Opening Bathymetric Data to the Public: Optimizing Big Data for Access through the Web

Caitlyn Raines

Esri

Geodata Services

Current GEBCO download

- Start at the GEBCO website
- Define an area of interest
- Add that data to a basket
- Get redirected to the BODC site
- Download it as a netCDF, geoTIFF or Esri ASCII raster

Or you can access it through a web service provided by a 3rd party

GEBCO by the numbers..

- **GEBCO 2014**

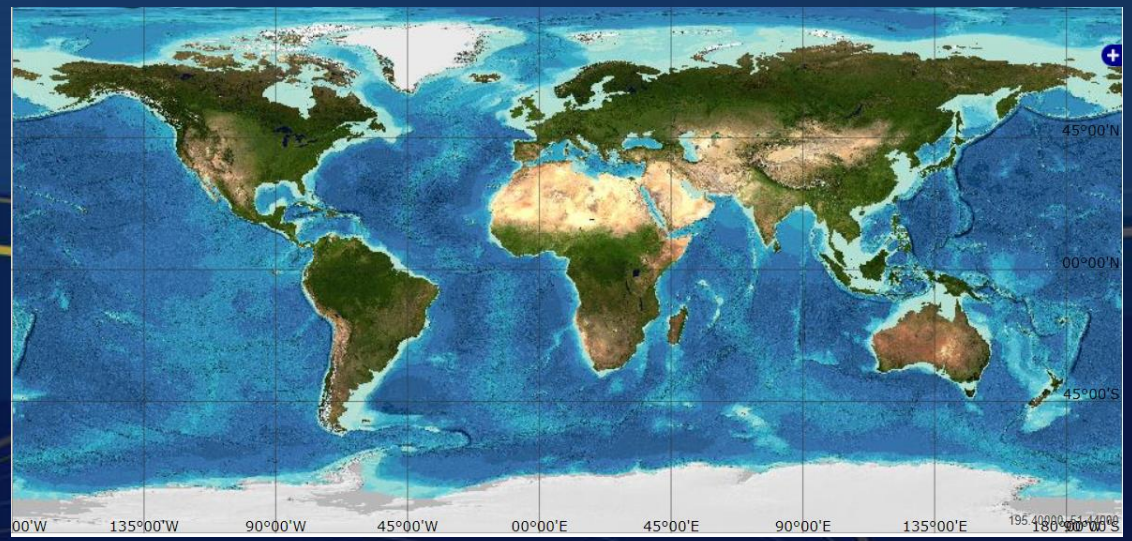
(30 arc-second grid)

  - **43200 columns/ 21600 rows**
  - **Compressed 6.95 GB**

- **Seabed 2030**

(100 meter grid)

  - **400,750,000 columns/ 200,000,000 rows**
  - **No direct correlation, but could increase this by an order of 5**



Suddenly the idea of just downloading and storing the data is less tenable…

Thinking about Data Storage for Seabed 2030

- **Format**
  - Easily read
  - Platform agnostic
  - Easy to compress and decompress
- **Access**
  - Easily discoverable
  - Fast
  - Open to the public
  - Aligned with new and emerging technologies

MetaRasterFormat

Limited Error Raster Compression

Data format

MRF - LERC

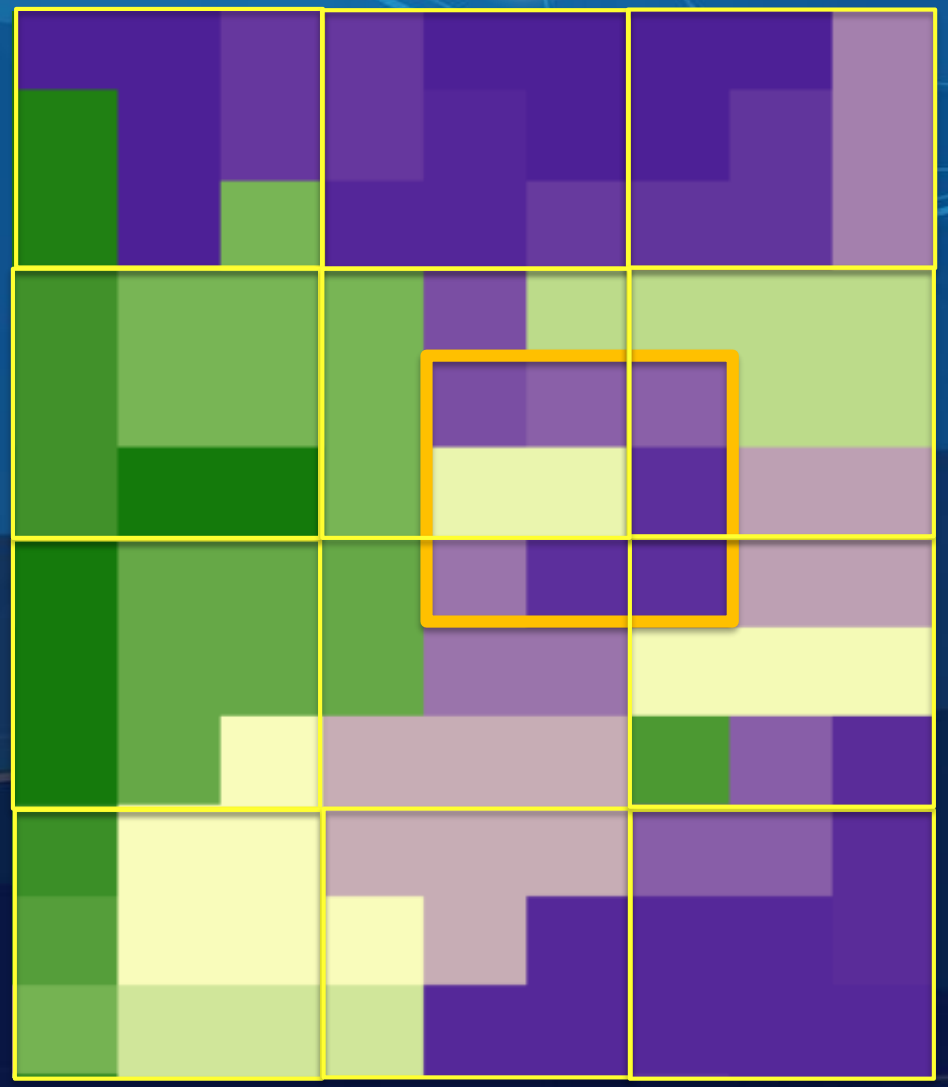
Compression

NASA JPL

Esri

Traditional Raster Data formats vs MRF

- Traditional raster data formats (netCDF, geoTIFF, BAG)
  - 2 part files
    - XML Header
    - Raster
- MRF
  - 3 part file
    - XML header
    - Tiled raster files
    - Index file
  - Can access information on the individual tiles



The Brilliance of LERC

- Lossy compression algorithm designed specifically for elevation
- User-defined compression
  - Pixel will only be allowed to change by the tolerance defined by the user
  - In the real world unit of the raster
- User controls the error introduced by the pixel
- Explicit Data Mask optimizes compression in sparse or projected swath data
- LERC decompression can also can be implemented in JavaScript

MRF - LERC

Fast

Controlled Error

GDAL Compliant

Platform Agnostic

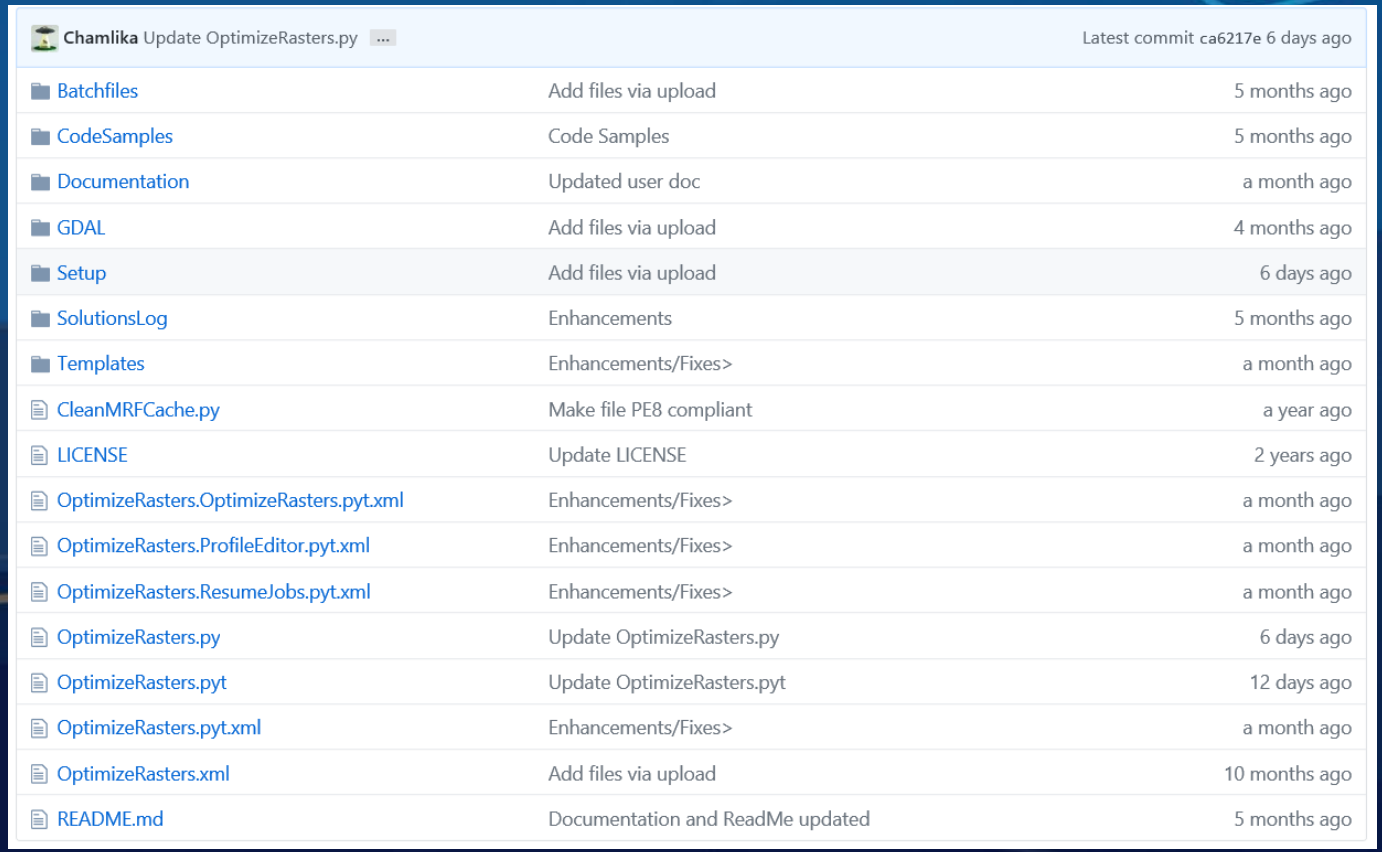
Web Service Friendly

Cloud Storage Friendly

Documented and Tested

Optimize Rasters

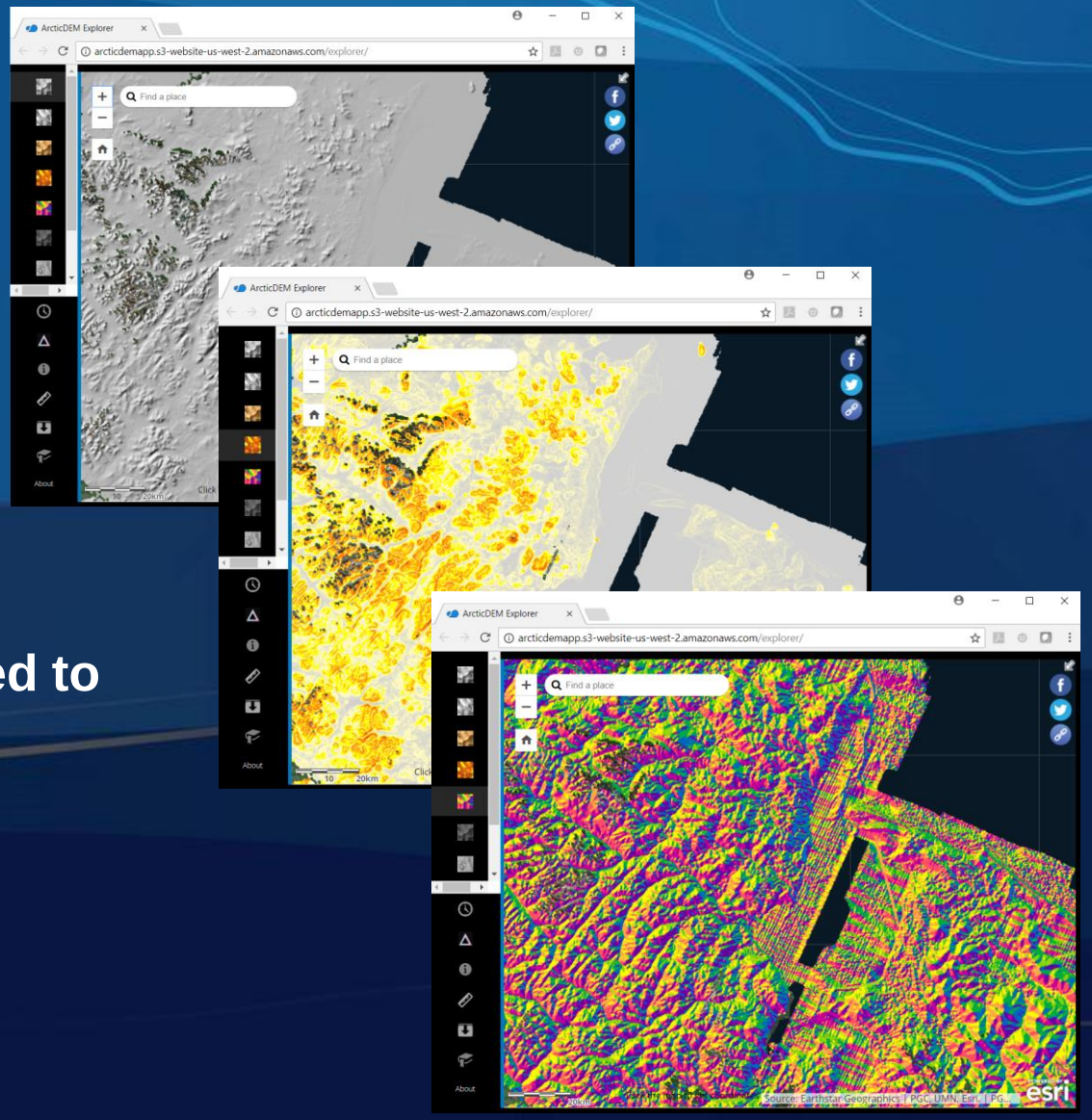
- Free to the public
  - No Esri License? Run by command line
  - Esri License? Run as Geoprocessing tool
- Available on GitHub
- Can convert files locally or stored in cloud storage



Chamlika Update OptimizeRasters.py ...		Latest commit ca6217e 6 days ago
Batchfiles	Add files via upload	5 months ago
CodeSamples	Code Samples	5 months ago
Documentation	Updated user doc	a month ago
GDAL	Add files via upload	4 months ago
Setup	Add files via upload	6 days ago
SolutionsLog	Enhancements	5 months ago
Templates	Enhancements/Fixes>	a month ago
CleanMRFCache.py	Make file PE8 compliant	a year ago
LICENSE	Update LICENSE	2 years ago
OptimizeRasters.OptimizeRasters.pyt.xml	Enhancements/Fixes>	a month ago
OptimizeRasters.ProfileEditor.pyt.xml	Enhancements/Fixes>	a month ago
OptimizeRasters.ResumeJobs.pyt.xml	Enhancements/Fixes>	a month ago
OptimizeRasters.py	Update OptimizeRasters.py	6 days ago
OptimizeRasters.pyt	Update OptimizeRasters.pyt	12 days ago
OptimizeRasters.pyt.xml	Enhancements/Fixes>	a month ago
OptimizeRasters.xml	Add files via upload	10 months ago
README.md	Documentation and ReadMe updated	5 months ago

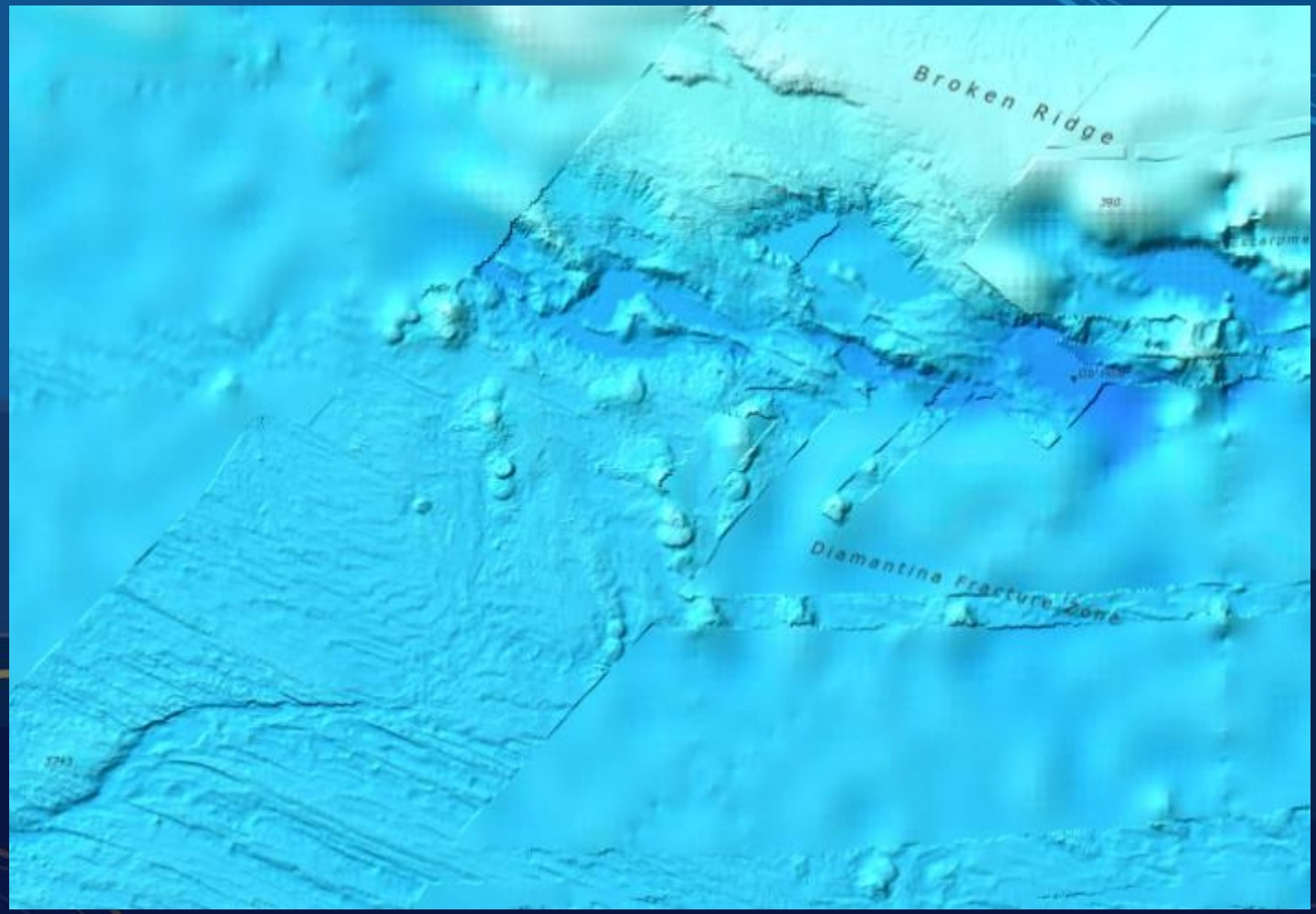
Data Access through Web Services

- Increased access for:
  - New Industries
  - Applications
- Increased metrics
- Analysis in the cloud – no longer limited to desktop computing
- Platform agnostic



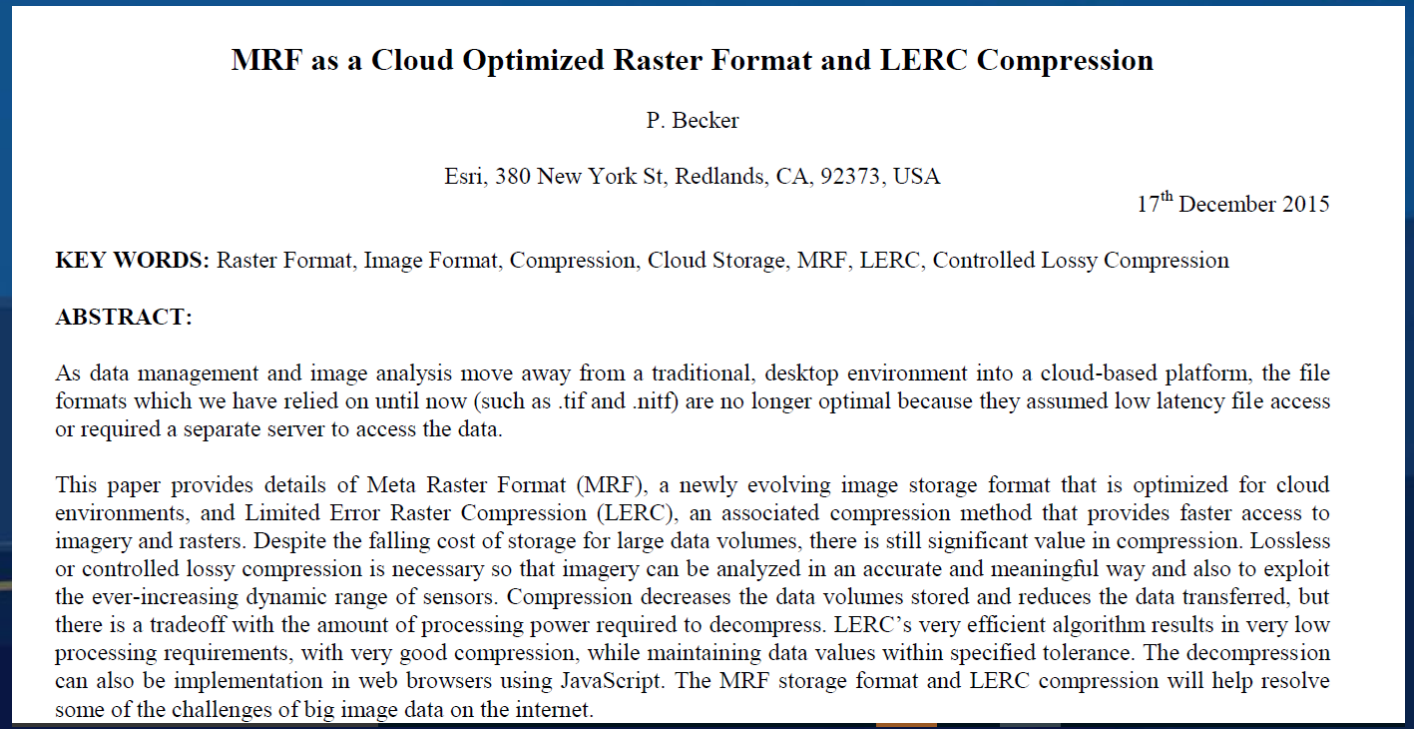
Recommendations for moving forward

- MRF-LERC
- Cloud Storage
- Web Services



Want to Learn More?

- ArcGIS Blog
- GeoNet Community
- Variety of Videos at Esri 360
- GitHub pages for MRF and LERC



MRF as a Cloud Optimized Raster Format and LERC Compression

P. Becker

Esri, 380 New York St, Redlands, CA, 92373, USA

17th December 2015

KEY WORDS: Raster Format, Image Format, Compression, Cloud Storage, MRF, LERC, Controlled Lossy Compression

ABSTRACT:

As data management and image analysis move away from a traditional, desktop environment into a cloud-based platform, the file formats which we have relied on until now (such as .tif and .nitf) are no longer optimal because they assumed low latency file access or required a separate server to access the data.

This paper provides details of Meta Raster Format (MRF), a newly evolving image storage format that is optimized for cloud environments, and Limited Error Raster Compression (LERC), an associated compression method that provides faster access to imagery and rasters. Despite the falling cost of storage for large data volumes, there is still significant value in compression. Lossless or controlled lossy compression is necessary so that imagery can be analyzed in an accurate and meaningful way and also to exploit the ever-increasing dynamic range of sensors. Compression decreases the data volumes stored and reduces the data transferred, but there is a tradeoff with the amount of processing power required to decompress. LERC's very efficient algorithm results in very low processing requirements, with very good compression, while maintaining data values within specified tolerance. The decompression can also be implementation in web browsers using JavaScript. The MRF storage format and LERC compression will help resolve some of the challenges of big image data on the internet.

Thank you

Questions?

Caitlyn Raines, craines@esri.com