ABoVE Science Cloud

Peter Griffith
Chief Support Scientist

Liz Hoy
Support Scientist

Carbon Cycle & Ecosystems Office

Mark McInerney
Data Services Lead

Dan Duffy
NCCS HPC Lead

Computational and Information Sciences and Technology Office (CISTO)

Contributors

High Performance Computing/ NASA Center for Climate Simulation (NCCS)

- Scott Sinno, System Architect and System Administrator
- Hoot Thompson, System Architect and System Administrator
- Garrison Vaughn, System Architect and Applications Engineer
- Brittney Wills, Computer Scientist
- Ellen Salmon, Computer Research and Development

Climate Model Data Services (NASA CDS)

- Laura Carriere, System Analyst
- Steven Ambrose, Principal Systems Engineer
- Julien Peters, Software Developer
- Eric Winter, Software Developer

And… but… therefore

- Science datasets are becoming more and more extensive, with intensive computation needed for data processing
- **AND** collaboration within diverse research groups is often needed,
- **BUT** it is often time consuming and expensive to process and share data with others
- **THEREFORE** the ABoVE Science Cloud (ASC) was created to meet the needs of ABoVE investigators and encourage collaboration within the field campaign.

Advanced Data Analytics Platform (ADAPT)
"High Performance Science Cloud"

High Performance Science Cloud is uniquely positioned to provide data processing and analytic services for NASA Science projects. A portion of ADAPT is dedicated to ABoVE (the ABoVE Science Cloud).

Adjunct to the NCCS HPC environment

- Lower barrier to entry for scientists
- Customized run-time environments
- Reusable HPC/Discover hardware

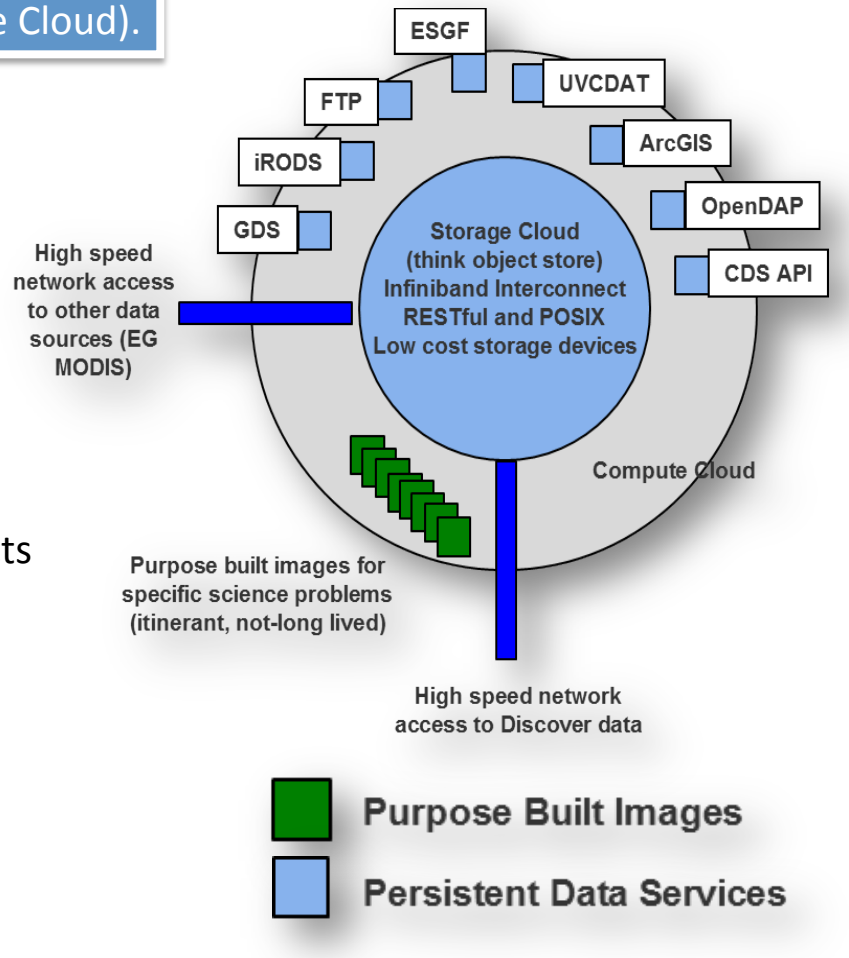
Expanded customer base

- Scientist brings their analysis to the data
- Extensible storage; build and expand as needed
- Persistent data services build in virtual machines
- Create purpose built VMs for specific science projects

Difference between a commodity cloud

- Platform-as-a-Service that comes close to matching HPC levels of performance
- Critical Node-to-node communication – high speed, low latency
- Shared, high performance file system
- Management and rapid provisioning of resources

High Performance Science Cloud Conceptual Architecture



ASC Services

- **Data Staging**
  - Common datasets in the ASC Environment
  - Folding ABoVE generated results into common ASC Environment
- **NGA/DigitalGlobe High Resolution Satellite Data Imagery staged on the ASC**
  - Available for US federal purposes
- **Distribution Services**
  - Products can be distributed within the Science Team
  - Long term data archive will be at a DAAC or other archive

Staged / Common Data Sets in ABoVE Science Cloud

Common datasets "Staged" for ABoVE investigators in ABoVE Science Cloud

- Staged and available for direct use
- Individual investigators don't have to invest time to locate and transfer data into system
- Avoids duplications of large datasets on system
- Additional datasets can be added, including generated data from ABoVE PI
- Data Services Manager to locate data

Example Staged datasets

- *Landsat*, Surface reflectance, 123 TB
- *MODIS*, Daily surface reflectance, 57 TB
- *MERRA* reanalysis, 80 TB
- NGA/DigitalGlobe dataset, >300 TB

Example Download Times For 80TB

ABoVE NGA/DigitalGlobe Data Services

- **Obtain Existing Imagery**: Alaska and entire Arctic north of 60N being tasked by ABoVE via NGA in stereo-panchromatic and multispectral
- **Direct tasking***: for ABoVE core and extended domain below 60N (2-3 year activity)
- **Bulk transfer** and store raw DigitalGlobe (NITF format) images in ABoVE Science Cloud
  - Ordering process
  - Networked connection direct from ABoVE Science Cloud to NGA
  - Over 300 TB now in ASC
- **Value Added NGA/DigitalGlobe Imagery***: Create ~0.5m panchromatic, orthorectified Arctic Mosaic of ABoVE core and extended domain including south of 60N
- **Elevation model*** of the ABoVE core and extended domain.

**Paul Morin will discuss DigitalGlobe datasets and products further in his presentation*

DigitalGlobe Imagery

- US National Geospatial Agency (NGA) licenses DigitalGlobe imagery which can be used for US federal government purposes

- NASA ABoVE researchers are able to access and analyze this imagery pursuant to the NextView license agreement NGA has developed with DigitalGlobe

- ABoVE researchers must sign the NASA-NGA data use policy acknowledging they understand the NextView license agreement terms

Imagery and Imagery Derived Products

Any new DigitalGlobe imagery or imagery derived products must contain **copyright** and **NextView License** notice.

Product Type	Subject to:	
	Copyright	NextView License
Imagery	✓	✓
Imagery Derived Product (e.g. image-based maps, reduced resolution images)	✓	✓
Derived Products* (e.g. maps, line drawings, DEM)	N/A	N/A

*Derived product methodology should attribute initial imagery to DigitalGlobe

Imagery Example:
Barrow Shoreline

© 2015 DigitalGlobe NextView License

ASC User Stats

- 20 new projects plan to use the ASC, 4 groups are continuing on the ASC
- ~50 new users, 13 current users
- Software needs (number of groups): IDL (9), MATLAB (4), PCI (3), E-Cognition (2), open source or NASA-licensed products such as R, ESRI, Python, and QGIS
  - Options for the use of non-NASA licensed products on the ASC are still being considered

Science Team Planned Geospatial Data Use

Common Satellite Datasets	No. of Projects
Landsat (surface reflectance, others?)	18
MODIS (multiple products)	15
DigitalGlobe Imagery	11
Radar: AMSR-E, SMMR, SSM/I, SSMIS, SAR, SAR/PALSAR/ERS, Radarsat-2, ERS, Sentinel, SMOS, SMAP, ALOS(2), GPR, InSAR, IfSAR	16

Variations in spatial and temporal scales across ABoVE projects.

Other satellite data mentioned: AVHRR, GPM, Hyperion, VIIRS, MISR, IceSat/IceSat-2/GLAS, GPM, Sentinel-2, GOSAT, OCO-2/3, GEDI, GRACE, Soumi NPP

Science Team Planned Geospatial Data Use

Aerial*
AHAP – 4 groups
LiDAR
CARVE
AirMoss

*Mentioned more than once

Other datasets/products mentioned: GIMMS3g, MEASURES, NASA/GEWEX Surface Radiation Budget, Reanalysis products (NARR, MERRA), WRF, HIPPO, AIRMETH, NAIP, AEM, PFA, ACG, ABLE3A, ARCTS-airborned, GMD

Ancillary Datasets
ASTER GDEM (or best available DEM) – 4 groups
Fire perimeters
Soils database
Climate variables
Lightning strikes
Lake location
Core variables (SNOTEL, CALM, AmeriFlux, BNZ LTER, GTN-P)
Vegetation/fuels maps
Pre-ABoVE datasets

Need for a Science Data Working Group

- ABoVE grid/projection for staging/sharing common datasets
- Data/metadata standards (for products and core variables, etc.)
- Developing DigitalGlobe tasking plan if needed
- Other?

Contact Information

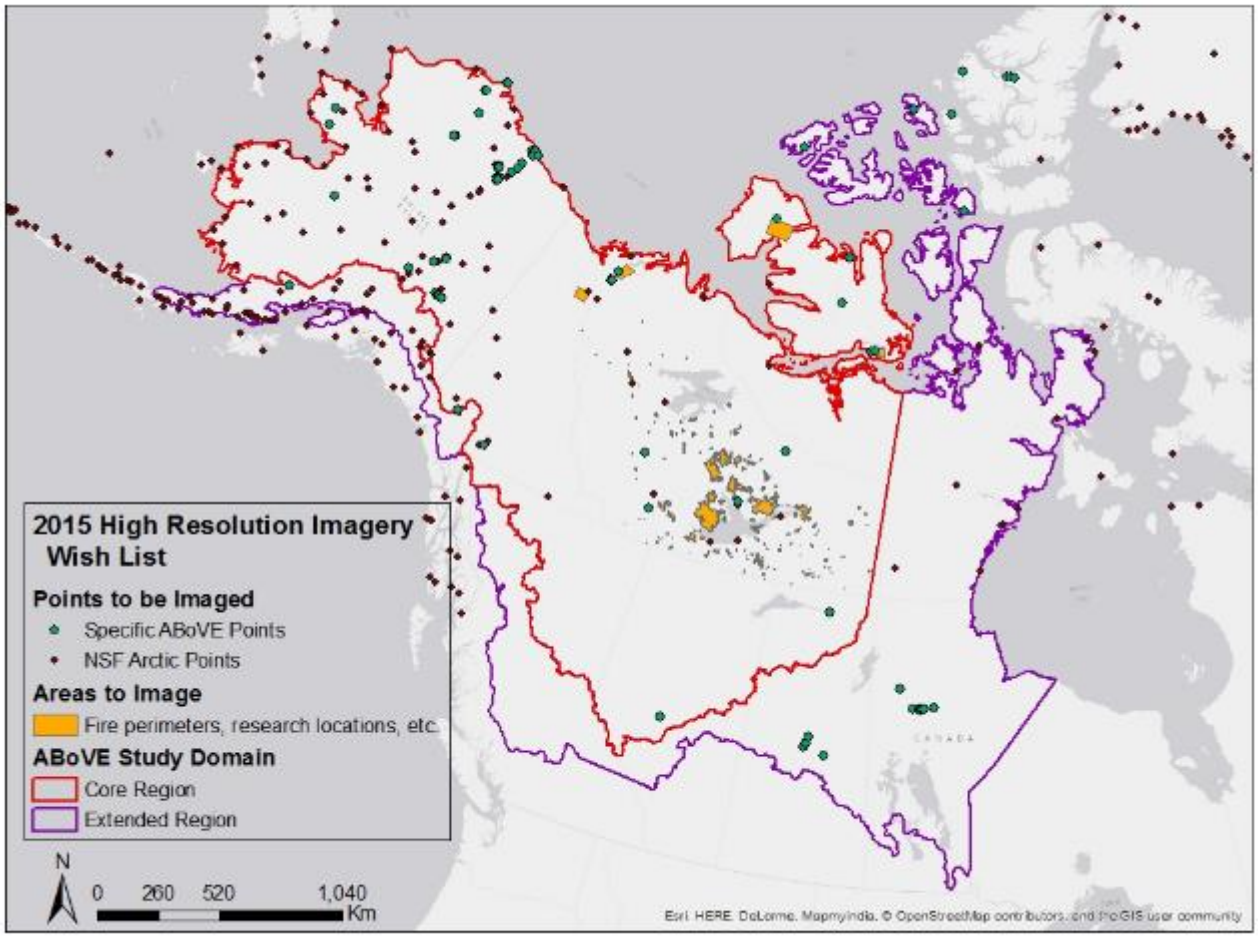
- Recent webinar on the ASC: http://above.nasa.gov/sciencecloud.html
- Contact Liz Hoy for:
  - Access to the ASC
  - Access to DigitalGlobe imagery
  - Input on DigitalGlobe tasking locations

elizabeth.hoy@nasa.gov, 301-614-6494

Additional Slides

ABoVE NGA/DigitalGlobe Imagery Tasking for 2015

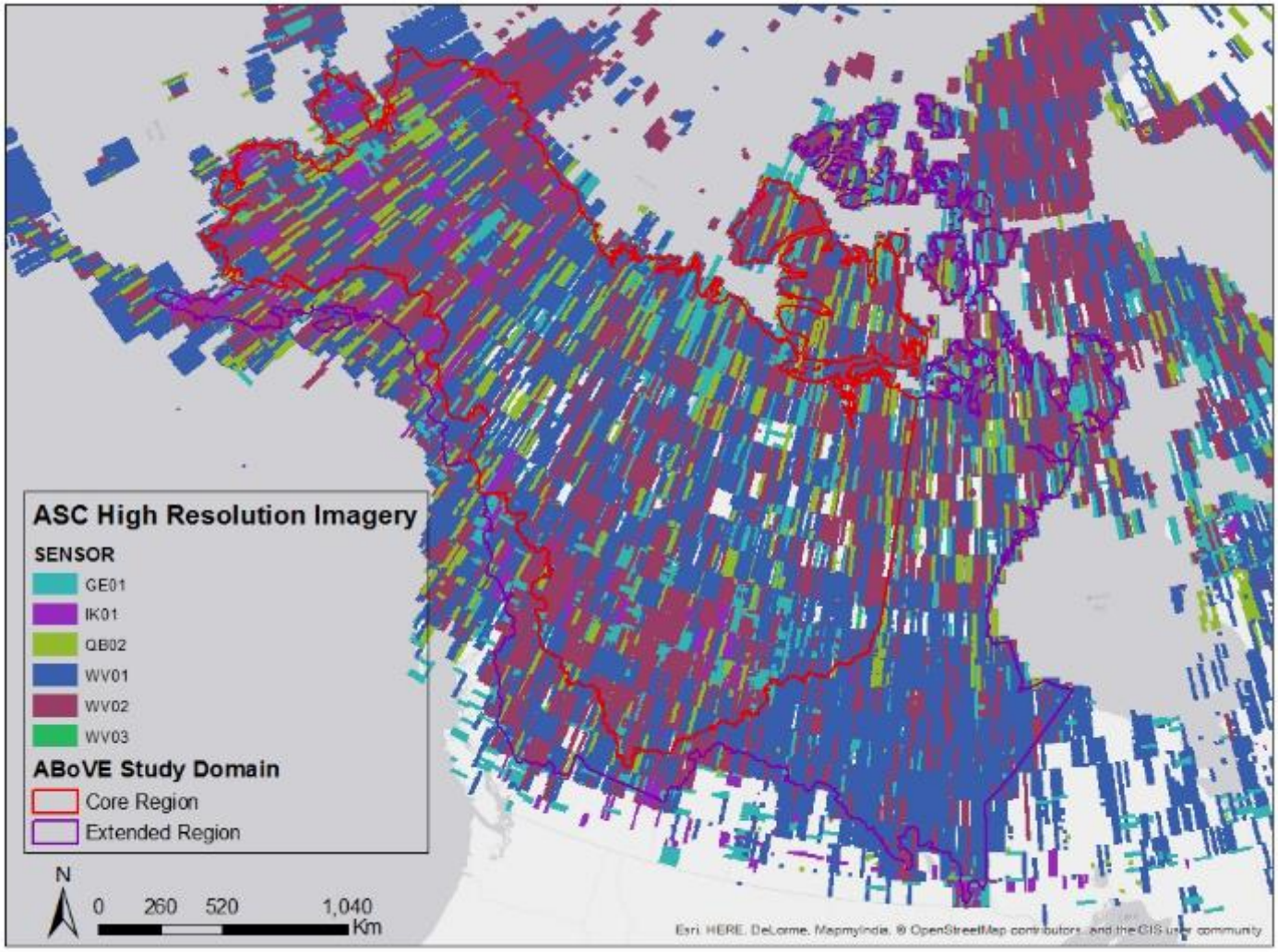
Tasking of new imagery coordinated through the CCE Office with help from PGC.



Tasking plan for 2016 will be developed with ABoVE Science Team support.

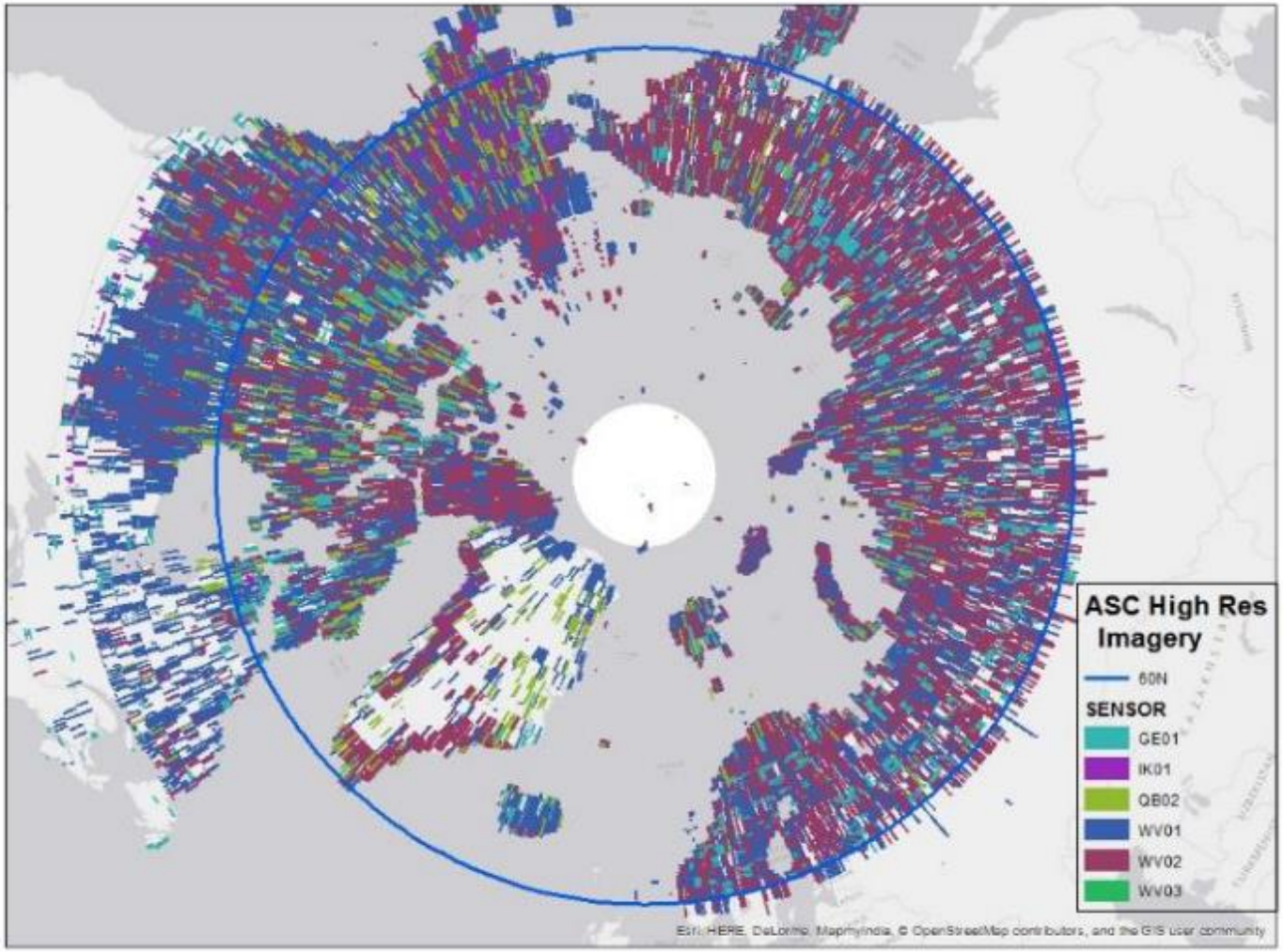
Interested research groups should contact Liz.

ABoVE Science Cloud DigitalGlobe Imagery: ABoVE Study Domain



Note: Imagery is included from all seasons and for years 1999-2015.

ABoVE Science Cloud DigitalGlobe Imagery: Circumpolar



Note: Imagery is included from all seasons and for years 1999-2015.

DigitalGlobe EnhancedView Web Hosting Service

ABoVE researchers have online access to view imagery through the DigitalGlobe EnhancedView Web Hosting Service

- Ability to view imagery, however not a space to process imagery
- Imagery from ~2011 and forward is available within the web-based system
- Must adhere to the NASA-NGA data policy and the NextView License to use
- Liz will discuss access with individual research groups

ESRI ArcGIS Services In NCCS Science Cloud

- Science cloud does have NASA enterprise licensed ESRI products installed
  - ArcGIS for Desktop
  - ArcGIS for Server
  - ArcGIS for Portal

- USE 1: ArcGIS services to stage select data by Carbon Cycle & Ecosystem office
  - NGA/DigitalGlobe, other

- USE 2: ESRI application / system level support for ABoVE projects
  - EG. Grant scientist / project ArcGIS server account for data publication (scientist/ project managed)

NASA Climate Model Data Services
Data Publication and Distribution Services

Data Publication Services	Protocol	Download	Subsetting	2D Visualization
Web Access For downloading small files	HTTP	✓		
File Transfer Protocol (FTP) Anonymous FTP supporting wget	FTP	✓		
GRads Data Server (GDS) Data subsetting and analysis services	OPENDAP	✓	✓	
Live Access Server (LAS) Data subsetting and analysis services	OPENDAP	✓	✓	
THREDDS Data Server (TDS) subsetting , analysis, & visualization	OPENDAP	✓	✓	✓
Earth System Grid Federation (ESGF) Data access to IPPC CMIP data	OPENDAP	✓		✓
Web Map Service (WMS) Data publication to IPPC CMIP Format	OPENDAP	✓	✓	✓