



# **USGEO Data Management WG Cloud Computing Recommendations for Earth Observation Data**

*- preview of draft document*

**Co-chairs: Jeff de La Beaujardiere (NOAA), Mark McInerny (NASA), Mike Frame (USGS)  
+ other DMWG members & contributors**

2017 July 27  
ESIP Summer Meeting



# About USGEO DMWG

- White House Office of Science & Technology Policy (OSTP)
  - Committee on Environment, Natural Resources, and Sustainability (CENRS)
    - **US Group on Earth Observations (USGEO)**
      - **Data Management Working Group (DMWG)**
        - *NOAA, NASA, USGS co-chairs*
        - *EPA, DOE, NSF members*
- Issued *Common Framework for Earth Observation Data* (2016) after pre-brief at ESIP 2015 (Monterey)
  - Standards for data discoverability, accessibility, usability
  - Focus on data hosted on-premises



# About USGEO DMWG

- White House Office of Science & Technology Policy (OSTP)
  - Committee on Environment, Natural Resources, and Sustainability (CENRS)
    - **US Group on Earth Observations (USGEO)**
      - **Data Management Working Group (DMWG)**
        - *NOAA, NASA, USGS co-chairs*
        - *EPA, DOE, NSF members*
- Issued *Common Framework for Earth Observation Data* (2016) after pre-brief at ESIP 2015 (Monterey)
  - Standards for data discoverability, accessibility, usability
  - Focus on data hosted on-premises
- New draft document focuses on data hosted in Cloud



# Document Scope and Status


- Focus: "If you have decided to use the Cloud, then consider the following recommendations."
  - *Not:* "This is why you should use the Cloud."
- Medium level of technical depth
  - Not a tutorial on how to use the Cloud
- Vendor-neutral, avoid naming specific vendor services
- Draft v0.3.3 in revision following comments from subject-matter experts (SME) in DMWG member agencies
  - Not yet achieved unanimous approval on content
- SME input from ESIP Cloud Computing Cluster welcome
  - Contact [jeff.deLaBeaujardiere@noaa.gov](mailto:jeff.deLaBeaujardiere@noaa.gov) if interested
- Future version will be cleared by USGEO and issued for public comment (RFI)



# Document Contents

- Introductory Material:
  - Definitions
  - Potential Cloud Benefits
  - Potential Cloud Disadvantages
- Recommendations:
  - Assessing When to Use the Cloud
  - Transferring Data to the Cloud
  - Data and Metadata Contents
  - Developing Applications in the Cloud
  - Cost Minimization
  - Security Considerations
  - Monitoring and Metrics
  - Agency Support
  - Earth Observations





# Highlights/Summary of Recommendations



# Assessing When to Use the Cloud

- Not all services suitable for migration to the Cloud.
  - Consider cost, performance, and development and maintenance effort for on-premises vs Cloud deployments.
    - Actual on-premises data center expenses include not only hardware but real estate, power and cooling, system administration, IT security updates, and network.
    - Actual Cloud expenses include storage, computation, egress, and personnel with relevant skills.
- Public data access and non-real-time analyses may be best suited for Cloud.
- Operational data processing and forecasting may be best done with dedicated on-premises hardware.



# Transferring Data to Cloud

- Consider alternate methods for large volumes.
  - Shipping physical media, dedicated network connection
- Traditional subset services (e.g., DAP, OGC) are not as relevant for bulk transfer.
- Provide inventory of all expected files.
- Provide notification of new/updated files.



# Data & Metadata

- Provide machine-readable **metadata** records.
- Include **versioning** information in metadata records and object metadata.
- Provide **checksums** or other means of verifying data integrity.
- Utilize **standard formats**.
- Provide **Cloud-enabled tools** to read data files.



# Developing Applications

- Use **Cloud-native** services when possible.
- Provide **containerized** applications when Cloud-native tools not available.
- Use **automation** to establish, scale, and monitor Cloud configurations.
- **Modularize** applications and loosely-couple components for scalability.
- **Prepare for failure** of individual components.
- Treat computing resources as **disposable**.



# Cost Minimization

- **Optimize configuration** to reduce costs of computing, storage, and egress.
- Use **cost controls** to avoid unbounded costs – for example, by monitoring traffic and limiting access if cost threshold is reached.
- Encourage **computing in place** rather than users downloading data to their local system (which would incur data egress charges).
- Consider having **users pay data egress** charges (instead of your agency paying them) if users choose to download instead of computing in place.



# Security

- Use only Cloud vendors and services authorized by Federal Risk and Authorization Management Program (**FedRAMP**).
- **Consider security** from the very beginning of design.
- Use Cloud vendor's **built-in security tools** for identify management, access management, and application firewalls.



# Agency Support

- Designate technical **point of contact** for each dataset.
- Establish **communication channel** to inform users of important updates or outages.
- **Prototype early**, leveraging free tier for low-volume applications.
- Understand that staffing roles and training are **different for Cloud** vs on-premises applications.
- **Consolidate logistics** functions (procurement, billing, authority to operate, etc) on behalf of multiple Cloud projects in agency.
- Consider **multi-agency collaboration**.



# Earth Observations

- **Collaborate** withing USGEO DMWG to share best practices and maximize compatibility.
- Provide data in **common coordinate systems** and projections.
- Establish **data index** to rapidly find data files based on location, time, observed quantity, etc.
- Use **consistent naming conventions** for objects to facilitate integration of data from multiple sensors.
- Include **EO-specific attributes** in dedicated object metadata fields.
- Enable retrieval of **full geospatial metadata** records.



# Summary

- This briefing includes only highlights from the complete (draft) document.
- Subject-matter expert input from ESIP Cloud Computing Cluster welcome
  - Contact [jeff.deLaBeaujardiere@noaa.gov](mailto:jeff.deLaBeaujardiere@noaa.gov) if interested