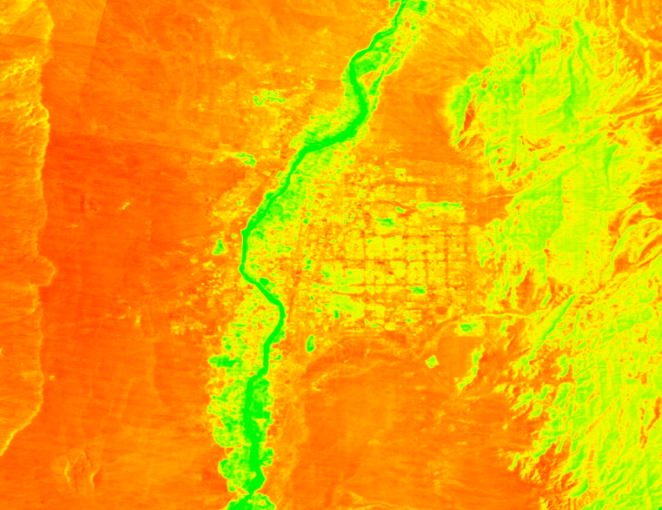
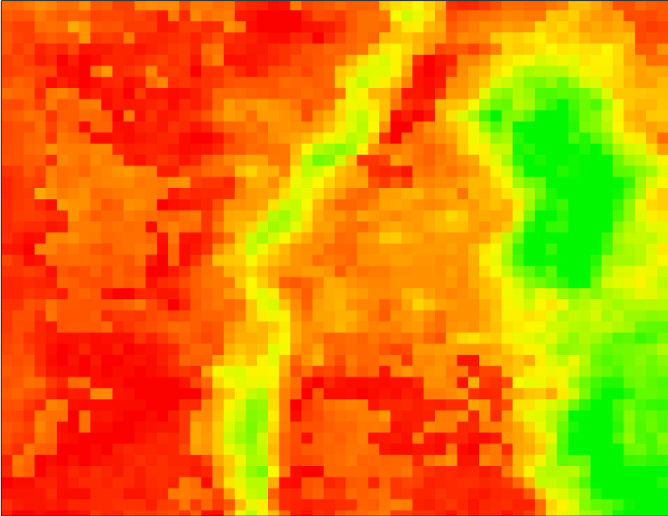
## horizontal line

ET Research & Science (ETRS)

Projects: FEWS NET, WaterSMART, and Landsat WaterBalance

# Overview

Increasing demand for ET data at a higher temporal and spatial resolutions is leading to the need of more storage to be able to process ET over varying extents and time periods.

# Goal

Have a designated Drive (in SMB3) with 100 TB of storage to meet project requirements.

# Specifications

Requesting adequate storage space for ET Research & Science efforts for projects: FEWS NET, WaterSMART, and WaterBalance. Currently we are sharing the location FEWSPSNFS2 on Stornext, which has about 64.9 TB, with the FEWS PPG team. The allocated storage for WaterSMART project is currently borrowing from FEWS because of our storage needs. Therefore, we would like to request a storage drive with 100TB for our ET Research and Science efforts.

## 

## 

## Current Storage:

## 

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2015 assigned space** | **Current used space as of 12/10/2018** |  |
| **WaterSmart** | **20 TB** | **34.28 TB** |  |
| Data |  | 12 TB | hosts all input data requirement for ET processing accessible to all team members therefore avoiding duplicate copies for individual users |
| Users |  | 18 TB | Individual team member processing environments |
| **FEWS PPG** | **43.50** | **26 TB** |  |

## 

## Future Storage needs (as of December 2018):

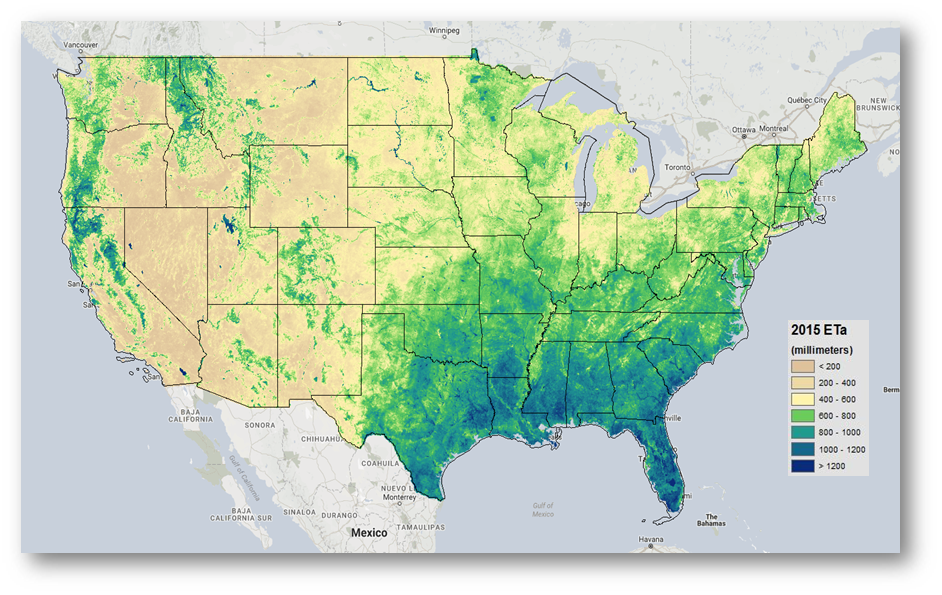
Note: All storage needs here are our estimates to the best of our knowledge of project needs and can vary.

|  |  |  |
| --- | --- | --- |
| **FEWS PPG** | **64.9 TB** | For PPG and KnowledgeBase |

🔺

🔻

|  |  |  |
| --- | --- | --- |
| **ET Research and Science (ETRS)** |  |  |
| **Data** | **20 TB** | Shared data repository for all team members |
| **MODIS ET** | **20 TB** |  |
| Global 10 day ET model research and development | 10 TB | Ca 1 TB for the ETa product incl all seasons/all years  Processing ET currently and needing 6 TB for intermediate data  Operational data stored on FEWS PPG and scienceweb1 |
| Global Daily ET product research and development | 10 TB | New project deliverable for WaterSMART  Ca 5 TB for all 17 years just ET without considering additional products |
| **Landsat ET** | **60 TB** | Estimated 60 TB is the lower limit for expected onsite product storage. This includes gridded Daily Landsat-scale ET and Count/QA bands for CONUS extent. 30m resolution, 1985-present. |
| For CONUS and Global Landsat ET, processing of the archive including inputs (Remote Sensing + global weather datasets) and model intermediate data is fully implemented within Earth Engine cloud platform | 50 TB | <https://developers.google.com/earth-engine/datasets/>  <https://developers.google.com/earth-engine/datasets/catalog/landsat/>  <https://github.com/Open-ET/openet-ssebop-beta> |
| For Water Balance analysis and ET partitioning work at both Landsat and MODIS scale | 10 TB |  |



**Landsat-based 30m Annual ET product for CONUS (preliminary)**

Discussion Questions:

1. Increased storage is going to have additional costs to our projects, that much we know. But we have not discussed money or funding with the government staff (are we allowed to?). We have talked to the PI Gabriel Senay and he has communicated that we will continue to scale up our processing and data needs and will continue to do so as time goes on (perhaps CONUS-wide Landsat-scale monthly ET to be processed and then put up on the SSEBop website in the end-2019 timeframe; daily CONUS Landsat ET in 2020 perhaps)
   1. Is there money or funding available from the government due to our increasing data storage needs? (Gabriel may not be the person to ask; Jim Rowland? Vaughn Ihlen?)
   2. How do we request for that?
2. What options for storage are available to us?
   1. What is the maximum allowable data storage we can acquire in the 2nd quarter of 2019?
   2. Would that storage be expandable?
   3. How easy is it to increase storage capacity in the future?
   4. Is there a maximum limit?
3. We are beginning to host ET on the SSEBop website. This data is stored on scienceweb1.
   1. How is that handled if we add 15 TB of data to the website? Or how about more than 15 TB in the years to come (2019-2021)?
4. What is the expected life-cycle of the current Stornext disk infrastructure which houses existing project datasets and files (March 2019?)?
   1. What actions should we do to prepare for a network storage transition?
5. If expandable storage is not possible or too expensive on standard server storage, what are the cloud-based solutions available such as Amazon or Google cloud storage buckets?
   1. Currently we collaborate with external partners from the Desert Research Institute who uses Google Cloud Storage. Would that be a viable cloud storage solution for us to use long term? GCS is an ideal solution for integrating a seamless pipeline with cloud processing platforms such Earth Engine, and is also [FedRAMP compliant](https://marketplace.fedramp.gov/#/product/google-services-google-cloud-platform-products-and-underlying-infrastructure?sort=productName). Is that something that the USGS would agree to going forward or would they insist on Amazon due to prior commitments?
   2. Would Amazon cloud storage be more affordable or even available for our use?
   3. How would access to that storage work with our processing and with posting data to the website especially in terms of access?
   4. USGS Data Release standards seem to imply that data that is hosted on a USGS website need to be physically located on USGS servers. How do we reconcile that if the cloud-based solution is the best?
6. What are the next steps we need to take?