terrabyte Café
Welcome!

terrabyte team
18.04.2024

terrabyte

EOC
Earth Observation Center

DLR

terrabyte – status quo
Hardware

Data storage (LRZ)
- Tranche 1: 36 PB net
- Tranche 2: 50 PB net (Q1/2024)
- Currently 6 PB dedicated for users
- GPFS Filesystem

Compute environment (LRZ)
- Tranche 1: ca. 10.000 CPU vCores with 61 TB RAM and 60 GPUs with 15 TB RAM + 4,5 TB GPU RAM
- Tranche 2 (Q1/2024): ca. 44.000 CPU vCores with 273 TB RAM and 188 GPUs with 47 TB RAM + 14 TB GPU RAM

Network line between OP and LRZ
- 10 Gbit/s line (upgrade to 100 Gbit/s possible)

Internet line between LRZ and external hubs
- 20 – 40 Gbit/s to Copernicus Data Space Ecosystem
- 10 Gbit/s to NASA Alaska Satellite Facility

terrabyte – status quo
Data volumes

▪ Sentinel-1 SLC:	6.0 Petabyte	
▪ Sentinel-1 GRD:	3.7 Petabyte	→ more Petabyte to come
▪ Sentinel-1 NRB:		→ 7 - 10 Petabyte more to come
▪ Sentinel-2 Col. 0 L1C:	1.5 Petabyte	
▪ Sentinel-2 Col. 1 L2A:	8.5 Petabyte	→ 8,5 Petabyte more to come
▪ Sentinel-3:	0.6 Petabyte	→ more Petabyte to come
▪ Sentinel-5p:	1.2 Petabyte	→ more Petabyte to come
▪ Landsat Col. 2 Level-2:	2.7 Petabyte	→ 3 Petabytes more to come
▪ MODIS:	1.0 Petabyte	→ more Petabyte to come
▪ VIIRS/NPP:	130 Terabyte	→ more Petabyte to come
▪ Meteosat:	68 Terabyte	
▪ DEM:	8 Terabyte	
Total	**~25.4 Petabyte**	**min. 44 Petabyte**

terrabyte – status quo
Data

Continuous download for Sentinel, Landsat, MODIS, VIIRS will start in May 2024

▪ Sentinel-1 SLC:	1,2 Mio / 2,5 Mio	On demand
▪ Sentinel-1 GRD:	2,5 Mio / 2,5 Mio	All data between 2014 and **March 2024** available
▪ Sentinel-1 NRB:		Processor finalization (test data available for Germ
▪ Sentinel-2 Col. 1 L1C:	0,1 Mio / 25 Mio	Waiting for network line for D-SDA archiving
▪ Sentinel-2 Col. 1 L2A:	14,2 Mio / 25 Mio	2023, 2022, 2021, 2020, 2019
▪ Sentinel-3:		Initial download for OLCI, STAC metadata in preparation
▪ Sentinel-5p:	0,7 Mio / 1,0 Mio	Finalizing data integration (STAC metadata)
▪ Landsat Col. 2 Level-2:	3,9 Mio / 8,4 Mio	Europe, rest of the world downloading from USGS

Step-by-step with ESA processing (descending in time)

- MODIS: MOD09GQ, MOD10A1, MOD13A2, MOD13A3, MOD13Q1
  MYD09GQ, MYD10A1, MYD13A2, MYD13A3, MYD13Q1
- VIIRS/NPP: VNP09GA, VNP13A1, VNP15A2H
- Meteosat: SEVIRI L1.5 DISC / HRV / VISIR
- DEM: Copernicus DEM 30m, 90m
  SRTM DEM, Aster DEM

Ready
In progress
Blocked

Roadmap

Version 1.0

terrabyte launch

- STAC API
- Ingestion
- HPC SLURM
- Compute Portal
- Tutorials
- Documentation
- Support forum
- User data transfers
- Initial data integration

from 08/2023 to 05/2024

Expansion of web services
Development of operations

- Documentation website
- User STAC API
- Compute & storage extension
- Public data roadmap
- STAC improvements
- Onboarding workshops
- Data integration
- Operations concept
- Service deployment concept

Version 1.2

from 06/2024

Additional functionalities & services

- Standardization of data analysis & processing with HPC Slurm
  - OpenEO API
  - OGC EO Application Package
  - On-Demand processors
- D-SDA data transfers
- SLURM REST interface
- Application as a Service (e.g., Database instance)
- S3 Interface for data science storage
- STAC API Subscription service
- Web-based visualization service
- Dashboards
- Service deployment with Kubernetes

Release 1.2 – coming soon!

Breaking changes!

- Sentinel-1 data move from `/dss/dsstbyfs01` to `/dss/dsstbyfs03`
- New STAC item asset names for (STAC API “Test” available soon)
  - Sentinel-2
  - Landsat
- Support-Forum available under https://forum.terrabyte.lrz.de
- Service Desk available under servicedesk@terrabyte.lrz.de

Release 1.2 – coming soon!

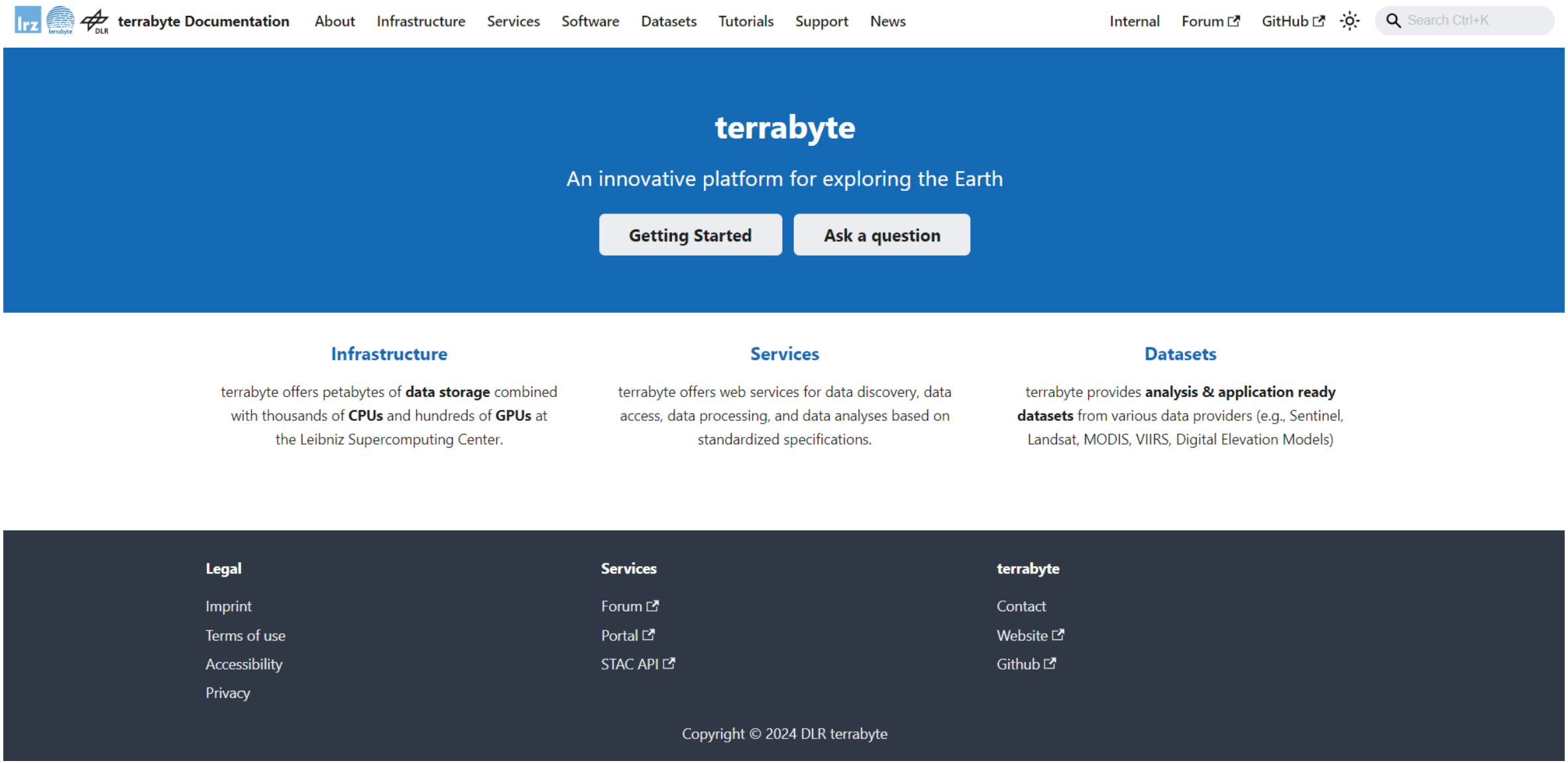
New tools and services

- Public documentation website
- Regular export of STAC metadata to GeoParquet files
- Private STAC API for users and projects
- terrapi command line and Python library
- Continuous data ingestion (hourly/daily)

Public documentation website (instead of DLR-internal Wiki)

https://docs.terrabyte.lrz.de

Publicly available soon!



Release 1.2 – coming soon!

New tools and services

- Public documentation website
- Regular export of STAC metadata to GeoParquet files
- Private STAC API for users and projects
- terrapi command line and Python library
- Continuous data ingestion (hourly/daily)

GeoParquet STAC metadata files

```
cql2_filter = {
    "op": "and",
    "args": [
        {
            "op": ">=",
            "args": [
                {
                    "property": "end_datetime"
                },
                '2020-03-28T20:05:46+02'
            ]
        },
        {
            "op": "<=",
            "args": [
                {
                    "property": "start_datetime"
                },
                '2020-03-28T22:06:15+02'
            ]
        },
        {
            "op": "=",
            "args": [
                {
                    "property": "sar:instrument_mode"
                },
                'IW'
            ]
        }
    ]
}
```

```
from pystac_client import Client as pystacclient
catalog = pystacclient.open('https://stac.terrabyte.lrz.de/public/api')
results = catalog.search(
    collections=['sentinel-1-grd'],
    filter=cql2_filter
)
items = results.item_collection()
```

You will query the central catalogue web service, which has limited resources shared with all requests!

```
import duckdb
geoparquet = '/dss/dsstbyfs01/pn56su/pn56su-dss-0008/Sentinel-1/GRD/geoparquet/*.parquet'

sql_where = DuckDBEvaluator({
    'sar:instrument_mode': 'sar:instrument_mode',
    'end_datetime': 'end_datetime',
    'start_datetime': 'start_datetime',
}, {}).evaluate(json_parse(cql2_filter))

sql_query = f"SELECT * FROM '{geoparqet}' WHERE {sql_where}"
df = duckdb.query(sql_query)
```

You will query static vector files (GeoParquet) with your own computing resources!

Release 1.2 – coming soon!

New tools and services

- Public documentation website
- Regular export of STAC metadata to GeoParquet files
- Private STAC API for users and projects
- terrapi command line and Python library
- Continuous data ingestion (hourly/daily)

terrabyte Release Version 1.2

V1 Summer 2023 | V1.2 Spring 2024 | V1.4 Autumn 2024 | Backlog

Infrastructure		
COMPLETED	Compute extension	Additional compute resources have been made available.
COMPLETED	Storage extension	Additional storage capacity have been made available for curated datasets.
Services		
COMPLETED	2FA	Login to all services with Two-Factor-Authentication
COMPLETED	Access for external project partners	External project partners have the possibility to get a terrabyte / LRZ account for project collaborations.
COMPLETED	Remote Desktop on compute nodes	The remote desktop application with QGIS is now available on terrabyte compute nodes.
DEVELOPING	STAC API for user and project data	To allow users to create and manage their own STAC collections and items, we will extend the STAC API with user authentication and enable per user private as well as shared data catalogs.
DEVELOPING	STAC metadata exports	Curated data catalogues are available in GeoParquet/DuckDB files in addition to the STAC API.

https://docs.terrabyte.lrz.de/about/roadmap/

Software		
DEVELOPING	terrabyte library `terrapi`	We will compile general functions to use our terrabyte services into a Python library.
Data		
DEVELOPING	Continuous data ingestion #1	New EO data are ingested hourly for Sentinel-1, Sentinel-2, and Landsat Collection 2 Level-2 data.
PLANNED	Continuous data ingestion #2	New EO data are ingested continously for Sentinel-3, Sentinel-5p, MODIS, and VIIRS data.
DEVELOPING	Roadmap for data provisioning	A roadmap for data availabilities will be provided.
Support		
DEVELOPING	Documentation webpage	To allow external users to access the documentation, we will migrate from the Wiki to a dedicated documentation webpage, which is accessible to all users with a terrabyte account.
DEVELOPING	Workshops	We will provide workshops for specific topics (e.g., STAC API, Data Transfers, Data Cubes, Containers, Environments)
COMPLETED	Onboarding workshops	We will start with workshops about how to use terrabyte.
SCHEDULED	terrabyte Café	We plan to have a regular exchange with terrabyte users for questions & answers and exchange between users within a terrabyte Café.

DLR

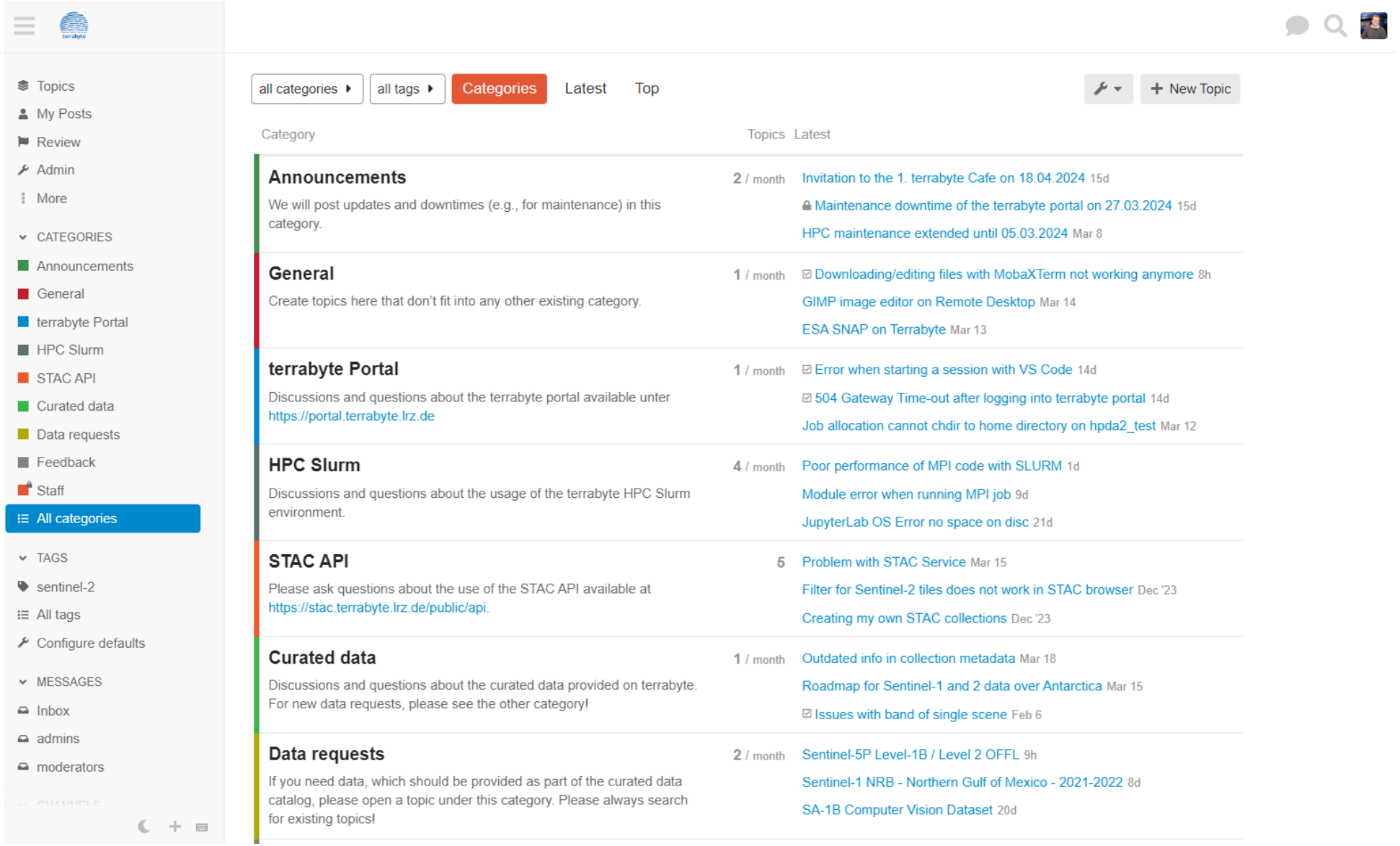
https://docs.terrabyte.lrz.de/about/roadmap/

terrabyte Release Version 1.4

V1 Summer 2023 | V1.2 Spring 2024 | **V1.4 Autumn 2024** | Backlog

DEVELOPING	HPC Slurm REST API	HPC Slurm commands can be send via a RESTful web API.
DEVELOPING	OGC EO Application Package	Allow users to develop and run OGC EO Application Packages on terrabyte HPC.
DEVELOPING	Data ordering from D-SDA	Data from the D-SDA archive can be automatically loaded on demand via a dedicated web service.
SCHEDULED	STAC API optimizations	Based on your feedback and our experiences with the initial version of the STAC API we will further work on performance optimizations.
SCHEDULED	Extended Python tutorials	We will further extend the basic Python tutorials. Please also share your experiences with us!
DEVELOPING	OpenEO API	Provide an OpenEO API for data processing on terrabyte HPC.
DEVELOPING	S3 interface for data storage	Offer an S3 compatible interface to read and write within Data Science Storage containers.

https://docs.terrabyte.lrz.de/about/roadmap/

You have questions and issues?
Please use our Support-Forum!

https://support.terrabyte.lrz.de