

Bernhard Bauer-Marschallinger

Remote Sensing Group
 GEO Department
 TU Wien



Greetings from Planet Big Data

Publishing global satellite imagery at the TU Wien Research Data Repository

remote sensing at TU Wien



TU Wien

- GEO Department
 - Remote Sensing Group
 - ~ 20 scientific/student staff
 - Earth observation
 - satellite imagery
 - signal analysis
 - geophysical variables on
 - vegetation
 - soil
 - land cover & water bodies
 - ...
 - <https://www.geo.tuwien.ac.at/>

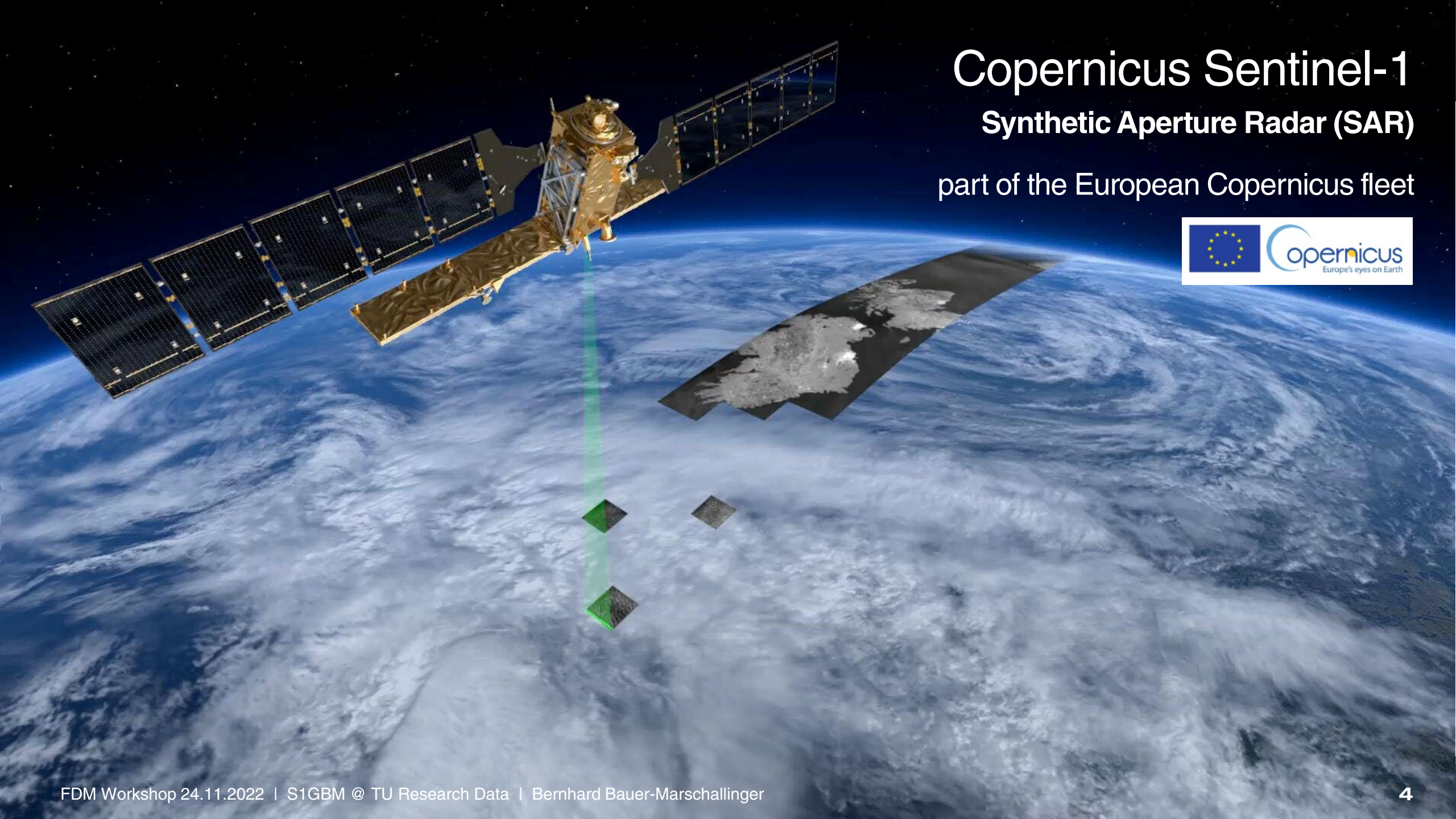


Satellite radar imagery!?

Copernicus Sentinel-1

Synthetic Aperture Radar (SAR)

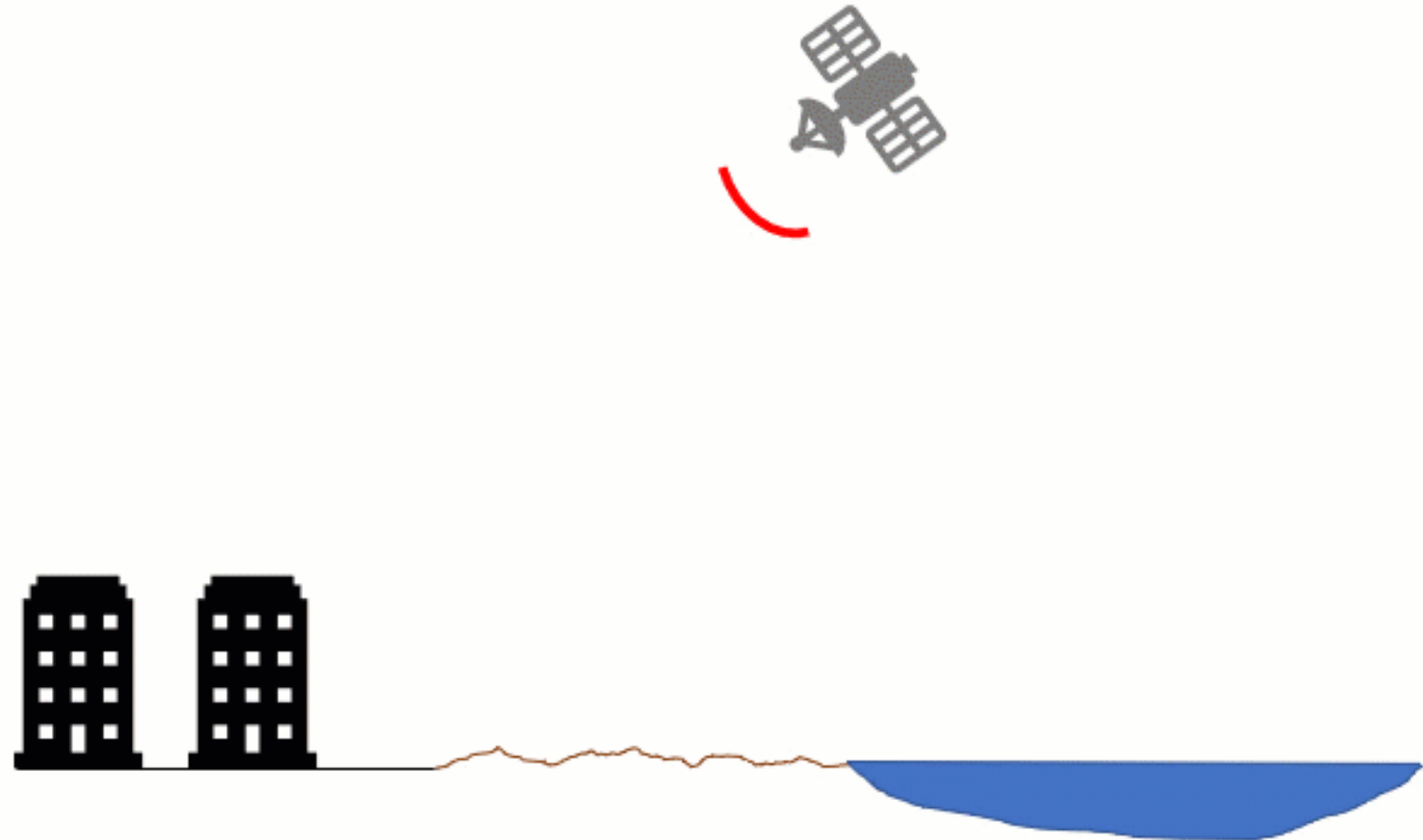
part of the European Copernicus fleet



Synthetic Aperture Radar (SAR) principle

The measurement principle

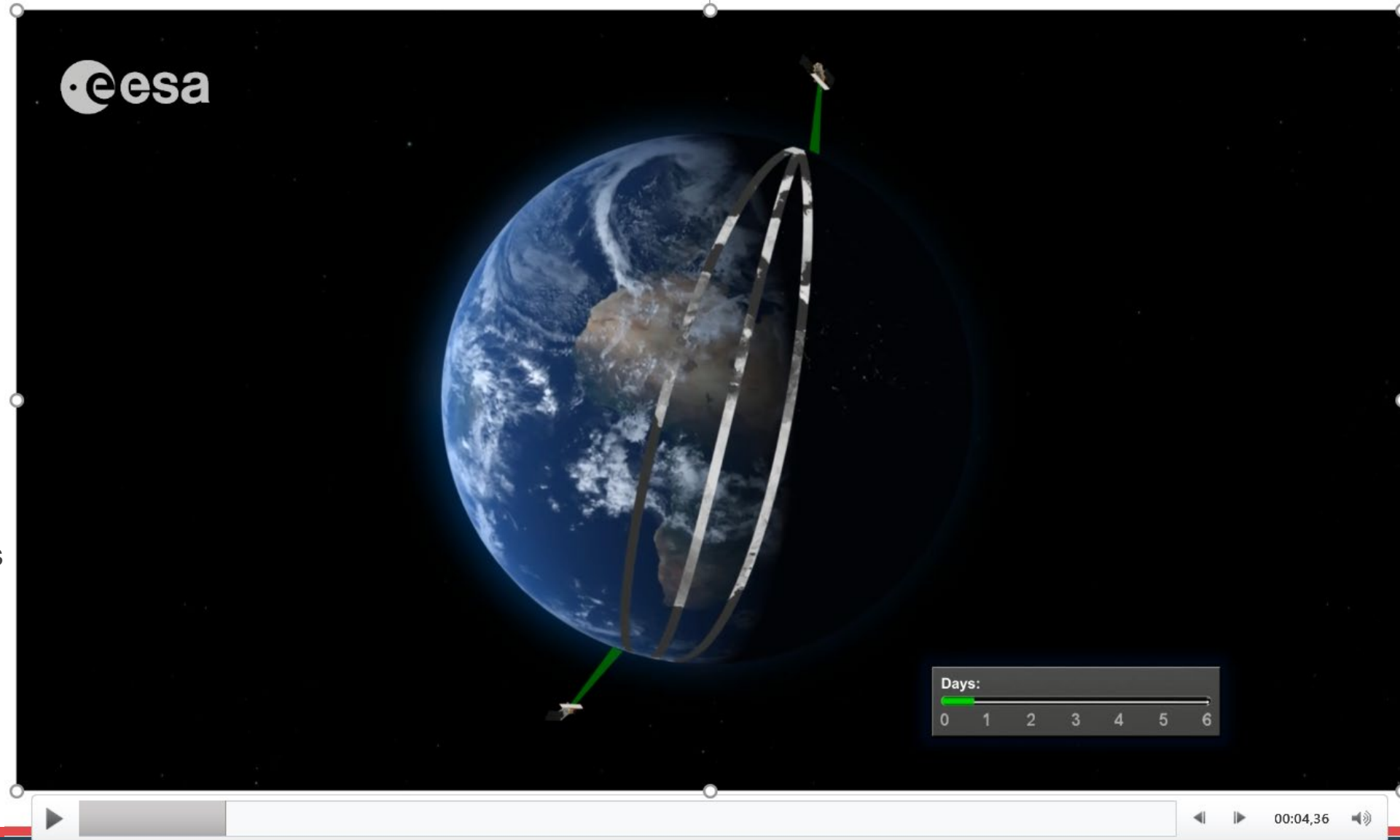
- emitting a **radar pulse** with a **synthetic antenna**
 - synthetic: using the motion of the satellite
- measuring the intensity of the **backscatter** signal
- active systems: independent from weather and light conditions.



Earth observation with Sentinel-1 radar

Application examples

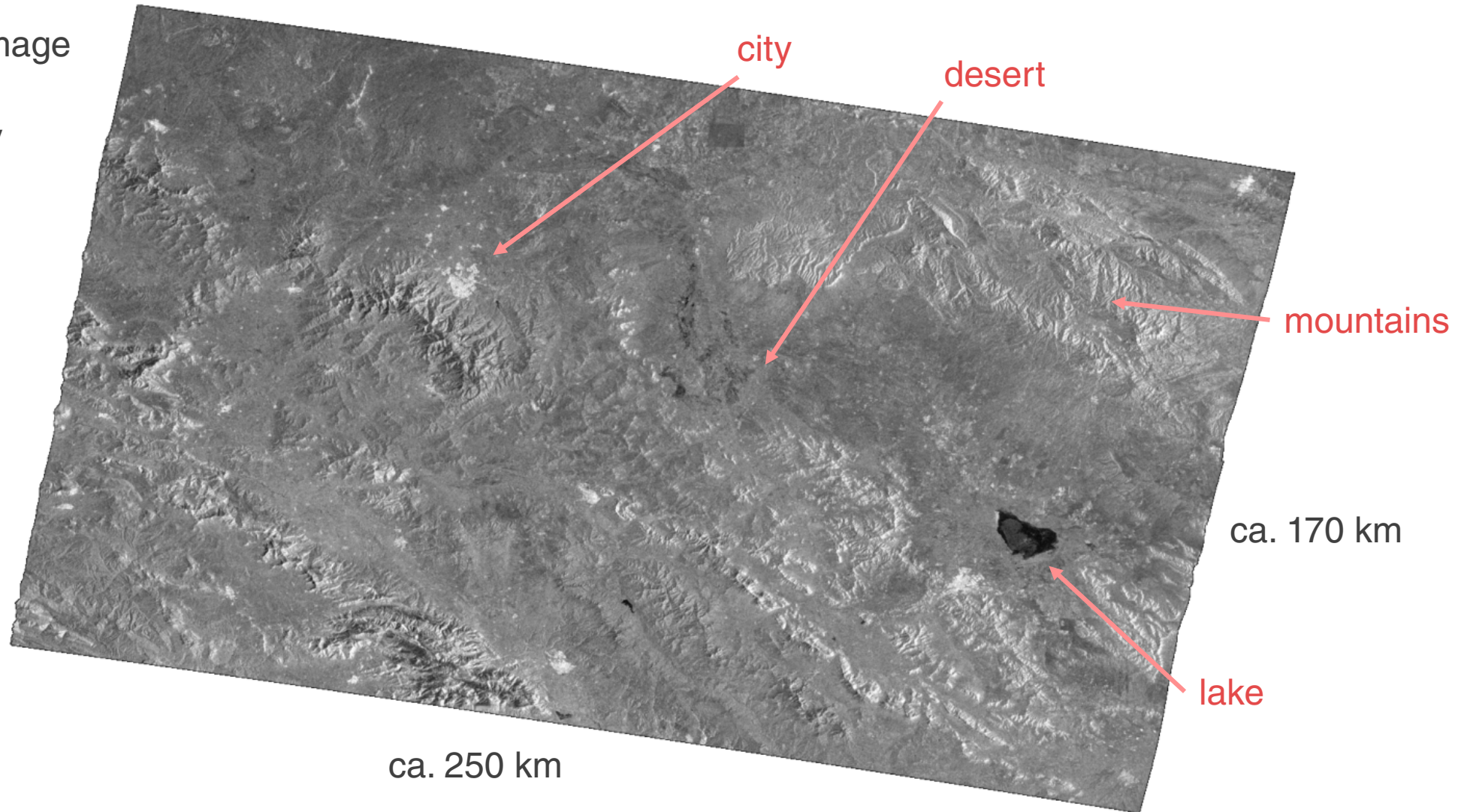
- precise terrain analysis
 - elevation models
 - tectonics
 - volcanic activities
- vegetation status & health
 - yield predictions
 - forest development
- land cover
 - water body mapping
 - flood alerts
 - geological structures
 - desertification
- soil moisture conditions
 - drought estimation
 - irrigation
- ship detection
- etc



example Sentinel-1 image over western Iran

1 typical S-1 GRDH image

- ca. 1 GB
- ca. 1000 images / day



Sentinel-1 datacube in Vienna

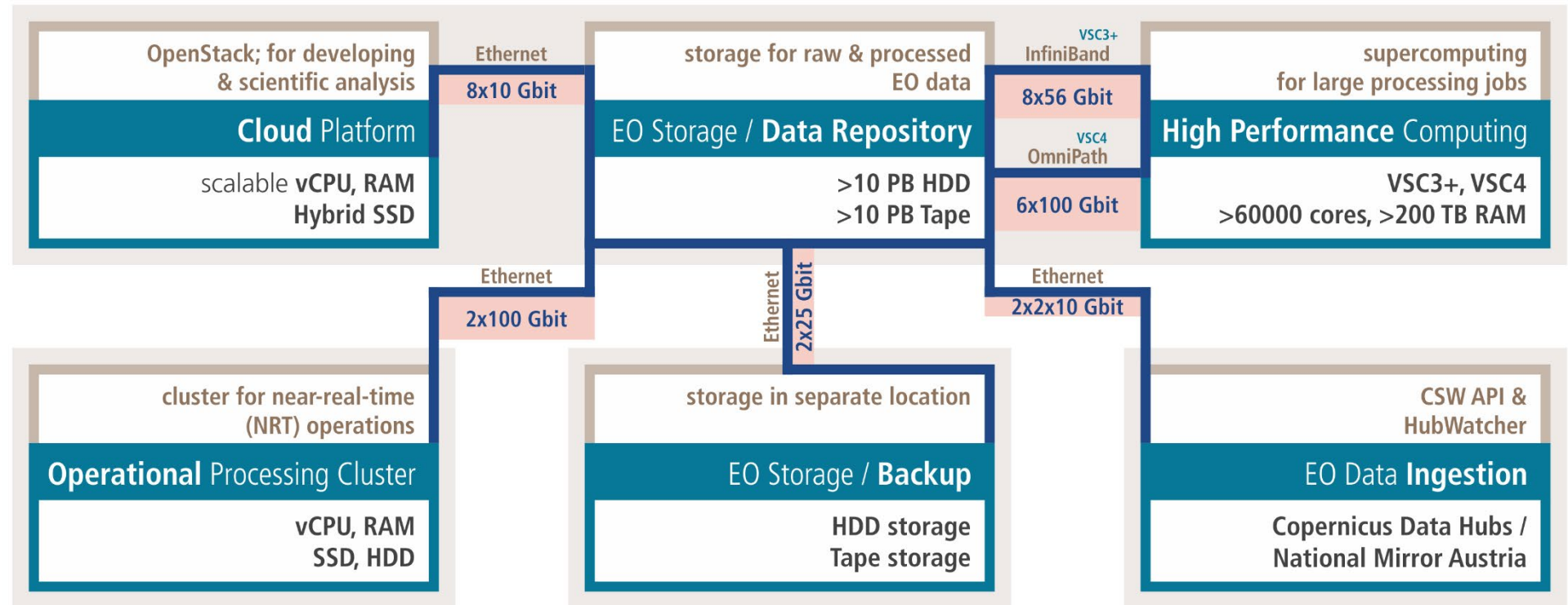
big data from satellites? high demands on hard/software!

<https://eodc.eu/>



Technical Infrastructure for Sentinel-1 ARD datacube

- petabyte storage
- gigabit ethernet
- super-computer
- cloud platform
- etc.



Wagner et al. (2021) A Sentinel-1 Backscatter Datacube for Global Land Monitoring Applications, *Remote Sensing*, 13, 4622.

datacube: processing & modelling

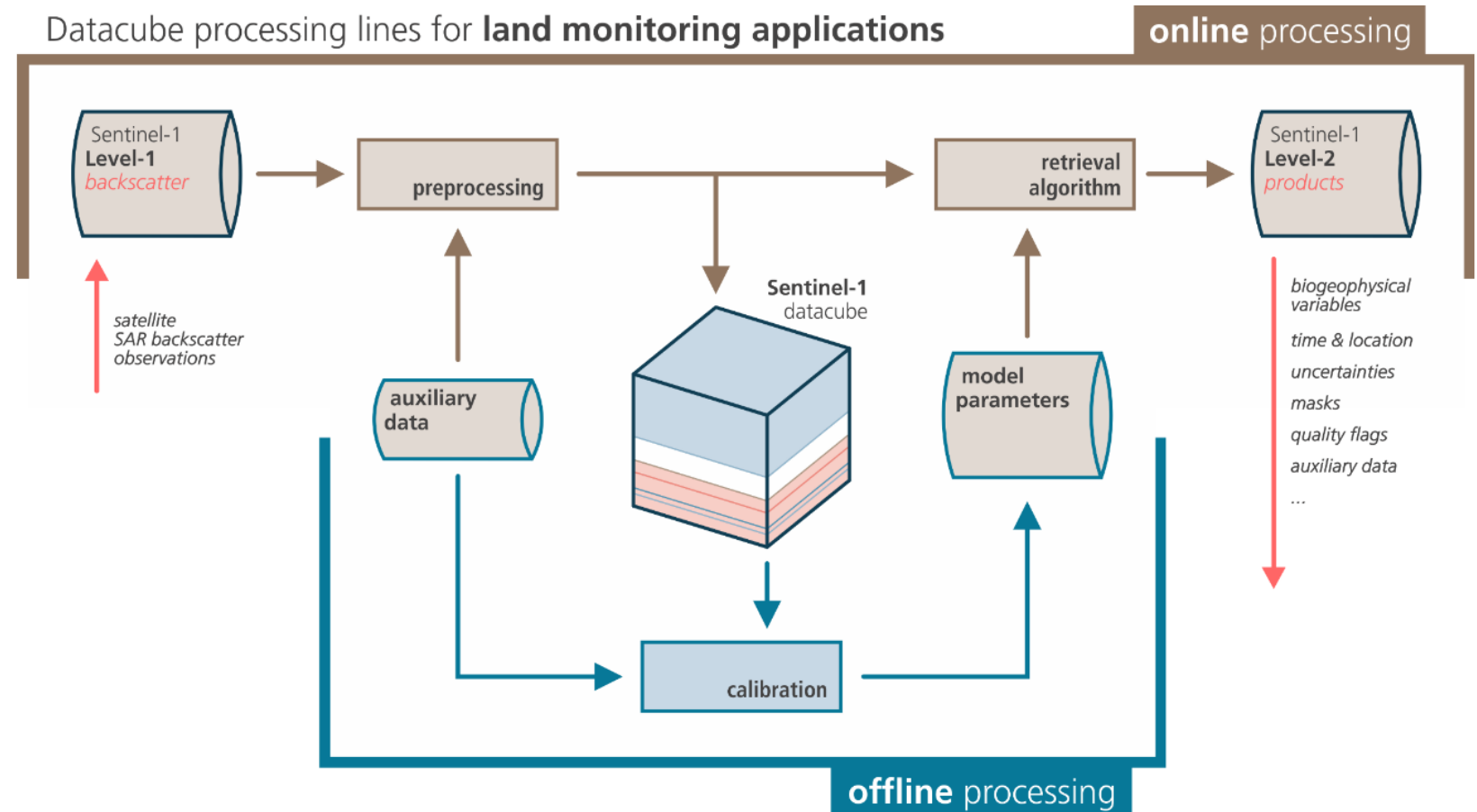


starting from scientific results

- algorithms for...
 - SAR processing
 - variable estimation
 - water bodies
 - soil moisture
 - vegetation health & density
 - etc
- operational services
 - global
 - in Near-Real-Time (NRT)
 - open & accessible

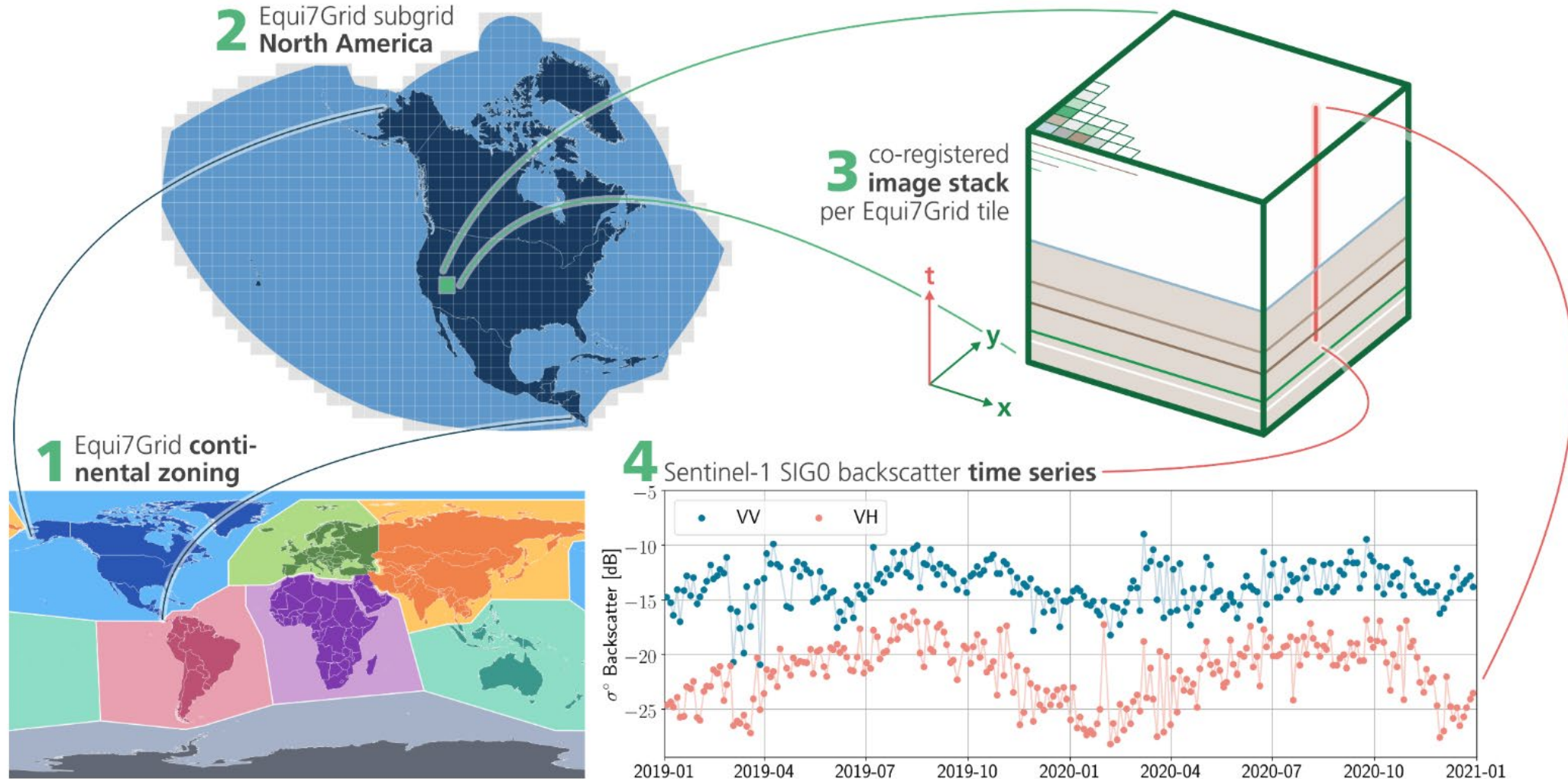


Datacube processing lines for **land monitoring applications**



datacube architecture: Equi7Grid & time series

Sentinel-1 ARD datacube: Concept of **Equi7Grid data structure & time series access** | Example for T3-tile over the USA



Bauer-Marschallinger et al. (2014) Optimisation of global grids for high-resolution remote sensing data, *Computers & Geosciences*, 72, 84-93.

Figure in "Wagner et al. (2021) A Sentinel-1 Backscatter Datacube for Global Land Monitoring Applications, *Remote Sensing*, 13, 4622."

<https://s1map.eodc.eu/>

- Sentinel-1 (MENSIG38_VV)
- Sentinel-1 (MENSIG38_VH)
- OpenStreetMap

Mosaic from 2016-17 scenes
500 000 Sentinel-1 images
→ 1.1 Petabyte

Bauer-Marschallinger et al. (2021) **The normalised Sentinel-1 Global Backscatter Model** – mapping Earth's land surface with C-band microwaves, *Scientific Data*, 8, 277.

the S1GBM deposit at
TU Research Data
repository

the S1GBM

mosaics for Earth's land surface

- radar VV & VH reflectivity
- statistical layers

the first global
(normalised/harmonised) SAR
mosaic

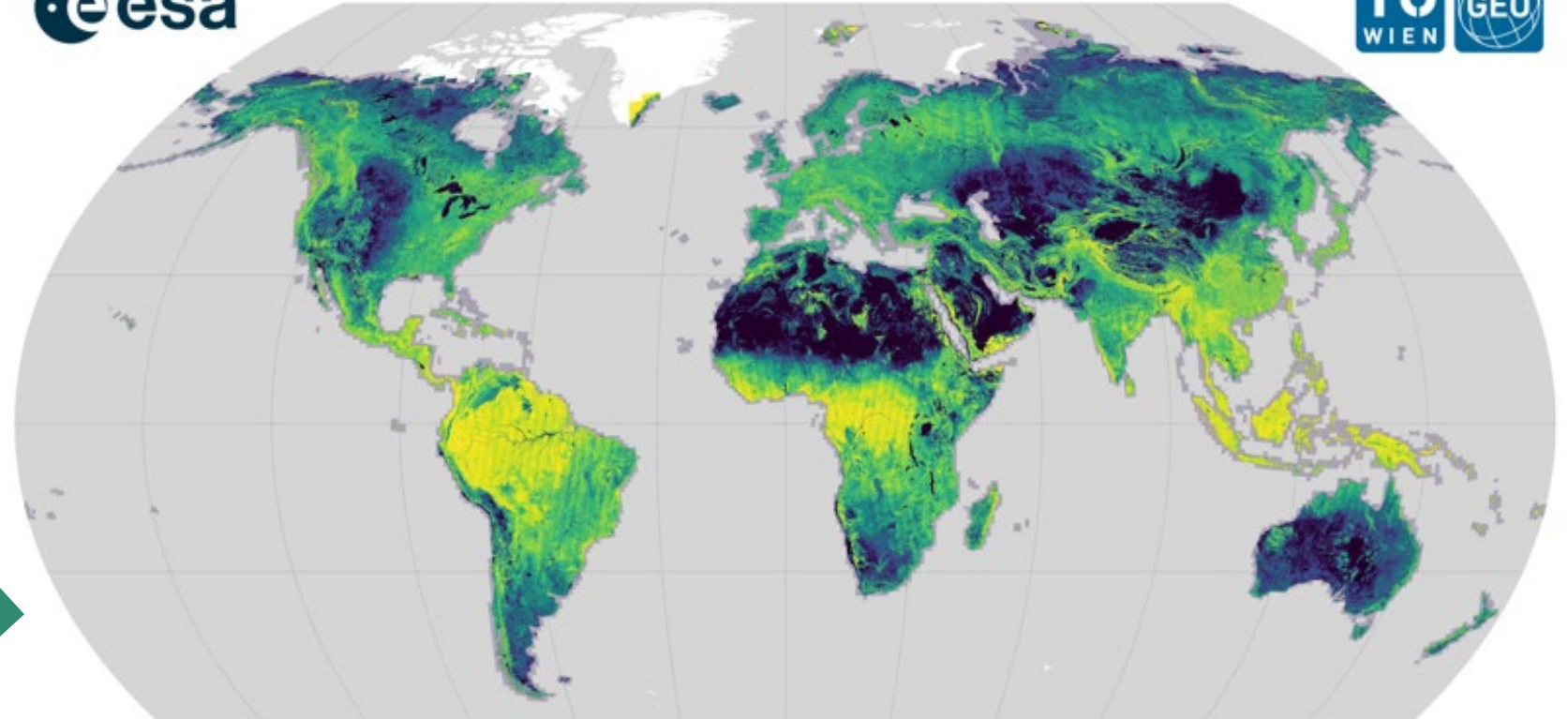
the first big-data-set we saw
meaningful to share with
community, used for

- land cover classification
- water body delineation
- human settlement mapping
- morphological/geological analysis
- soil typ determination
- more to come?

...added value



The Sentinel-1 Global Backscatter Model (S1GBM)



the S1GBM

mosaics for Earth's land surface

- radar VV & VH reflectivity
- statistical layers

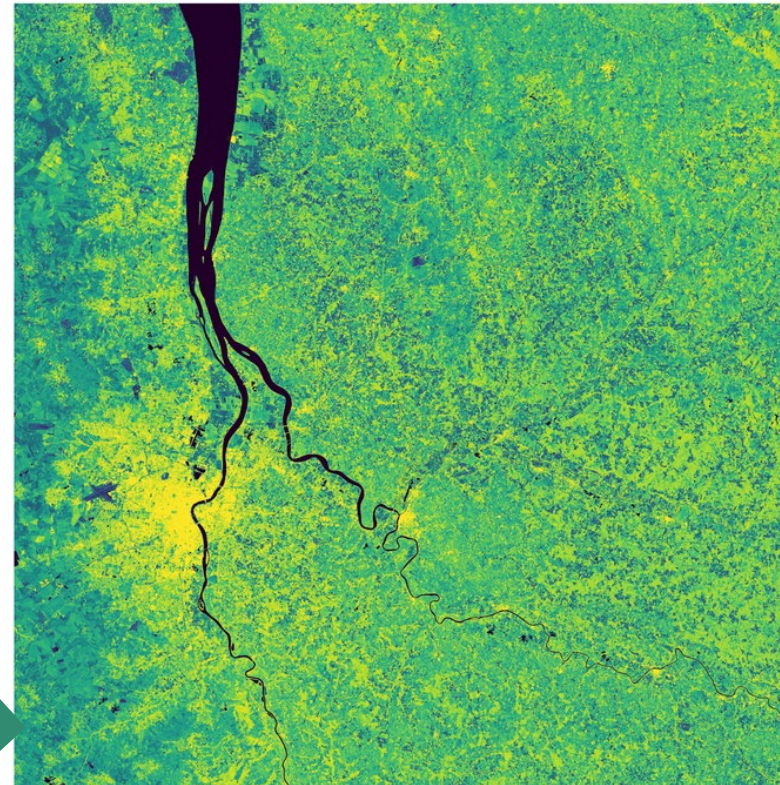
the first global
(normalised/harmonised) SAR
mosaic

the first big-data-set we saw
meaningful to share with
community, used for

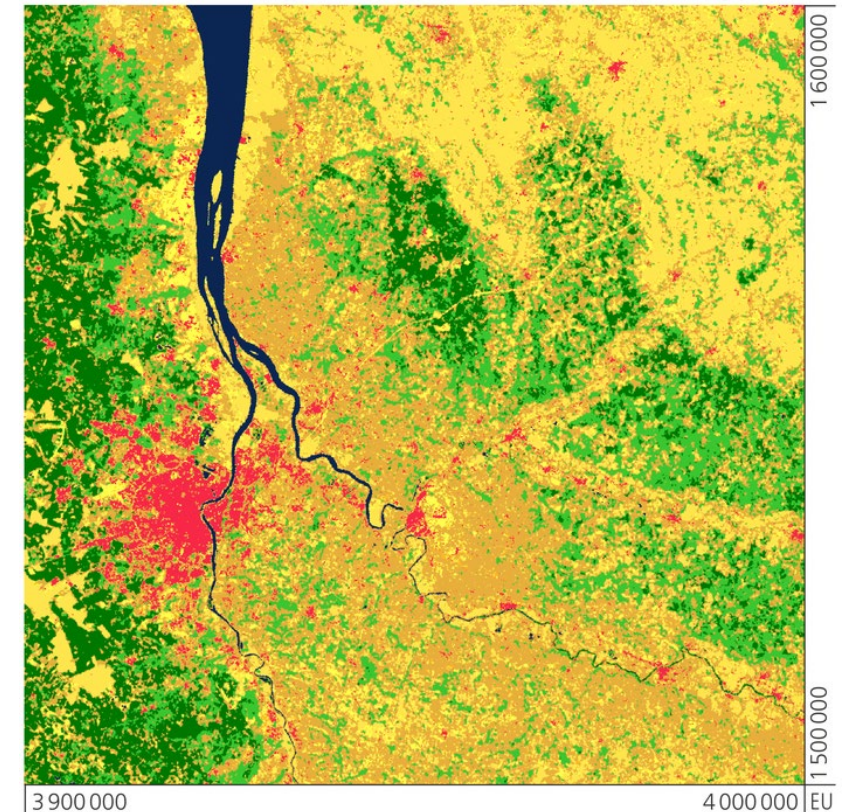
- land cover classification
- water body delineation
- human settlement mapping
- morphological/geological analysis
- soil typ determination
- more to come?

...added value

b) S1GBM VV backscatter mosaic



c) CGLS LandCover100



the history of this project

2018: European Space Agency (ESA) project on mosaics successfully finished

- ~1 PB input & intermediates
- 75 TB output layers

...

2020: shouldn't we publish this cool dataset?

- just a paper?
- selected mosaics?
 - ~ 3 TB

but who can host such a big dataset?

- ESA had no resources/capability to host it
- we don't like Google & co...
- contacts to SurfSARA
 - but we're not Dutch...
 - but this is really big...

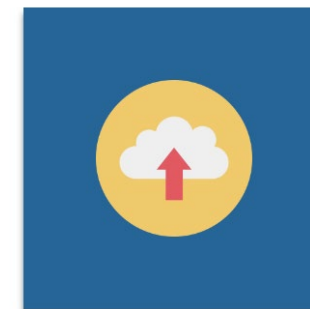
...

- contacts to TU Center for Research Data Management
- our requirements can be met
- pilot user

2021

- deposit successfully online
- first under **embargo**
- then, after peer-review publication, fully released

Welcome to TU Wien Research Data



Deposit

the S1GBM V1.0 dataset

1 DOI

- generated by TU Research Data

2 global mosaics

- VV + VH radar bands

6 continents

- tiled per 100km squares

12 zipped collections

- grid + tile structure as folder + file structure
- file nomenclature

32138 GeoTIFF files

- 10m pixel size
- geo-metadata

Collection file name	Continent	Tiles VH	Tiles VV	Volume
S1GBM_VH_mean_mosaic_v1_EQUI7_AF010M	Africa	3775		320 GB
S1GBM_VV_mean_mosaic_v1_EQUI7_AF010M			3776	336 GB
S1GBM_VH_mean_mosaic_v1_EQUI7_AS010M	Asia	4457		378 GB
S1GBM_VV_mean_mosaic_v1_EQUI7_AS010M			4457	379 GB
S1GBM_VH_mean_mosaic_v1_EQUI7_EU010M	Europe	1339		100 GB
S1GBM_VV_mean_mosaic_v1_EQUI7_EU010M			1339	98 GB
S1GBM_VH_mean_mosaic_v1_EQUI7_NA010M	North America	2669		223 GB
S1GBM_VV_mean_mosaic_v1_EQUI7_NA010M			2670	215 GB
S1GBM_VH_mean_mosaic_v1_EQUI7_OC010M	Oceania	1786		144 GB
S1GBM_VV_mean_mosaic_v1_EQUI7_OC010M			1788	139 GB
S1GBM_VH_mean_mosaic_v1_EQUI7_SA010M	South America	2041		169 GB
S1GBM_VV_mean_mosaic_v1_EQUI7_SA010M			2041	169 GB
	Total	16067	16071	2.67 TB

Table 1. The ESA S1GBM data publication is organised at the TU Data repository in twelve collections, with information on tile count per continent and polarisation, and data volume. Note that for each of the 32138 tiles, an additional quicklook-file is provided, yielding a total number of 64276 files.

the S1GBM landing page

the TU RD deposit's landing page comprises:

- the DOI and version
- dataset citation
- textual high-level description
- preview file (!)
- 10 download links for the collections
- core facts of the datarecord description & available codes
- link to the browser-based data viewer
- link to the peer-reviewed article
- link to funding organisations
- links to related works

→ more than we had thought we need initially 😊

The screenshot displays the landing page for the S1GBM dataset on the TU Wien Research Data platform. The page includes a navigation bar with 'Log out', 'HOME', 'COMMUNITIES', 'MY DASHBOARD', 'REQUESTS', and 'SETTINGS'. The main content area features the dataset title, authors (Bauer-Marschallinger, Bernhard; Cao, Senmao; Navacchi, Claudio; Freeman, Vahid; Reuß, Felix; Geudtner, Dirk; Rommen, Björn; Vega, Francisco Ceba; Snoeij, Paul; Attema, Evert; Reimer, Christoph; Wagner, Wolfgang), and a citation section. The citation is displayed in APA style: Bauer-Marschallinger, B., Cao, S., Navacchi, C., Freeman, V., Reuß, F., Geudtner, D., Rommen, B., Vega, F. C., Snoeij, P., Attema, E., Reimer, C., & Wagner, W. (2021). The Sentinel-1 Global Backscatter Model (S1GBM) - Mapping Earth's Land Surface with C-Band Microwaves (1.0) [Data set]. TU Wien. <https://doi.org/10.48436/n2d1v-gqb91>. The details sidebar on the right shows the DOI (10.48436/n2d1v-gqb91), version (1.0), and publisher (TU Wien).

the summary & acknowledgement slide

following requirements could be met:

- open access
- citation / DOI / license / version control
- hosting of ~3 TB
- keeping tile/file-structure
- handable download (collections)
- description & documentation
- linking to related works & funding
- long-term accessibility

through direct communication with TU Center for Research Data Management:

- initial FTP upload from EODC storage
- download user statistics
- help & support in all related questions (!)

some lessons learned

a **DOI-link** is persistent!

- a link to a repo, or a paper, perhaps not...

when planning a data-focused project:

- estimate dataset volumes
 - discuss early (!) access and license options
 - think about shareability and how to make it usable
- consider drafting a Data Management Plan (DMP)

generally, when planning a research project:

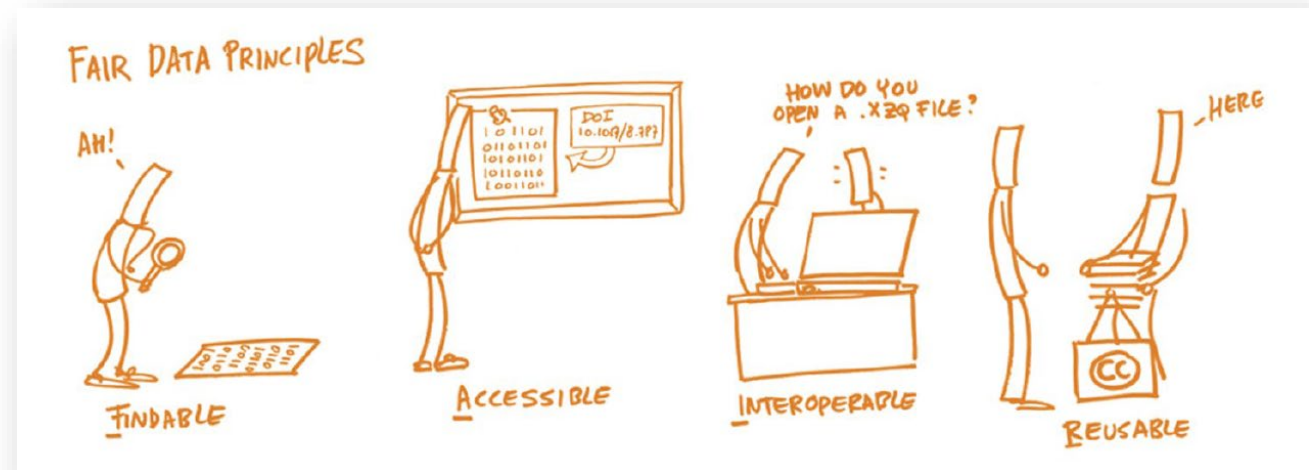
- should the obtained data follow the FAIR principles?

DMP makes you think

- what data you will use and where you get it from
- what infrastructure, software, licenses are needed
- what will be the output of your research
- how you will share your research outputs

Internal continuing education course at TU Wien

Tomasz Miksa, Christiane Stork, Center for Research Data Management, 15 November 2022



Thanks for your attention!
