) button. The Google logo is visible in the bottom left corner, and the bottom right corner contains copyright information: "Kartendaten © 2017 GeoBasis-DE/BKG (©2009), Google Inst. Geogr. National, Mapa ©Israel, ORION/ME 200 km 1:1000000 Nutzungsbedingungen".

(FUTURE) ACCESS TO EO DATA?

Alternatives to downloading scenes from space agencies?

- dedicated EO data centers / cloud services that provide access at a
 - low-level (file-based): AWS, EODC,
 - high-level: Google Earth Engine, SciDB,
- Processing where the data is stored

SAMPLE DATASET

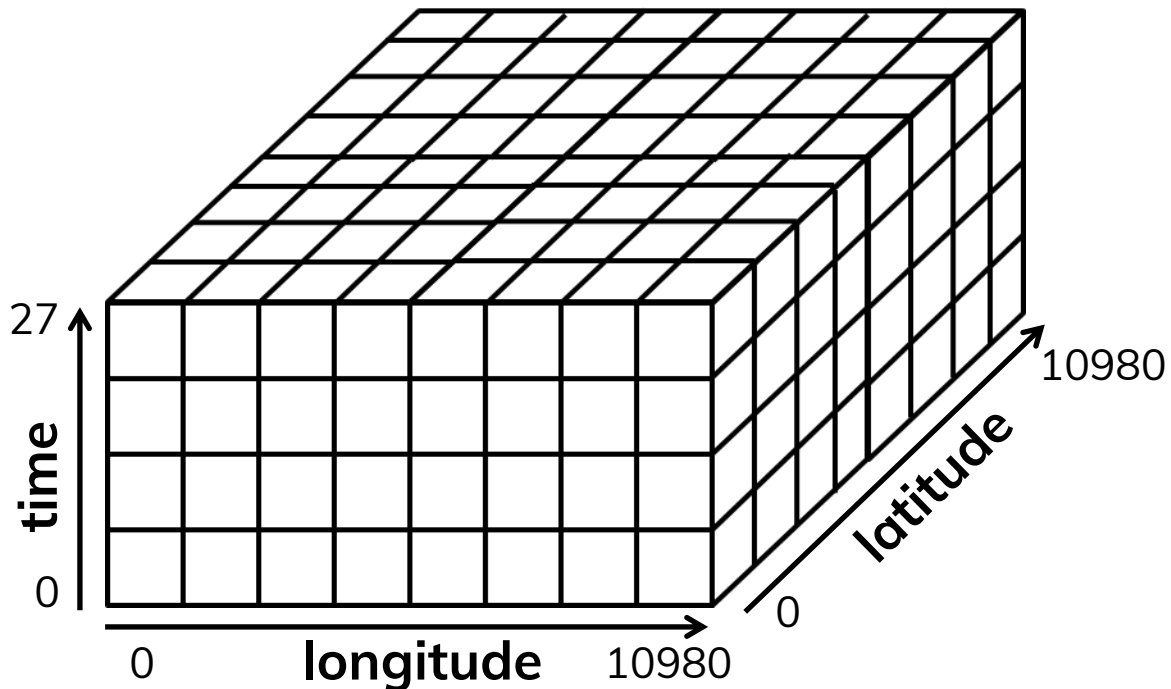
- 26 scenes from Sentinel 2A Level 1C
- Tile 34KGD (Okavango Delta region)
- Captured between 2016-12-12 and 2017-08-29
- approx. 12×10^9 measurements, approx 22 GB

SAMPLE DATASET



SAMPLE DATASET

- SciDB representation has three dimensions (10980 x 10980 x 27 cells)
- Each cell has 4 attributes (bands)
- Cells cover 10m x 10m x 10d
- Two arrays with the same data but different chunk organization:
 - S2_OKAVANGO_S
 - S2_OKAVANGO_T



BASIC SCIDB CONCEPTS

Multidimensional Arrays (n dimensions, m attributes)

$$A: D \rightarrow V$$

$$D \subset D_1 \times D_2 \times \cdots \times D_n$$

$$V \subset V_1 \times V_2 \times \cdots \times V_m$$

- Dimensions are ordered and finite
- SciDB dimensions are always 64bit integers \rightarrow space or time must be transformed
- SciDB attributes can be of any type, and nullable if desired

BASIC SCIDB CONCEPTS

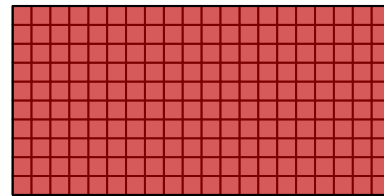
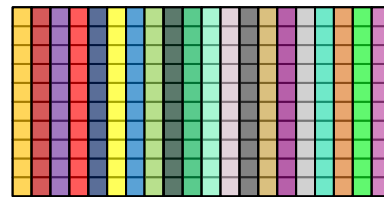
Examples:

- RGB image: two dimensions, three attributes
- multivariate time series: one dimension, multiple attributes
- hyperspectral EO image time series: four dimensions, one attribute
- DEM: two-dimensions, one attribute

BASIC SCIDB CONCEPTS

Chunking

- equally sized rectangular subsets of the data within an array
- specified by number of cells per dimension
- defined during array creation
- optional chunk overlap
- strongly affects scalability and performance



BASIC SCIDB CONCEPTS

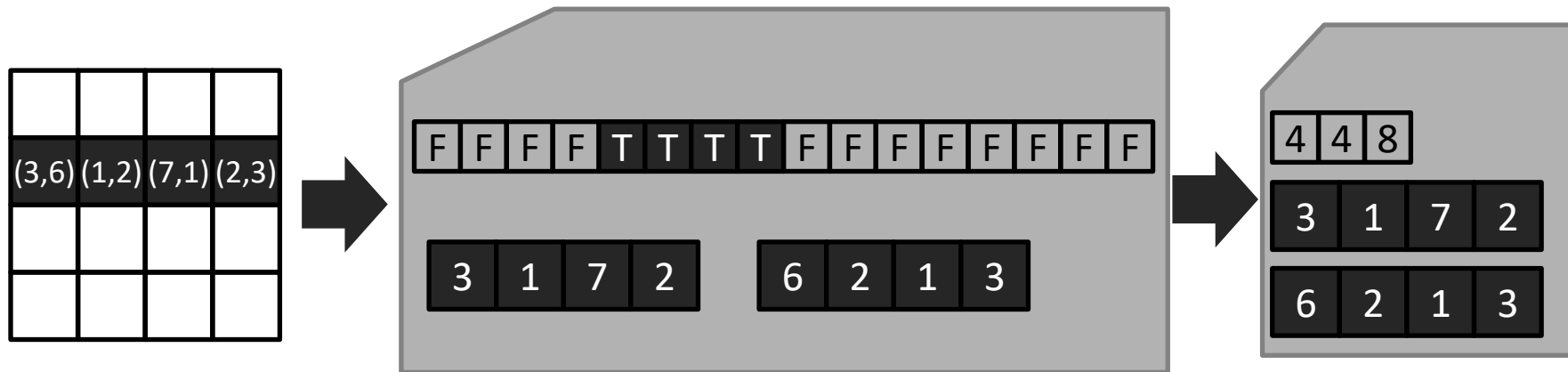
Nearly Sparse Chunk Storage

- **Empty cells:** no value or any of the attributes
- **Empty vs Null:** cell value of a single attribute can be null but still have data for other attributes, null also adds a missing reason code and is stored

BASIC SCIDB CONCEPTS

Nearly Sparse Chunk Storage

- run-length encoded dense bitmaps and
- actual values of attributes for nonempty cells



SCIDB QUERY LANGUAGE AFL

→ `see scripts/1_AFL_basics.md`

SCIDB R CLIENT

→ see `scripts/2_SciDB_R.R`

SCIDBST R PACKAGE

→ see `scripts/3_scidbst.R`

EXAMPLE TASKS

1. Time series extraction
2. Image averaging
3. Greenest image
4. Water classification
5. Seasonality assessment