EARTH BIG DATA

Where Solutions Begin.

Cloud-Based Solutions for ScanSAR Processing

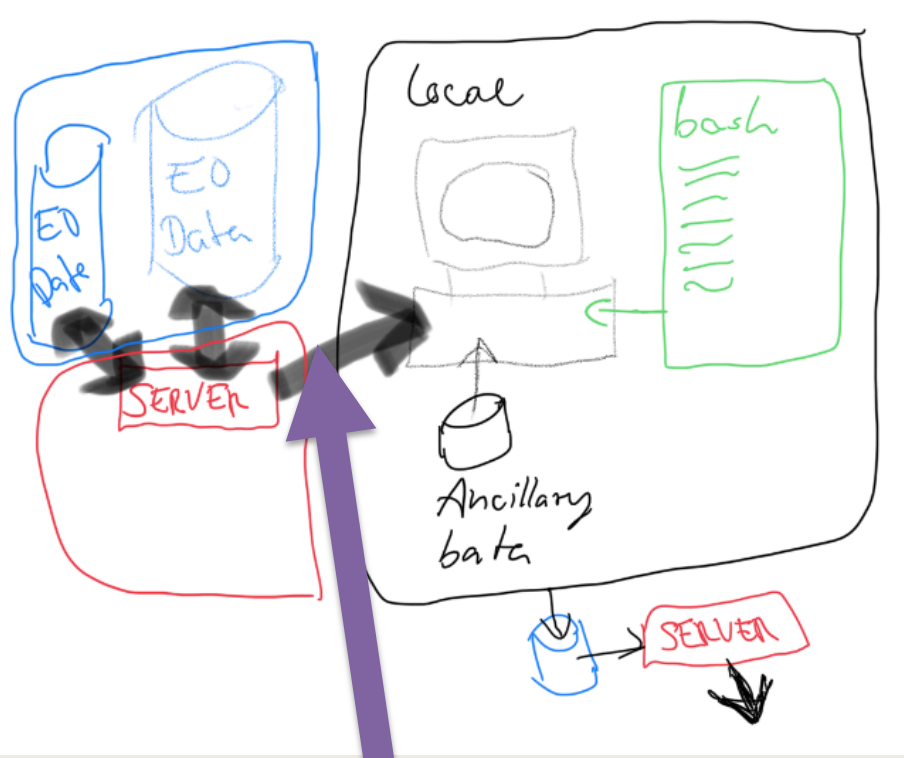
Josef Kellndorfer
President and Senior Scientist

What is EARTH BIG DATA (EBD)?

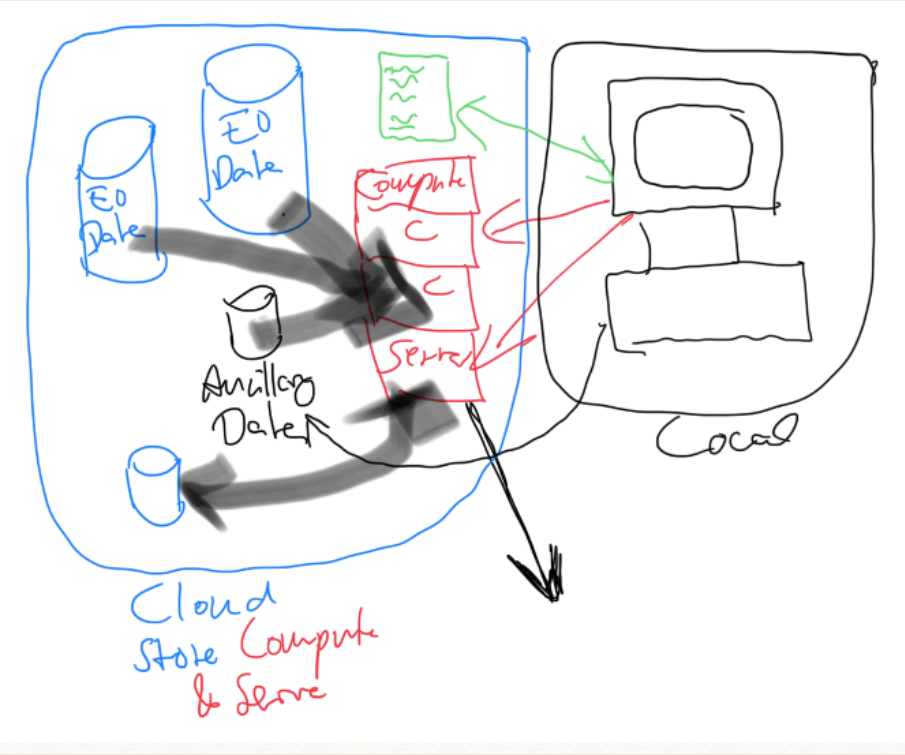
- Founded in 2015 by Dr. Josef Kellndorfer
- Vision: Develop and provide modern computational tools and approaches to mine **big Earth Observation data sets** efficiently with an **eye on solutions for the big environmental challenges** of our times.
- EBD is committed to harnessing Open Source software solutions and community engagement in geospatial software development
- EBD addresses the “big EO data challenge”

The Big EO Data Challenge

OLD Model:
Copy data to local and processes locally



New Model:
Copy instructions to cloud and "process next to the data"



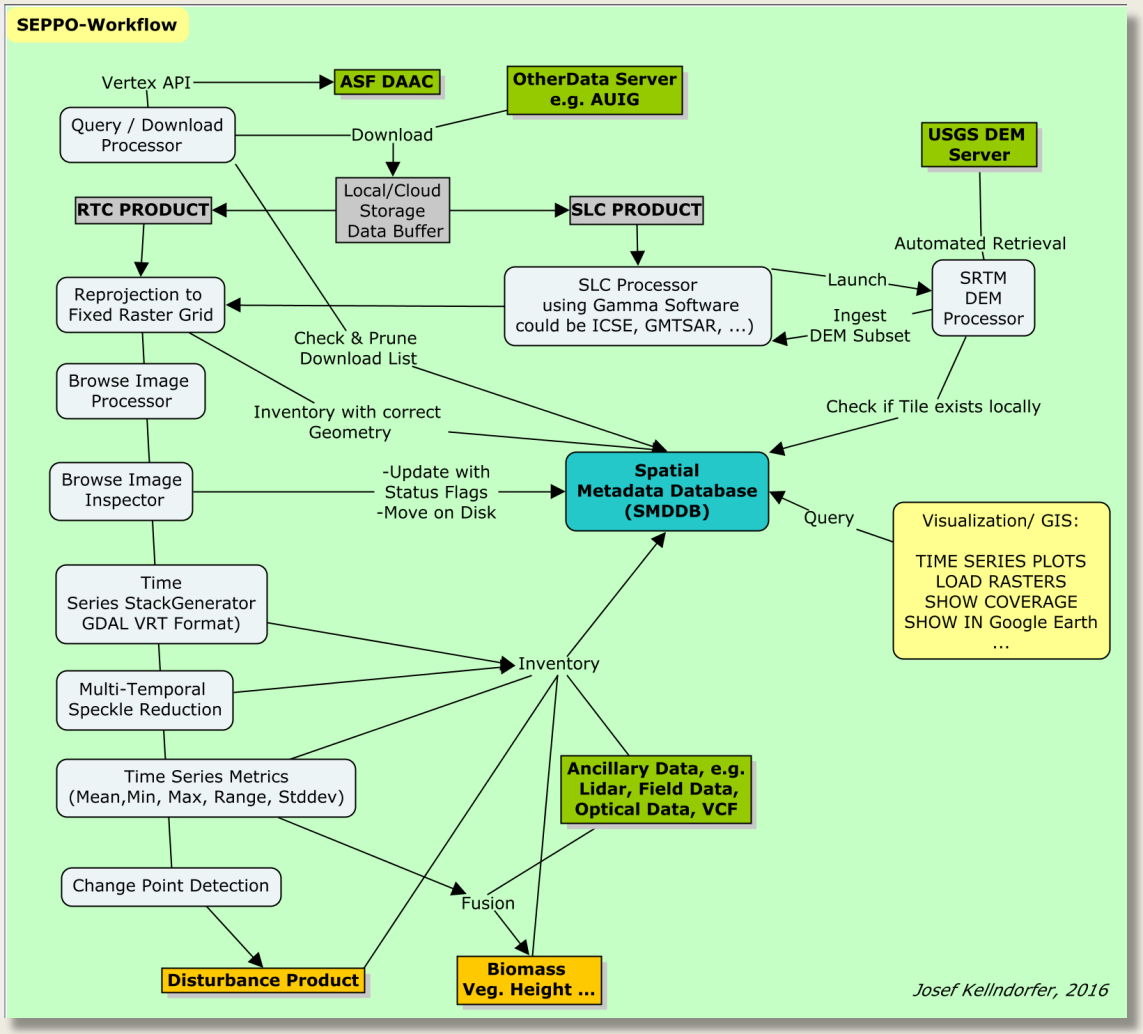
This is the big EO data challenge: To much data to handle
in the era of time series high-res data sets
from new SAR/optical/lidar … missions (e.g. NISAR 100TB/day)

EARTH BIG DATA
Where Solutions Begin.

Process Automation

SEPPO

Software for
Earth Big Data
Processing,
Prediction Modeling,
and
Organization



Automated work flow example for SAR based ecosystem disturbance mapping

EARTH BIG DATA
Where Solutions Begin.

What is **SEPPO**?

- Software package for automation of large volume SAR, optical, and lidar data processing, largely based in Open Source components
  - For SAR processing:
    - *Gamma Remote Sensing* software available (preferred solution)
    - ISCE, the official NISAR processing software
    - SNAP, the official Sentinel-1 processing software
- Automation of retrieval, pre-processing, and value-added processing
- Integrated with a searchable spatial metadata database (Postgres/PostGIS) for management of all processing steps
- Unix/python3 API (bash, python, R, …)
- Commandline, scriptable modules
- Fully cloud deployable (Amazon Web Services)
- Documentation processor with Spinx
- Some components as open source available as **openSAR** on github.com/earthbigdata

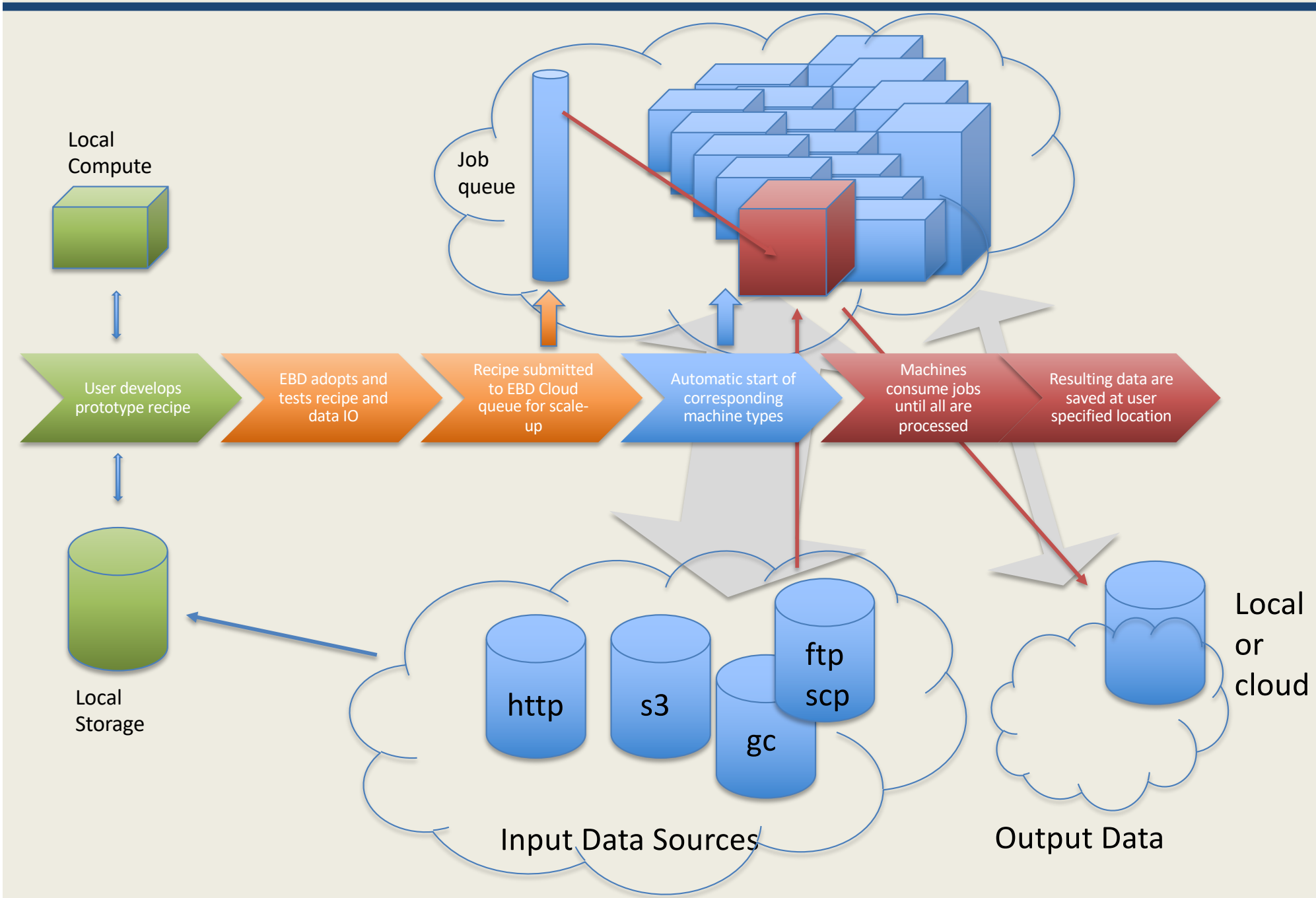
EARTH BIG DATA
Where Solutions Begin.

EBD CloudProcessor – Cost Factors

- Process Preparation, Initiation, Monitoring, and QC
  - Script development with client for cloud scaling
  - Rigorous script and process flow testing before scale-up
- Compute Instances
  - Cost minimization through smart deployment of appropriate machine types (cores, RAM, internet gateways)
  - Optimized use of lower cost instances if jobs are not time critical
- Data Egress
  - Work with client to optimize (minimize) data egress cost through data compression and balance of processing to data reduction
- Data Storage
  - Balance cost of data storage for optimized processing at various data storage models (simple storage solution s3, elastic file system EFS)
- Software licensing fees for non-open source software

EARTH BIG DATA
Where Solutions Begin.

EBD CloudProcessor Work Flow



EARTH BIG DATA
Where Solutions Begin.

Overview of Services

http://earthbigdata.com

Email:
info@earthbigdata.com

EARTH BIG DATA
Where Solutions Begin.

EBD SAR DATA SERVICES

- Supported SAR Sensors: Sentinel-1, JERS-1,ALOS-1, ALOS-2, NISAR, others …
- Radiometric Terrain Correction Processing (based on Gamma Software[(1)])
  - Customizable pixel spacing
  - Standard UTM or Geographic Coordinate Projections, customizable other projections
  - Custom DEM option. e.g. from Lidar; Standard is SRTM-30 DEM
  - Layover/shadow masks, incidence angle layers
- Time Series Generation
  - Multi-temporal speckle filtering
  - Time series metrics computation
  - Standard change-point detection
- Standard and Custom Tiling
  - Standard Lat/Lon 1x1 degree tiles
  - Standard Military Grid Reference System (MGRS) Sentinel-2 analogous (110x110 sqkm) [(2)]
  - Customized tiling and subsets (e.g. via Shapefiles, bounding box coordinates)
- Client Algorithm Application
  - e.g. change monitoring, biomass/carbon stock estimation, wetland mapping
  - Integration with optical data, e.g. Landsat, Sentinel-2
- Routine Processing for Monitoring Systems
  - Automated processing of newly acquired scenes
- Custom Data Delivery Options
  - Standard dedicated download links, shipment on media (DVD, Hard drives, thumb drives) upon request

[(1)] Partnership with http://ww.gamma-rs.ch GAMMA REMOTE SENSING

[(2)] See https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products

EARTH BIG DATA
Where Solutions Begin.

EBD OPTICAL DATA SERVICES

- Supported Optical Sensors: Sentinel-2, Landsat-8, others upon request
- Cloud-based processing from standardized products available via Amazon Web Services cloud stores
- Time Series Generation
  - Multi-temporal data stacking
  - Time series metrics (e.g. mean, min, max, range, variance, percentiles, vegetation indices)
- Standard and Custom Tiling
  - Standard Lat/Lon 1x1 degree tiles
  - Standard Military Grid Reference System (MGRS) Sentinel-2 analogous (110x110 sqkm) [1]
  - Customized tiling and subsets (e.g. via Shapefiles, bounding box coordinates)
- Client Algorithm Application
  - Integration with SAR data, e.g., Sentinel-1
- Routine Processing for Monitoring Systems
  - Automated processing of newly available scenes
- Custom Data Delivery Options
  - Standard dedicated download links, shipment on media (DVD, Hard drives, thumb drives) upon request

[1] See https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products

EARTH BIG DATA
Where Solutions Begin.

EBD Application Services

- EBD supports client applications with:
  - High-data volume throughput through our cloud scalable processing solutions -> Produce **Application Ready Data (ARD)**
  - Integration of **client algorithms into automated processing workflows**
  - Support of **client algorithm transfer to cloud based processing**
  - Transfer of complete processing **work flows to client-managed cloud infrastructure**
  - **Training on cloud processing** solutions
- Application examples:
  - Forest monitoring (Deforestation, Degradation, Biomass mapping)
  - Wetlands and inundation mapping and monitoring
  - Soil moisture monitoring
  - Land Cover change detection
  - Natural Hazards and disaster mapping and monitoring

EARTH BIG DATA
Where Solutions Begin.

ALOS-2 ScanSAR Processing

- Volume: ~120,000 scenes for 2019
- Gamma RS - based RTC Script
- Output: γ^0 backscatter for each scene,
- DEM: SRTM-1 arcsec scaled to desired output resolution
- Tiling into 1x1 degree blocks
  - Mosaic per cycle with far range over near range, potential trimming of far range edge
  - Acquisition date as raster layer corresponding to mosaicked pixels
  - Incidence angle map corresponding to mosaicked pixels
  - Output as Cloud-Optimized Geotiffs (COG)
  - Retain individual layers as compressed COG as option
  - Pixel as Area boundaries for seamless matching of tiles
- Stored in s3 bucket made available via https download URLs (to be discussed if retained on s3)

EARTH BIG DATA
Where Solutions Begin.

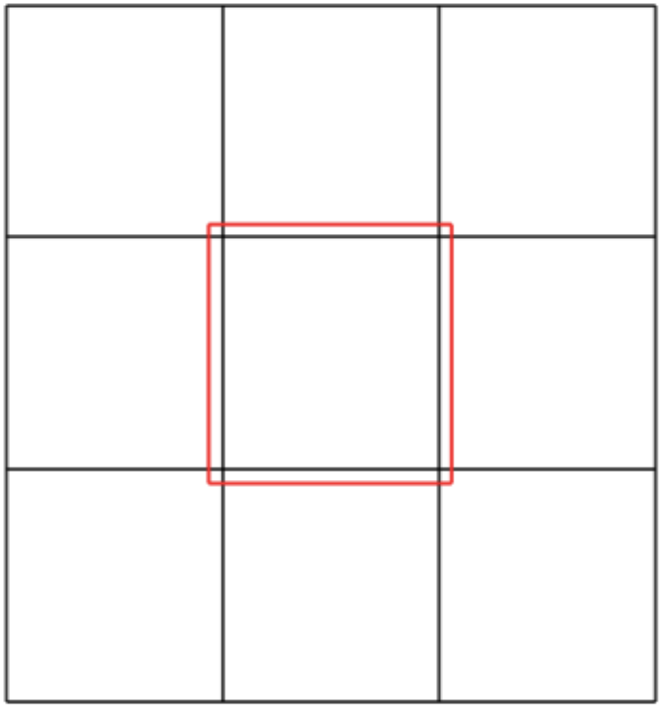
Tile Overlap or Not?

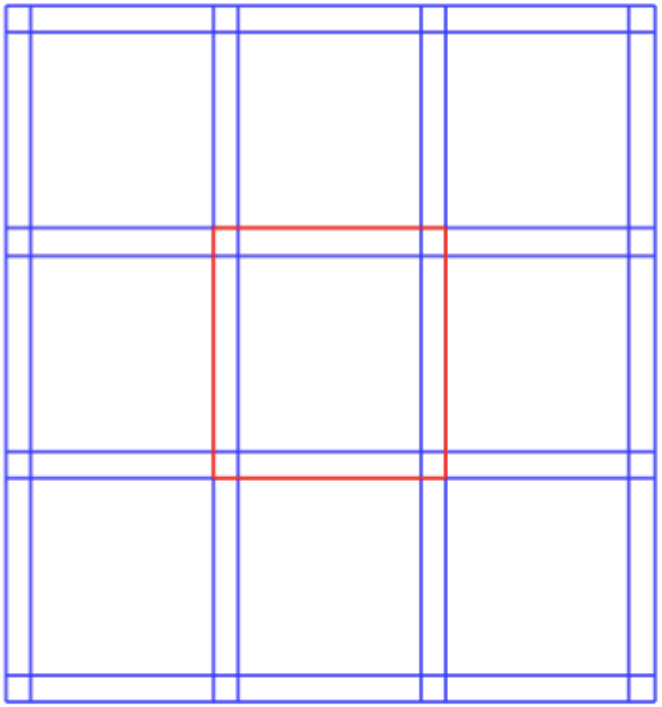
No overlap

- Advantage:
  - Seamless fit
  - No redundant data
- Disadvantage:
  - Need to mosaic neighboring tiles with any spatial kernel based estimates to avoid no-data edges

Overlap

- Advantage:
  - Retains seamless fit with kernels < overlap pixels/2
- Disadvantage:
- - Some data redundancy
- - overlap regions need to be accounted for





Black = No overlap, Blue and Red: Tiles with Overlap buffer

EARTH BIG DATA
Where Solutions Begin.

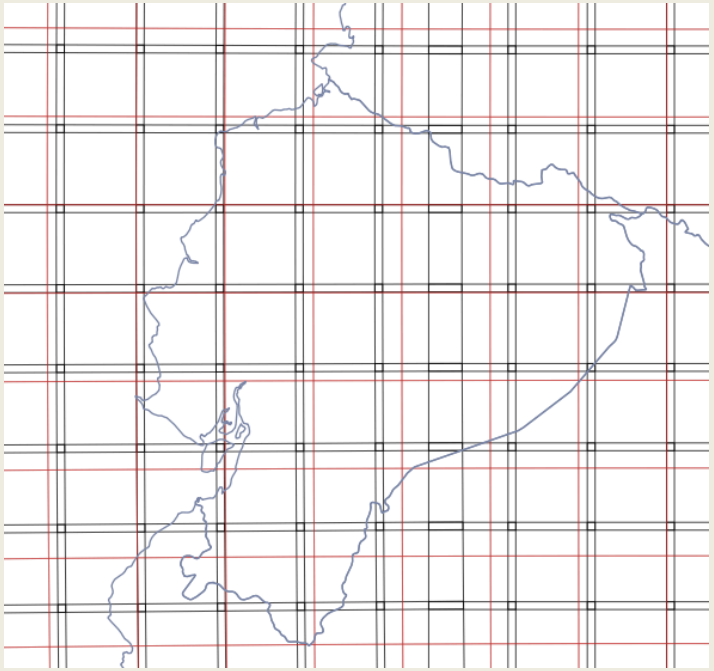
Tiling Schemes

- Lat/Lon 1x1 degrees (like SRTM)
  - Advantage:
    - Simple to reproject, mosaic
  - Disadvantage:
    - Variable longitudinal ground projection with latitude, thus oversampling required, leading to big data increase. Not suitable for high latitudes
    - Introduces variable statistics for spatial operations like speckle filtering
- UTM based Tiling
  - Advantage:
    - Used by Sentinel-2, Landsat efforts underway as Military Grid Reference System (MGRS)
    - Consistent pixel spacing across all regions, minimal ground projection distortion
    - Best effort to construct Application Ready Data Stacks
  - Disadvantage:
    - Cannot readily be mosaicked between UTM zones for large regions. However, most GIS packages have no problem with on the fly reprojection for neighboring UTM zones
    - Overlaps required for complete coverage, leads to some data redundancy

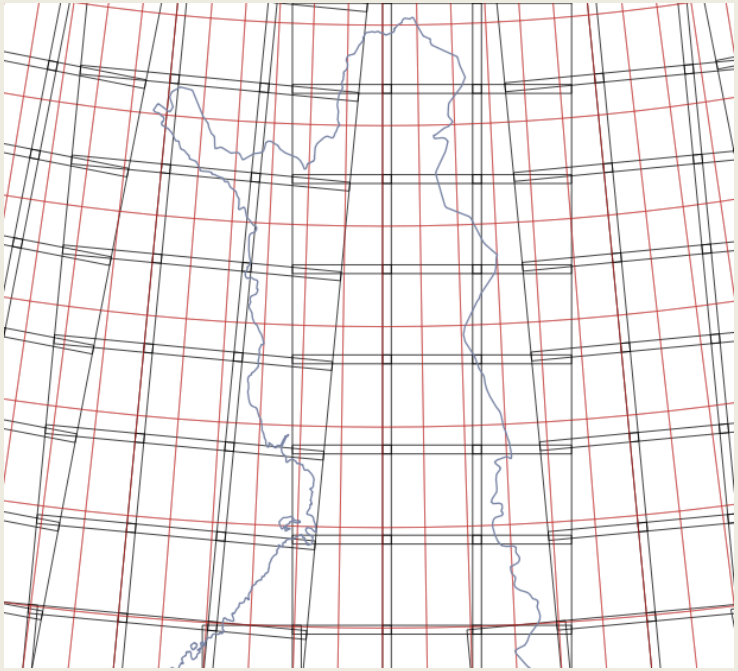
EARTH BIG DATA
Where Solutions Begin.

MGRS AND Lat/Lon Grid Comparison

ECUADOR



FINLAND



Lat/Lon 1x1 degree tiles
MGRS 110x110 sqkm tiles

EARTH BIG DATA
Where Solutions Begin.

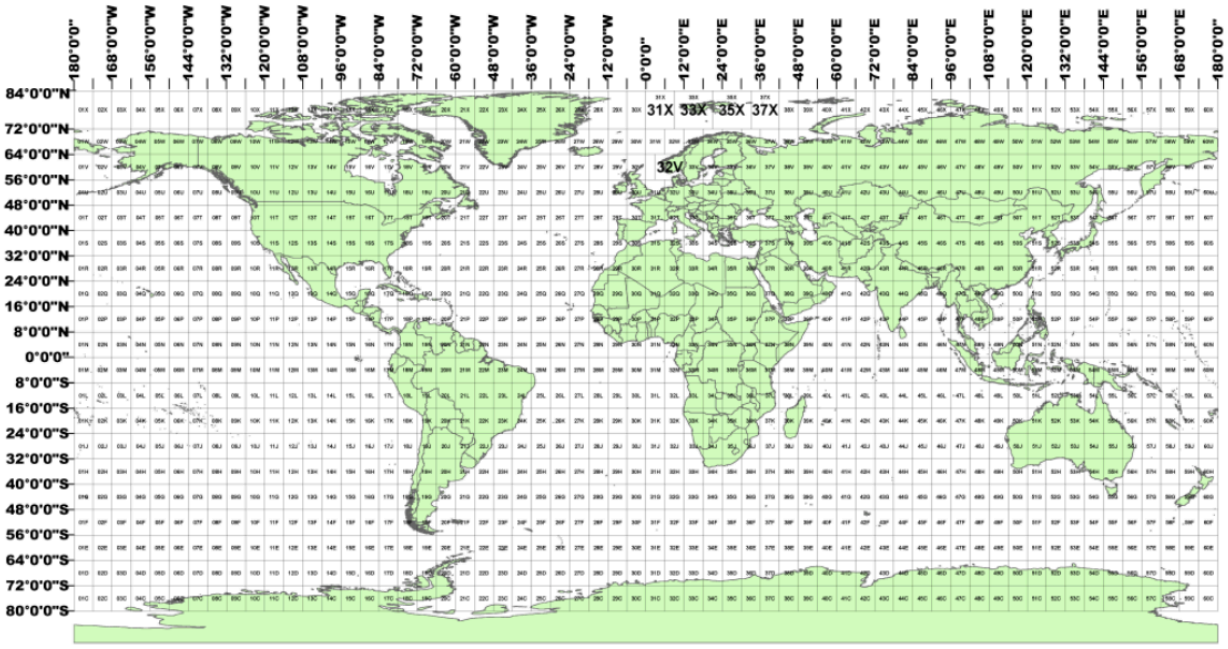
Landsat / Sentinel-2 Harmonization

Tiling System

Selected HLS tiling system is identical as the one used for Sentinel-2. The tiles dimension is 109.8km and there is an overlap of 4,900m on each side.

The system is aligned with the Military Grid Reference System (MGRS) and its naming convention derived from the UTM (Universal Transverse Mercator) system. The UTM system divides the Earth's surface into 60 vertical zones. Each UTM zone has a vertical width of 6° of longitude and horizontal width of 8° of latitude, as shown in the map below. Each UTM zone is subdivided in MGRS 100x100km zone. The first 2 digits and 1 letter correspond to the UTM zone, the two last letters to unique ID.

The kml edited by ESA with all tiles ID can be downloaded here.

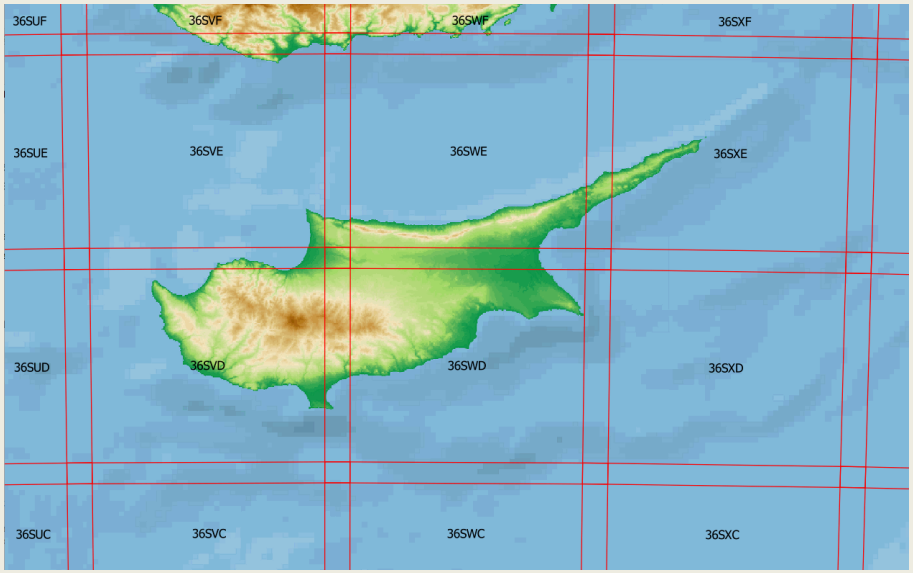


Military Grid Reference System

https://hls.gsfc.nasa.gov/products-description/tiling-system/

EARTH BIG DATA
Where Solutions Begin.

MGRS / Sentinel-2 Tiling



- 110x110 sqkm UTM projected tiles
- 20 m pixel spacing
- 5 km Overlaps between tiles
- Follows Military Grid Refernce System Standards coordinate standards
- Globally consistent

EARTH BIG DATA
Where Solutions Begin.

Decisions with Big Data Impact

- Pixel spacing: 10m vs 20m -> Factor 4
- Pixel encoding: Float32 vs. 16 Bit Amplitude -> Factor 2
- Subsetting of Data sets
  - Need to find viable geographic sampling units to process time series stacks
- Example. MGRS Tiling, RTC Products:

One year observation of Dual-Pol Data					
Tilesize	110	km			
Time Steps	30				
PixelSpacing	10	20	10	20	m
Pixel Coding	32	32	16	16	bit
One Tile Data Volume	27.0	6.8	13.5	3.4	GB
Global (41,000 Tiles)	1082.9	270.7	541.4	135.4	TB

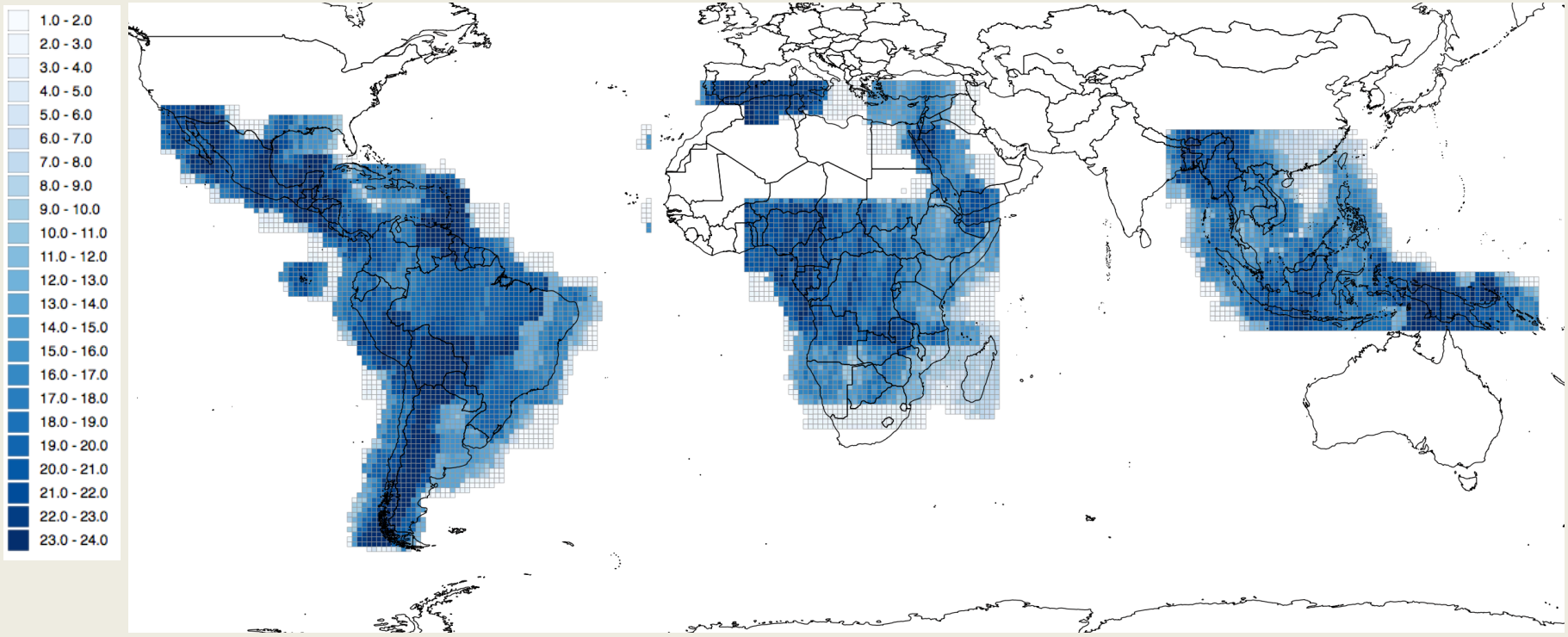
EARTH BIG DATA
Where Solutions Begin.

Decisions:

- Geographic coverage of tiles?
- Trimming of near and far-range edges? By how much?
- Multi-looking of ScanSAR data in pre-processing
  - Reduction in data ingress volume (60GB SLC, 10x1 looks: 6 GB)
- Provide tile mosaic as virtual mosaic on individual swath data using GDAL compatible virtual raster tables?
  - Can readily be ingested in any GIS, seamless to user
  - Advantage: No data loss from swath data
- UTM Projection?
  - Using UTM ensures consistence area coverage per pixel independent of latitude
  - Using MGRS tiling scheme like Sentinel-2 has immediate overlap with these data
- Lon/Lat?
  - Ensure upper left coordinate and pixel spacing for dataset seamlessly nest within full degree tiles (in particular when overlap is chosen
- Permanent access links instead of staged download (budget issue, EBD can provide cost-effective solutions)

EARTH BIG DATA
Where Solutions Begin.

KC_ScanSAR Tiles Coverage



EARTH BIG DATA
Where Solutions Begin.

Contact

EARTH BIG DATA

josef@earthbigdata.com

EARTH BIG DATA, LLC
P.O. Box 114
Woods Hole, MA 02543
p +1 508-444-0885
f +1-508-455-5118
w earthbigdata.com