

GeoDN: A Large-Scale Data Mining Perspective

Conrad Albrecht
Earth Observation Data Science, German Aerospace Center

January 18, 2023 – *IBM Climate Network Summit*



1 | *Geospatial Discovery Network* – data science opportunities

© 2023 Large-Scale Data Mining in Earth Observation, German Aerospace Center (DLR)

Interconnect National Geospatial Data Centers | I

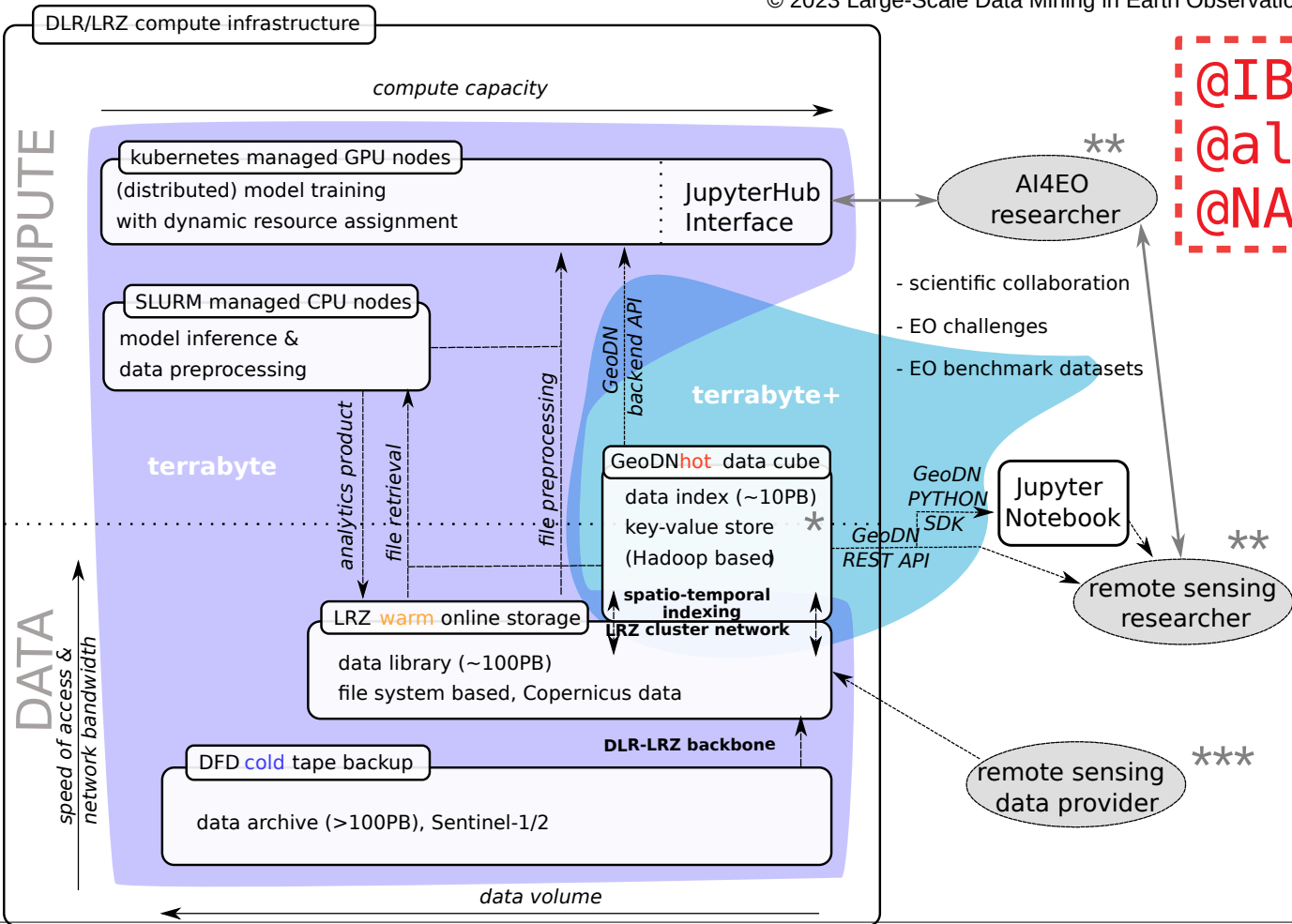
Perspectives for Large-Scale Geospatial Data Science | II

GeoDN's as International Climate Action Strategy Platform | III



2 | DLR terrabyte – geospatial AI analytics cluster

© 2023 Large-Scale Data Mining in Earth Observation, German Aerospace Center (DLR)



@IBM Research*
 @all GeoAI analytics**
 @NASA/NOAA/WMO/ECCC***

- scientific collaboration
- EO challenges
- EO benchmark datasets



2 | DLR terrabyte – geospatial AI analytics cluster

© 2023 Large-Scale Data Mining in Earth Observation, German Aerospace Center (DLR)



@NASA*
 @Argonne**
 @RedHat***
 @IBM****
 @Columbia*****

monitoring & user interaction
Web-IDEs





geospatial data cube services













geospatial & machine learning processing backends






compute resource & work flow management, monitoring

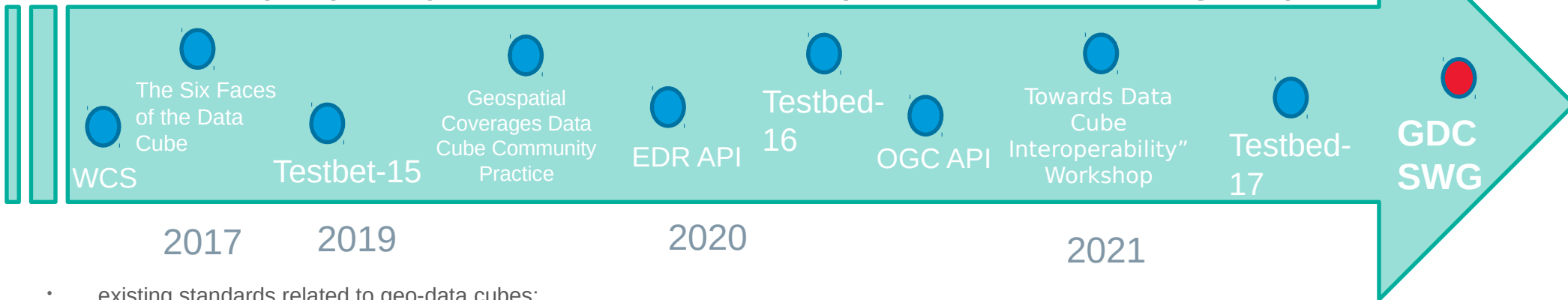

* Albrecht et al. "Open Source Science for Large-Scale Data Mining in Earth Observation", NASA Open Science Workshop (2022), <https://elib.dlr.de/185458>

3 | terrabyte@GeoDN – towards interoperable OGC datacube APIs

© 2023 Large-Scale Data Mining in Earth Observation, German Aerospace Center (DLR)

@OGC
@NASA
@ECCC

GeoDataCube (GDC) history @ OGC Earth Observation Exploitation Platform Working Group*

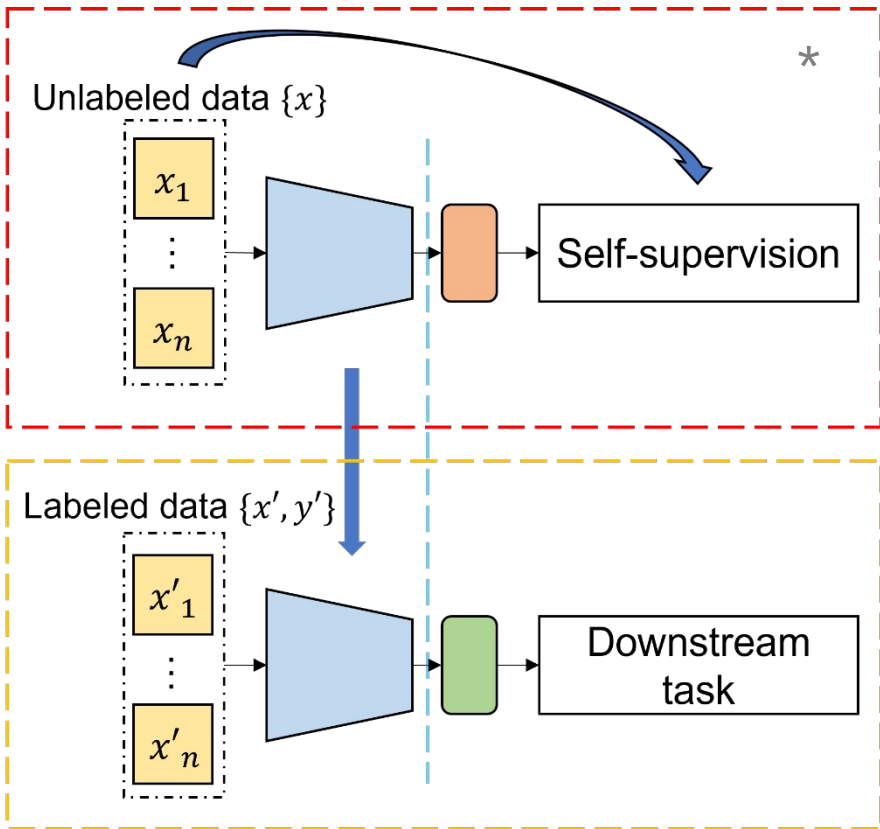


- existing standards related to geo-data cubes:
 - OpenGeospatialConsortium WebProcessingService: <https://www.ogc.org/standards/wps>
 - OpenGeospatialConsortium DataAccess(and)ProcessingAPI: <https://www.ogc.org/blog/4665>
 - SpatialTemporalAssetCatalogue API: <https://github.com/radianteearth/stac-api-spec>
 - SentinelHub processing API: <https://docs.sentinel-hub.com/api/latest/reference/#tag/process>
- Cross-platform, cross-datalayer spatio-temporal data cube join RESTful API
 - follow-up on OGC initiative „Towards Data Cube Interoperability“: <https://www.ogc.org/projects/initiatives/gdc>
 - objective:** pursue a user-centric, pragmatic approach to define an (extendable) API for data cube filtering and joining

4 | terrabyte@GeoDN – Overcome Data Gravity by AI Data Compression

© 2023 Large-Scale Data Mining in Earth Observation, German Aerospace Center (DLR)

@oxford**
 @IBM Research/ECCC**
 @all GeoData centers**



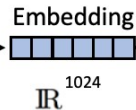
Datacenter 1:



Encoder S1

Encoder S2

- Data:**
- S1 – Europe
 - S2 – World-wide

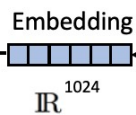


Datacenter 2:



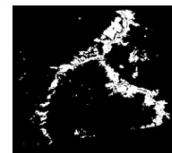
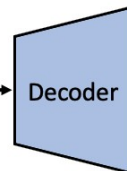
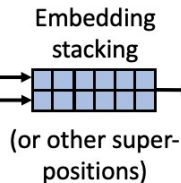
Encoder S1

- Data:**
- S1 – World-wide



**

User: Flood mapping US



5 | terrabyte@GeoDN – GeoAI services 4 global climate action

© 2023 Large-Scale Data Mining in Earth Observation, German Aerospace Center (DLR)

Williamsburg,
New York

3D LiDAR
point cloud

@NY Acad. Sci.
@IBM Research

Williamsburg
bridge

high-rise
building

trees

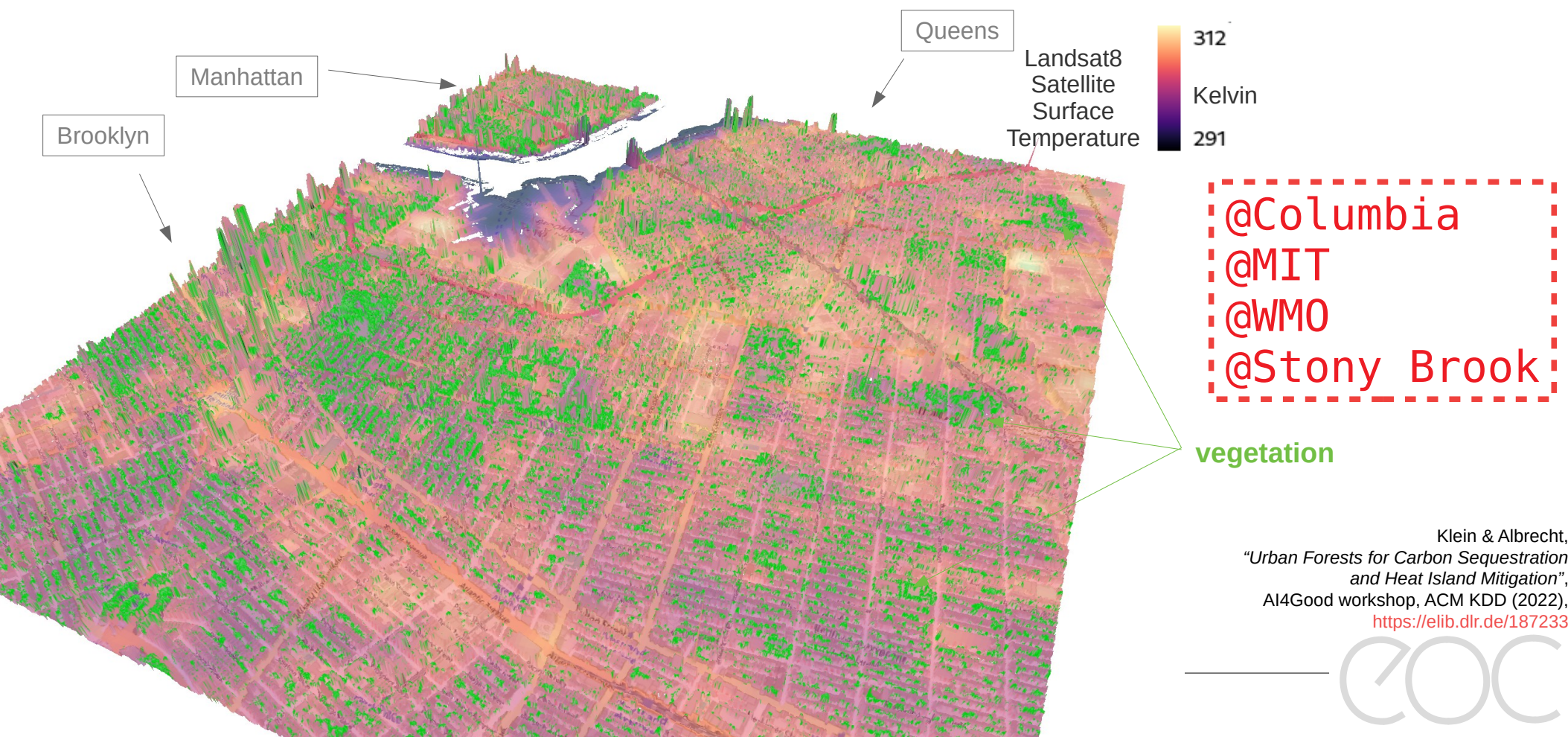
map legend:

- buildings
- vegetation
- road infrastructure
- water body
- bare ground



5 | terrabyte@GeoDN – GeoAI services 4 global climate action

© 2023 Large-Scale Data Mining in Earth Observation, German Aerospace Center (DLR)



Q&A session

