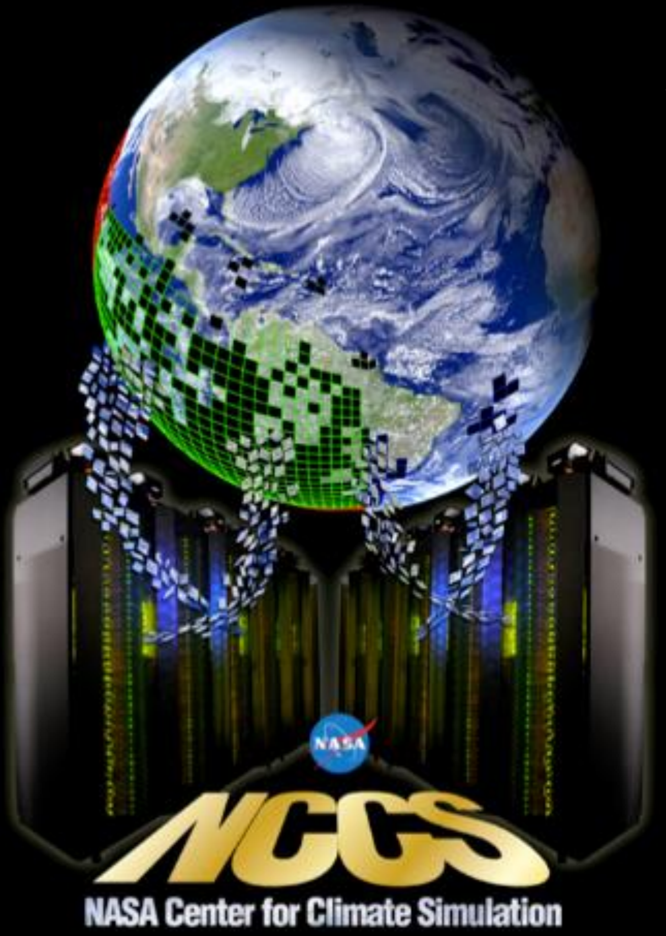
ABoVE Science Cloud on ADAPT/Explore

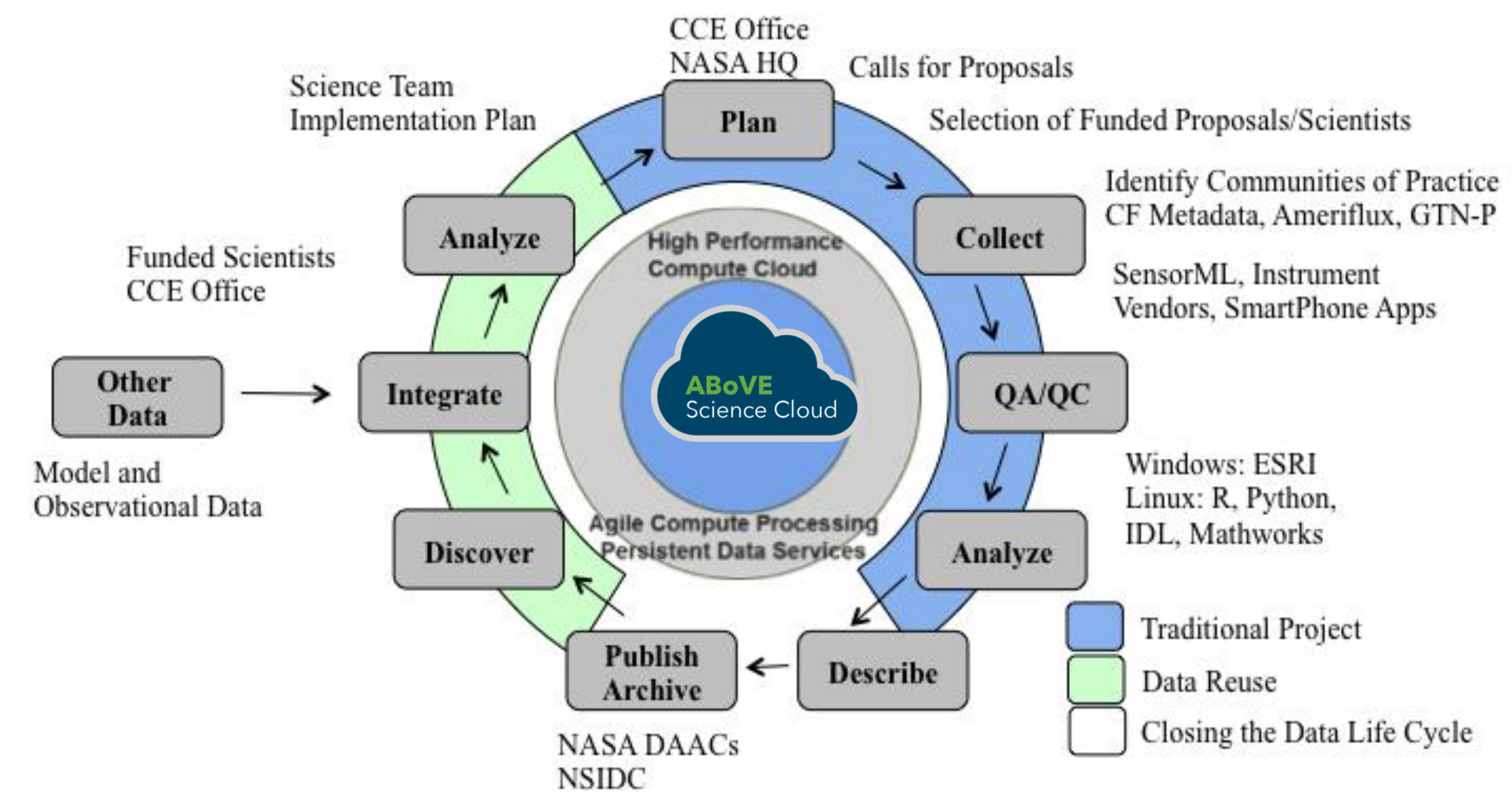
Liz Hoy - NASA Carbon Cycle and Ecosystems Office
Jim Shute, Ryan Forbes, Ellen Salmon, Matt Stroud & others - NCCS Team
Mark Carroll - NASA Data Science Group and Innovation Lab
& ABoVE Science Team



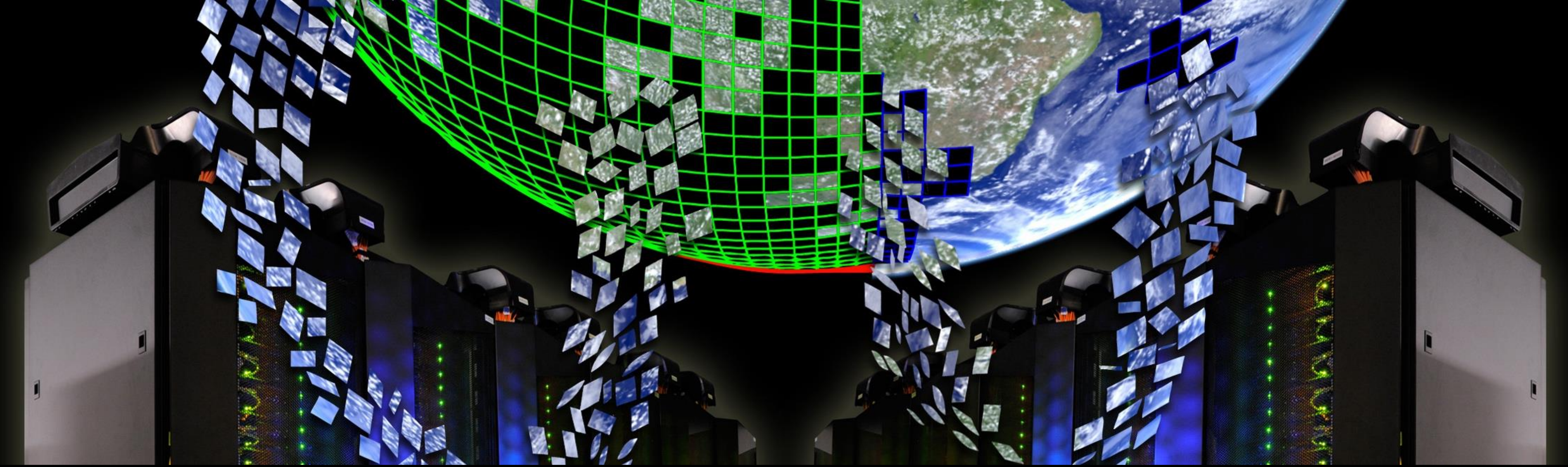
Overview

- Background
- NCCS Capabilities
- Large Datasets
- Analysis Ready Products from the Innovation Lab
- ASC Research Highlights
- Account Setup
- Demos
- Questions

ABoVE Data Workflow



Augmented from Rüegg et al 2014 in *Front Ecol Environ*



NASA Center for Climate Simulation (NCCS) Geospatial Capabilities Briefing

Jim Shute
NCCS Cloud Computing and Data Services (CCDS) Lead
james.k.shute@nasa.gov

Agenda

- NCCS Mission
- NCCS Systems
- Functional Area Overviews
- Spatial Platform
- GIS Development Paradigm
- GIS Development Options
- Example Workflows
- Accessing the Systems
- Demo
- Strengths and Challenges

NCCS Mission

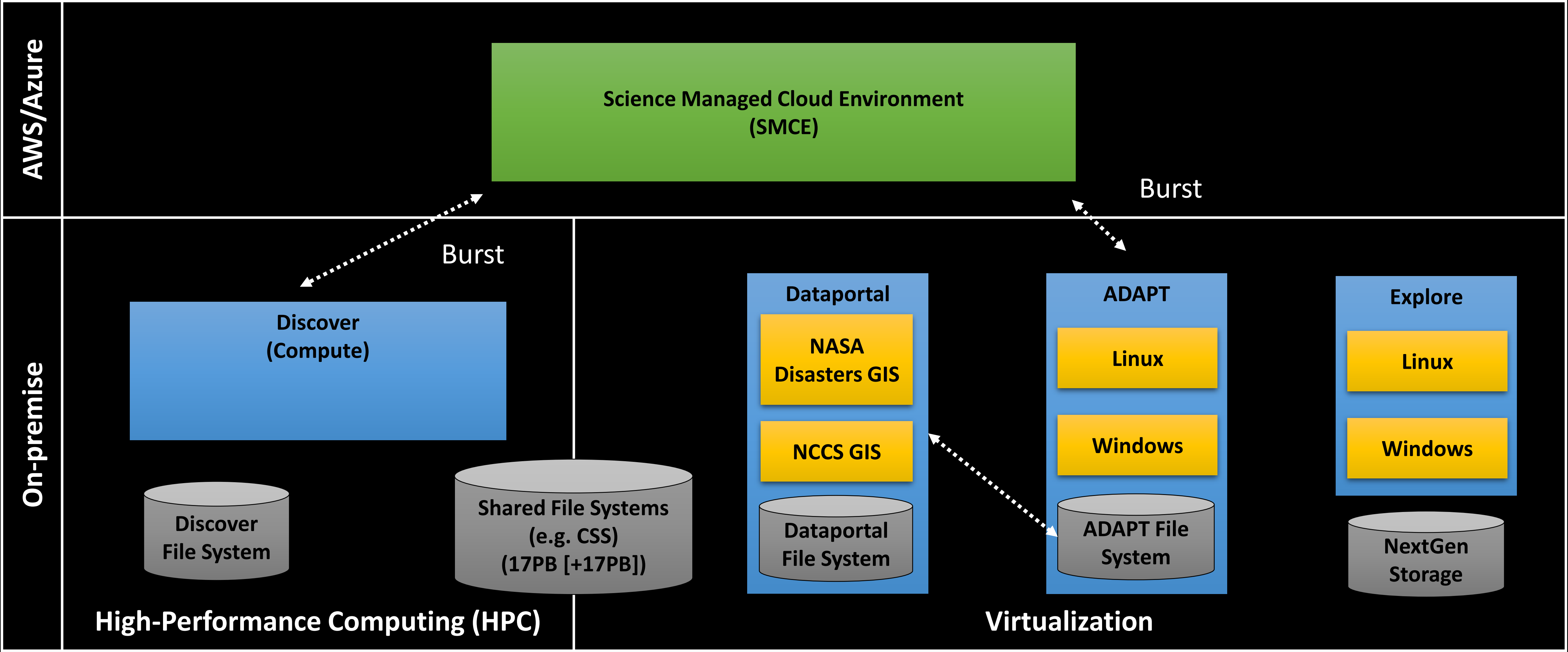
The NCCS provides high performance computing for NASA-sponsored scientists and engineers. Our integrated set of computational capabilities includes High Performance Computing, Cloud Computing, Analytics, Data Sharing and Tools, Visualization, and Climate Data Services. The purpose of the NCCS is to enhance NASA capabilities in Earth science, with an emphasis on weather and climate prediction, and to enable future scientific discoveries that will benefit humankind.

Building 28, Goddard Space Flight Center

Piers J. Sellers Data Visualization Theater

ABoVE Science Team Meeting (ASTM5), 2019

NCCS Systems



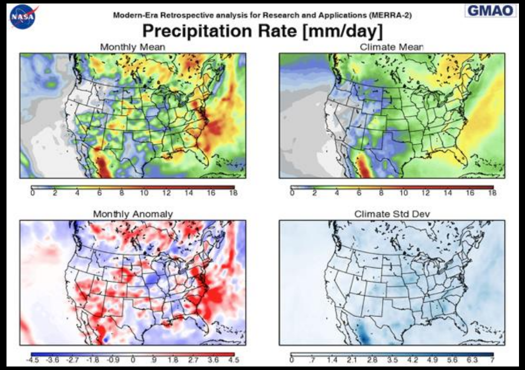
Functional Area Overview – Discover

- NCCS flagship high performance computing (HPC) cluster
- Upgraded every 1 to 2 years with scalable computing units (SCUs)
  - Recently deployed SCU16
  - SCU17 inbound (late '22)
- Platform specifications
  - 3,564 nodes
  - 127,232 CPU cores (5.57 PFLOPs)
  - 573TB RAM
  - 48PB GPFS file storage

Discover SCU 14

Discover SCU 16

Discover-based analysis



Centralized Storage System

- Centralized data store, accessible by all NCCS platforms/subsystems.
- CSS 1 – 15PB usable storage
- CSS 2 – 15PB usable storage
- CSS 3 inbound (Q2 ‘22)

CSS 1 and 2 (30PB total usable storage)

- ABoVE: 80 TB
  - Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) datasets, including over 130 archived as part of ABoVE.
  - National Snow and Ice Data Center (NSIDC) DAAC datasets, including LVIS products archived in support of ABoVE.
  - Alaska Satellite Facility (ASF) datasets related to L-band SAR.
- AMSR-2: 5 TB
- AVHRR/Polar: 40 TB on ADAPT and 10 GB on CSS
- CFHA: 250 TB
- CMIP5: 105 TB
- CREATE-IP: 79 TB
- CSDA-Spire: 30 TB

  Note: Access permission required, please contact NCCS support

- DSCOVR: 72 TB (L1B)
- DSCOVR: 72 TB (L2_CLOUD_03)
- FLDAS: 40 TB
- GeoMIP: 14 TB
- Geostationary (GOES): Ingest starting now, planning for 1 PB
- GEOS-IT: 420 TB

  Note: Public access coming soon.

- GEOS-5 Nature Runs (g5nr): 5 PB
- HIMAT Snow Reanalysis: 5 TB
- ICEBridge: 2 TB
- ICESat: 8 TB
- ICESat-2: 161 TB
- IMERG: 15 TB
- Landsat: 186 TB
- MAIAC: 107 TB

Subset of datasets stored on CSS

ADAPT/Dataportal

- Original virtualization environment
- Three subsystems
  - Legacy ADAPT hypervisors
  - Legacy OpenStack on-premise cloud
  - Dataportal data sharing applications/websites
- Platform specifications
  - 144 hypervisors
  - 3,456 CPU cores
  - 37TB RAM
  - 6PB GPFS file storage
  - Supports approximately 350 virtual machines
  - Multiple bare metal GPU systems

Nvidia DGX
Next generation 8-GPU node

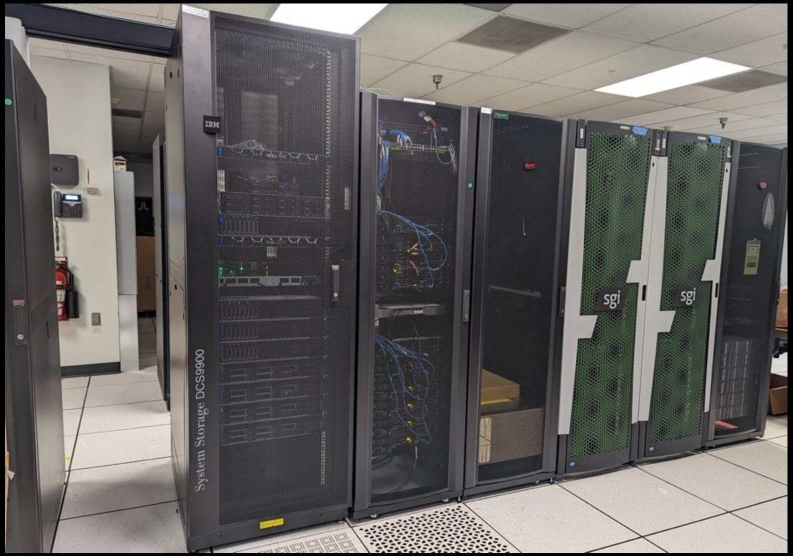
Data sharing application

PRISM
22-node GPU cluster

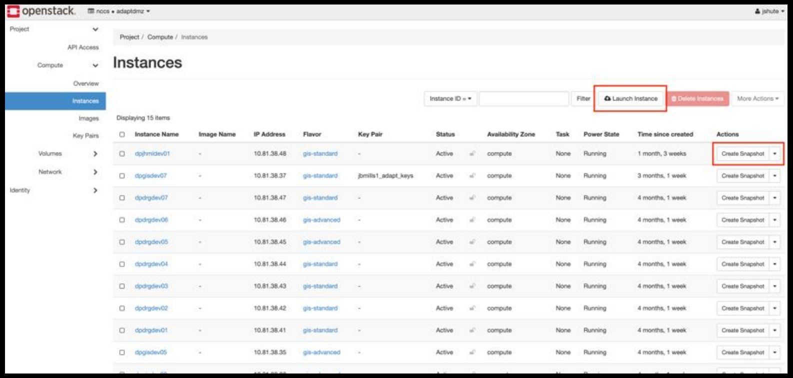
Explore

- Next generation on-premise cloud
- Multiple availability zones (B28 and B32)
- Current specifications
  - 104 hypervisors
  - 4,352 CPU cores
  - 28.6TB RAM
- Full operating capability specifications (Q4 ‘22)
  - 280 hypervisors
  - 8,704 CPU cores
  - 72TB RAM
  - 6PB next generation storage

Explore Cloud Control Plane and Initial Compute



Explore Cloud OpenStack Web Interface

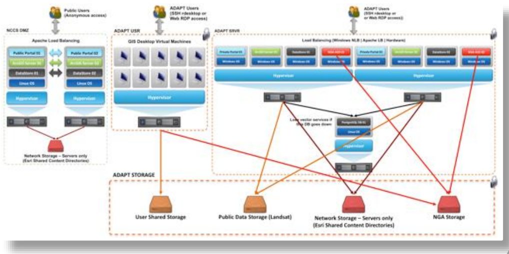


Functional Area Overview – Spatial Platforms

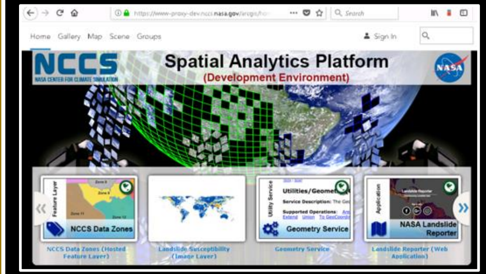
Spatial Analytics Platform (Public access)

Disaster Mapping Platform (Public access)

Common System Architecture and Tools

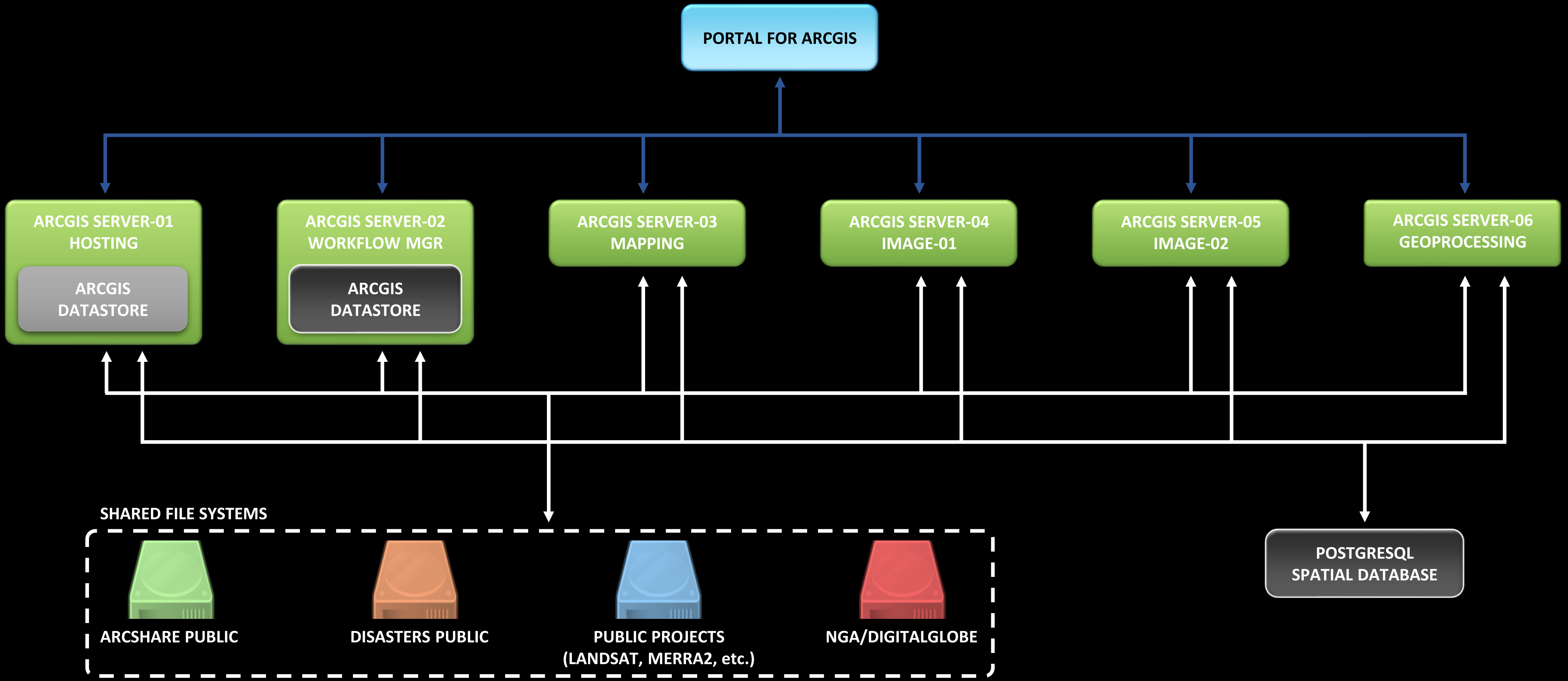


Spatial Analytics Platform DEV (NASA Network Only)

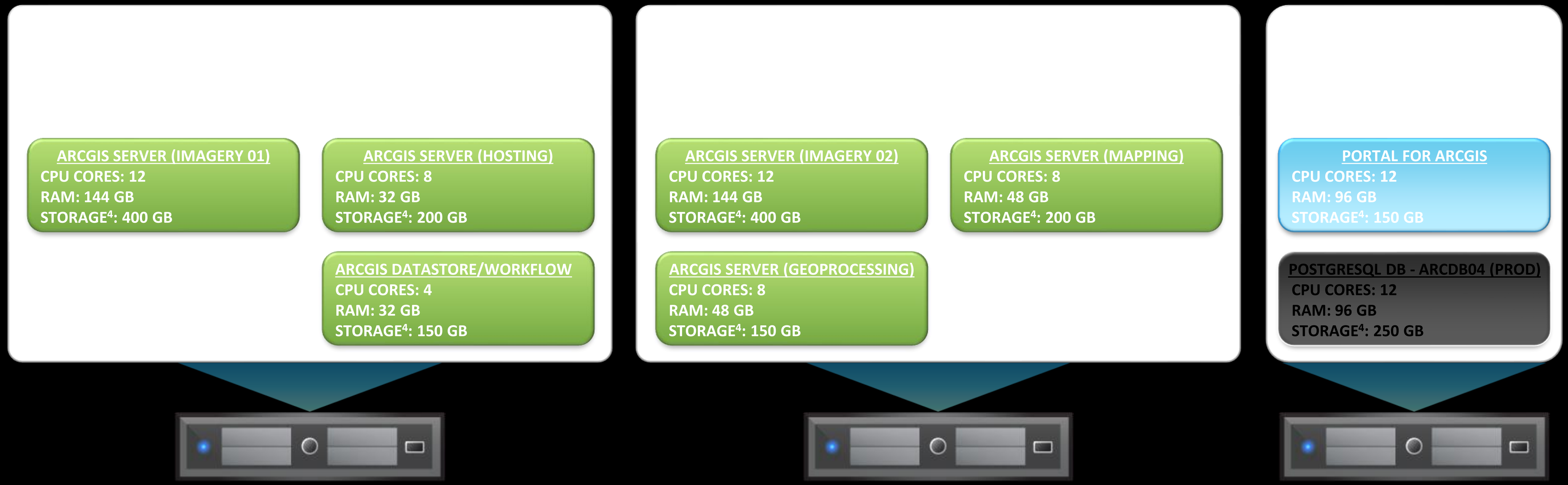


Disasters Mapping Platform DEV (NASA Network Only)

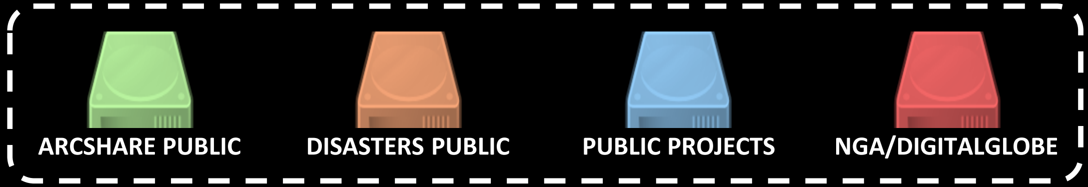
Spatial Platform Logical Architecture



Spatial Platform Physical Architecture[1]



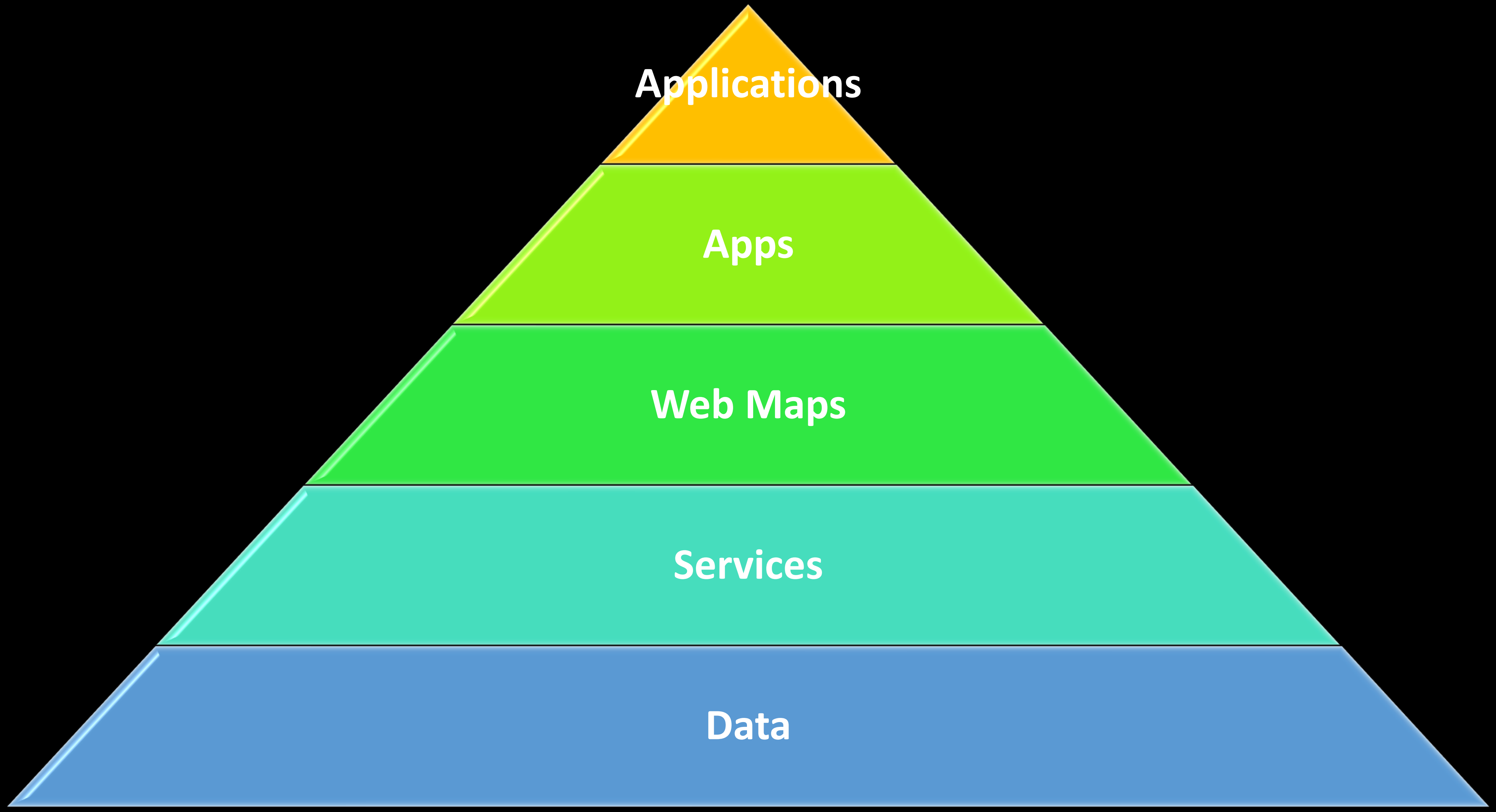
SHARED FILE SYSTEMS



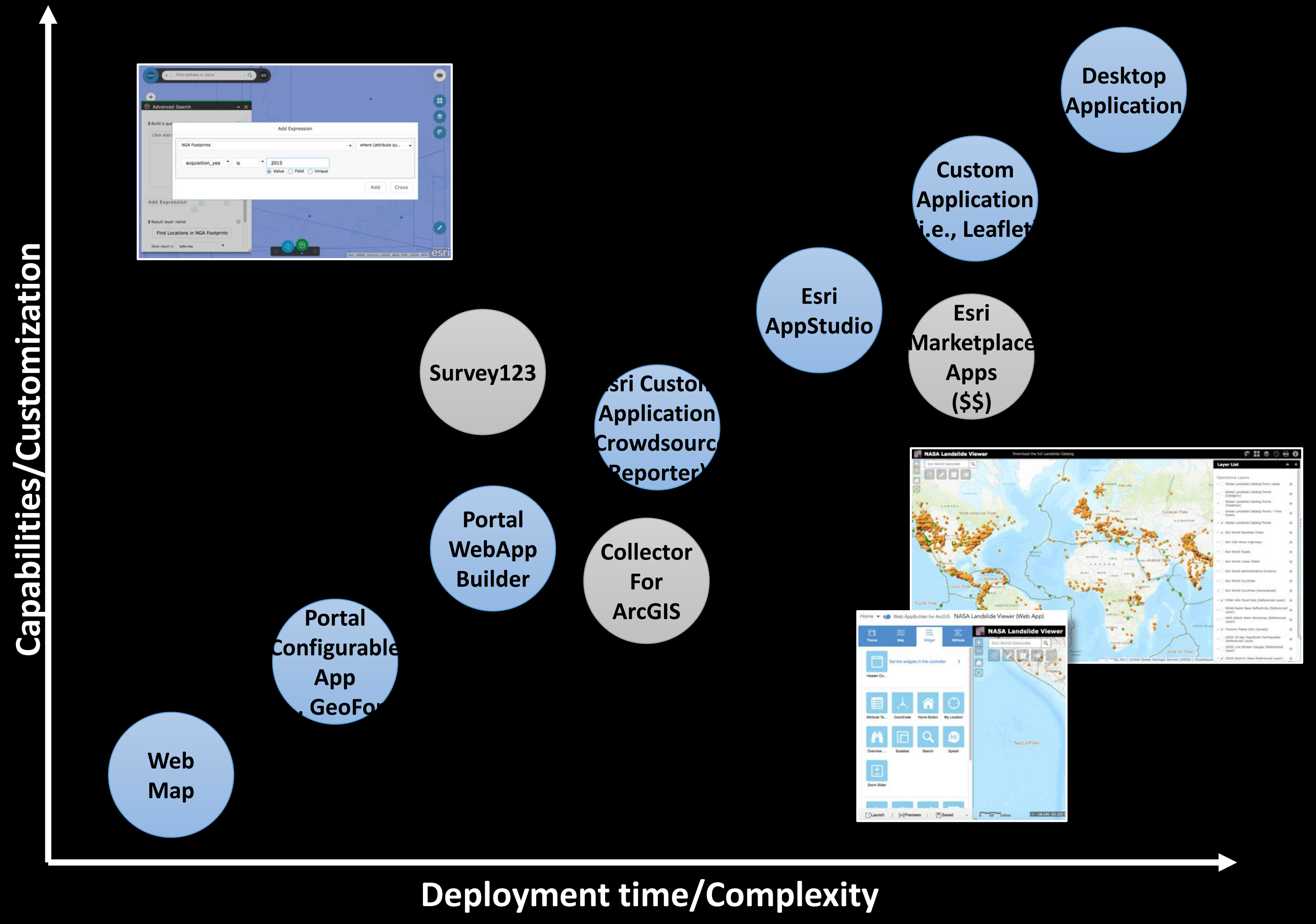
WINDOWS VIRTUAL MACHINES

NOTES:
[1] Most hypervisors have additional capacity to support expansion/growth.
[2] 28 hyperthreaded CPUs per system; 56 cores total.
[3] Hypervisors also include 2 x 600GB SSDs.
[4] Recommended values vary based on requirements.

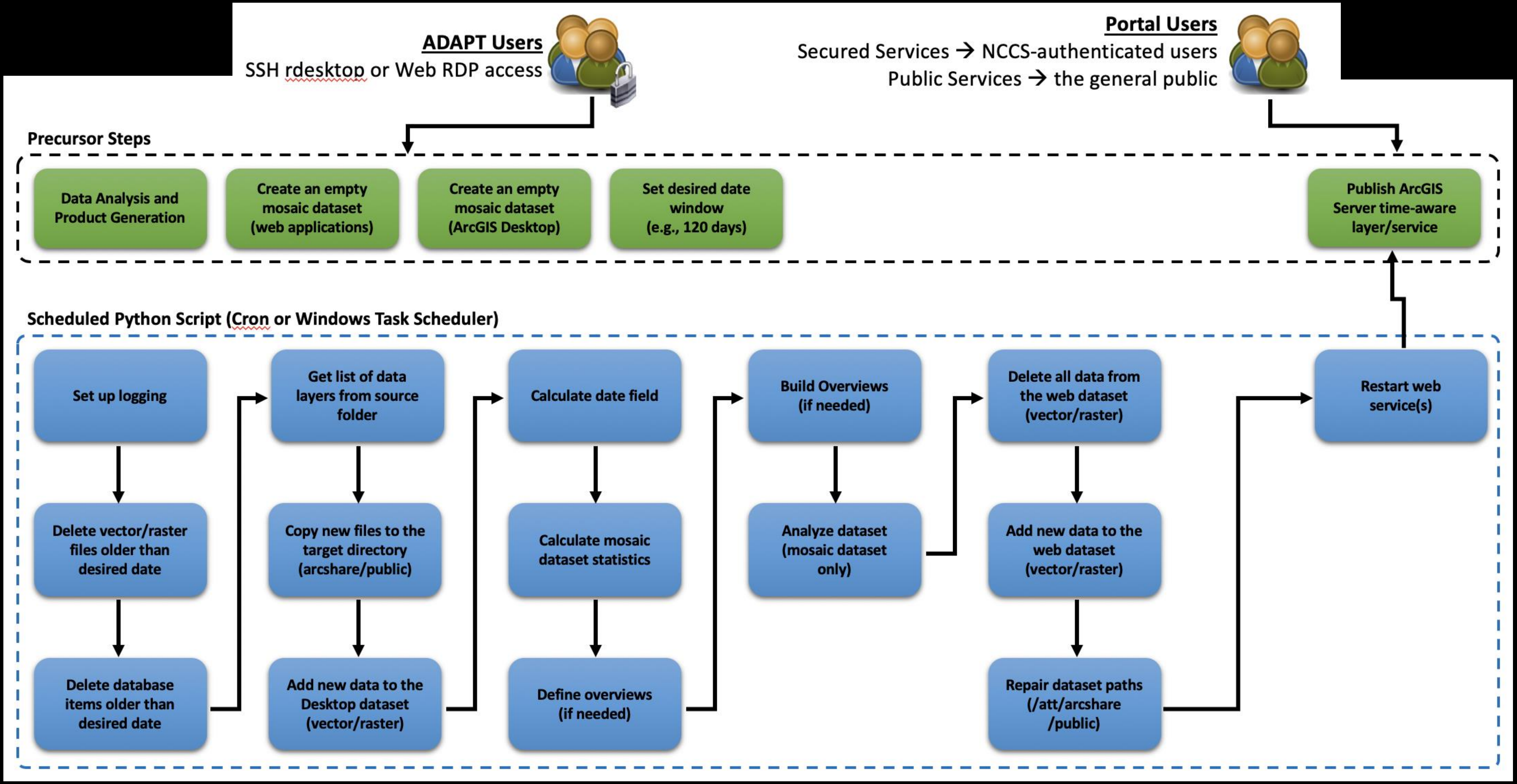
GIS Development Paradigm



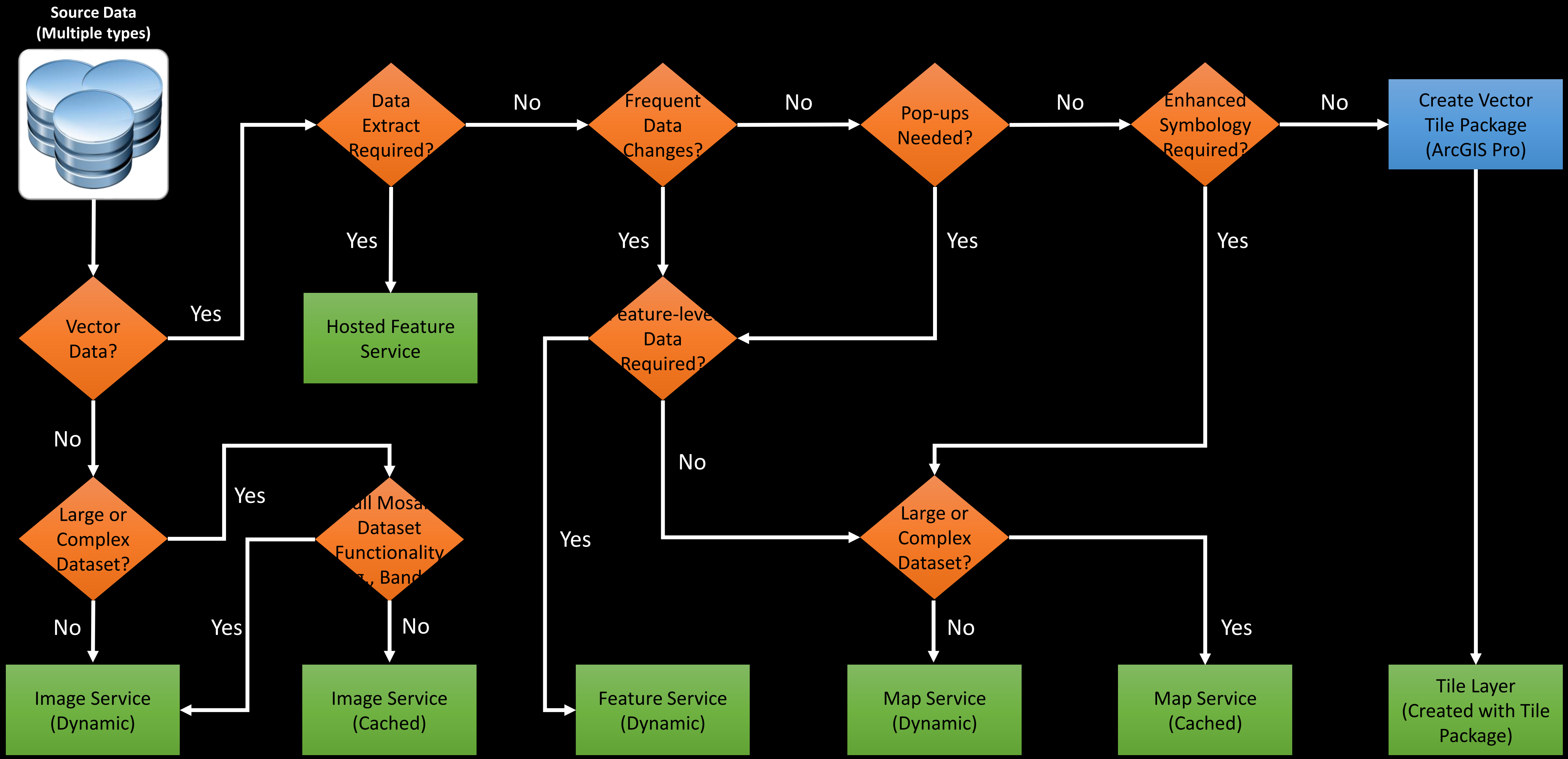
GIS Development Options



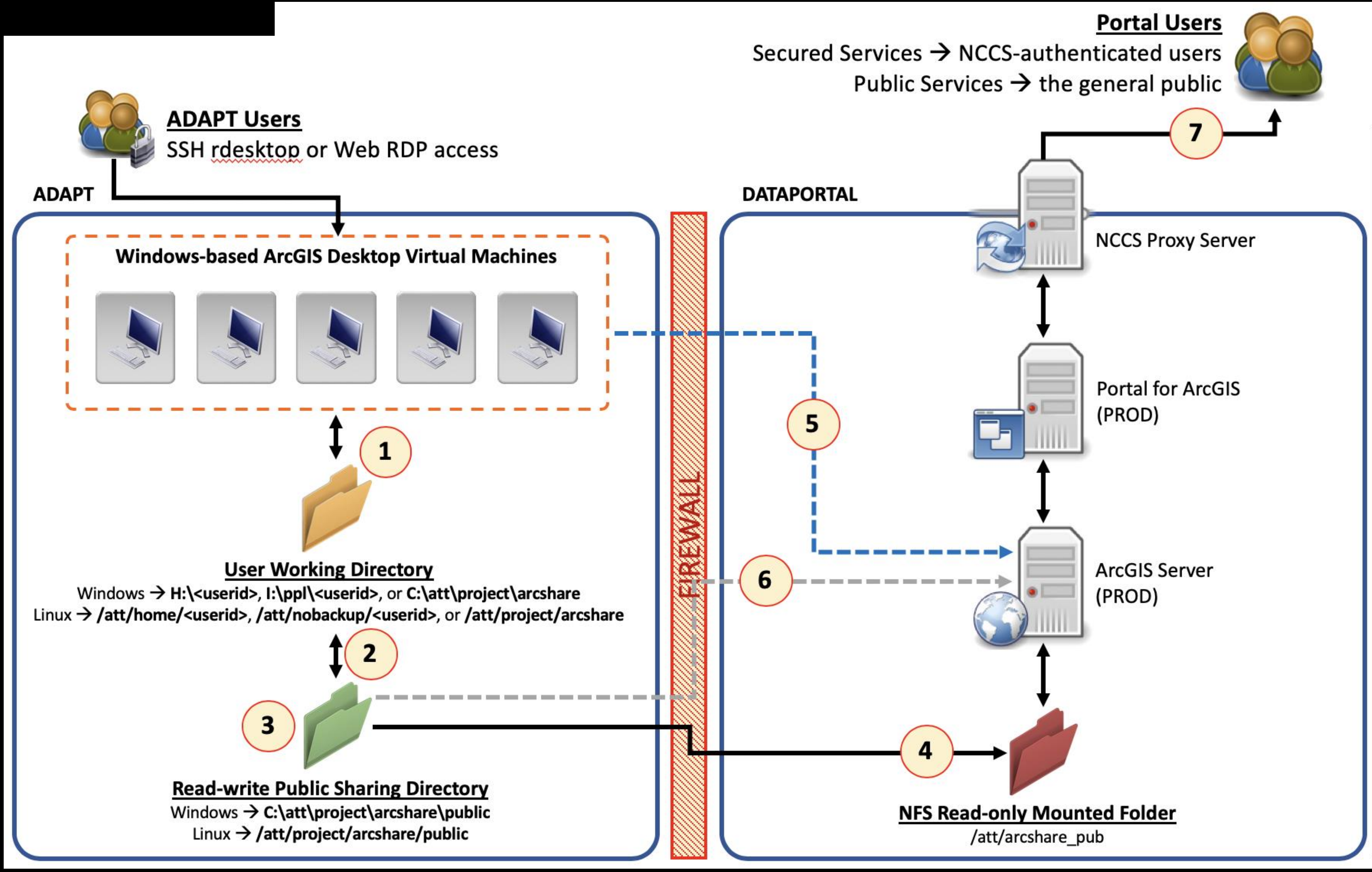
Example Workflow – Automation



Example Workflow – Service Publishing Decision Tree



Example Workflow – Service Publishing



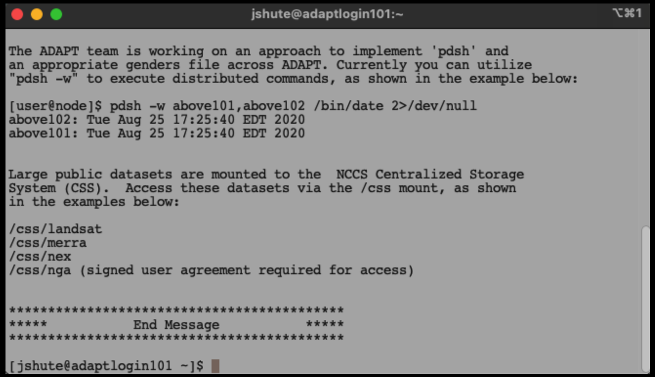
Publishing Steps

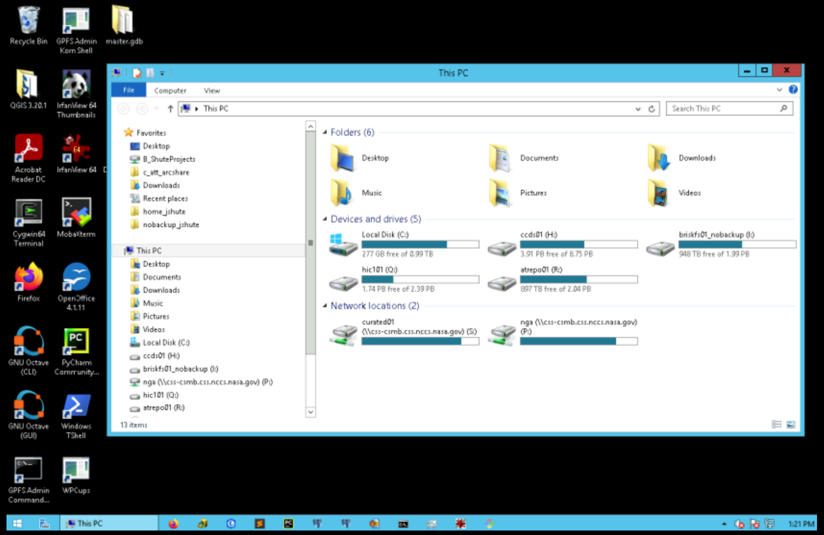
1. Using the Windows VM environment, user creates and manages data in their working directory (or shared ArcGIS folder)
2. When ready to publish, data is copied to the /arcshare/public folder
3. User sets file ownership and permissions
4. (System) NFS read-only mounted folder automatically updates with data changes
5. Using ArcGIS Desktop tools from the Windows VM environment, user publishes service(s) to the Production Portal. Note: Map documents, etc. should point to the /arcshare/public folder
6. (System) Folders are registered with ArcGIS Server and replace /arcshare/public data pointers with the NFS read-only mounted folder location
7. End user accesses services that reference the NFS read-only version of the data (/att/arcshare_pub)

Accessing the Systems

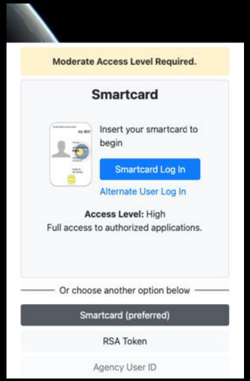
- Secure shell (ssh)
  - Majority of users
- Web services
  - Esri Portals
  - Custom applications
  - Data sharing services
- Windows desktop access
  - Guacamole
    (open source, Citrix equivalent)

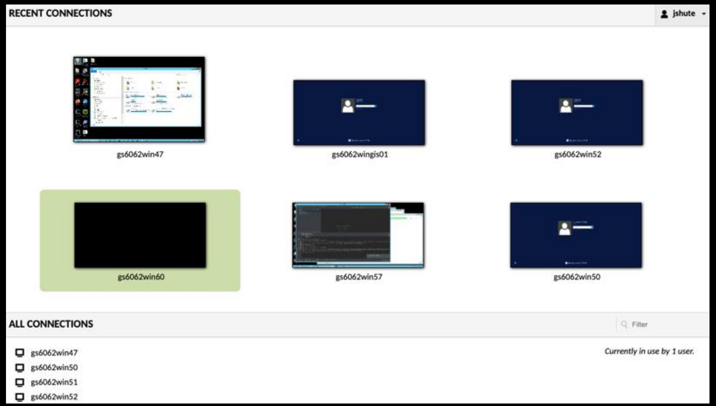
2FA-enabled secure shell (ssh) connection





2FA-enabled Windows access via Guacamole





Strengths

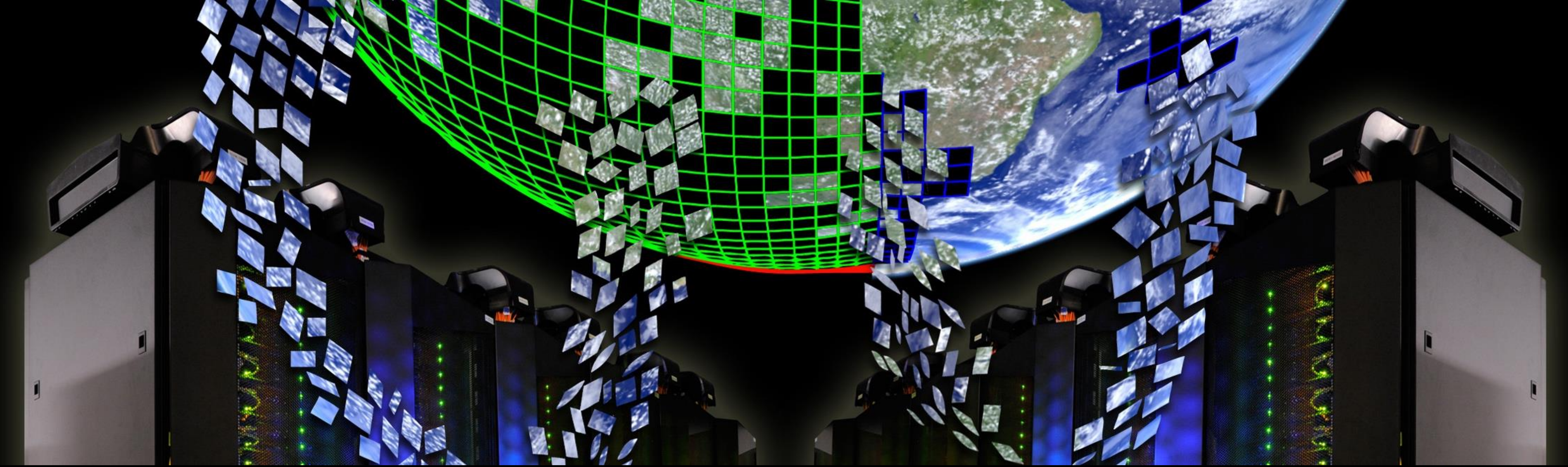
- Abundant compute resources
- Abundant storage resources
- NASA-wide Enterprise License Agreement (ELA)
  - Latest versions of the Esri software (10.9.1)
  - Enterprise software (ArcGIS Server and Portal)
  - Desktop software (ArcGIS Pro and ArcGIS Desktop)
- Proximity of data archives to GIS analysis and visualization capabilities
- Bridging the gap between Linux and Windows
- Full life-cycle support
  - Analysis (Linux) > Publication (Windows) > Visualization (Web)

Esri License Breakdown

Row Labels	Sum of License Count
3D Analyst	30
ArcGIS for Desktop Advanced	30
ArcGIS Image Server	8
ArcGIS Pro for Desktop Advanced	30
ArcGIS Server Enterprise Advanced	16
ArcGIS Server Network Analyst Extension Standard	4
ArcGIS Workflow Manager Extension Advanced	4
Geostatistical Analyst	30
Network Analyst	30
Portal for ArcGIS (qty is # of seats)	250
Spatial Analyst	30
Tracking Analyst	5
Workflow Manager	5
Grand Total	**472**

Challenges

- Most users do not consider their work to be GIS/geospatial
- Low adoption of:
  - Esri Portal application capabilities
  - Esri ArcGIS Server web service capabilities
- Filesystem stability
  - Will be mitigated with arrival of next generation storage platforms
- Finding time to automate critical workflows, due to context switching, etc.

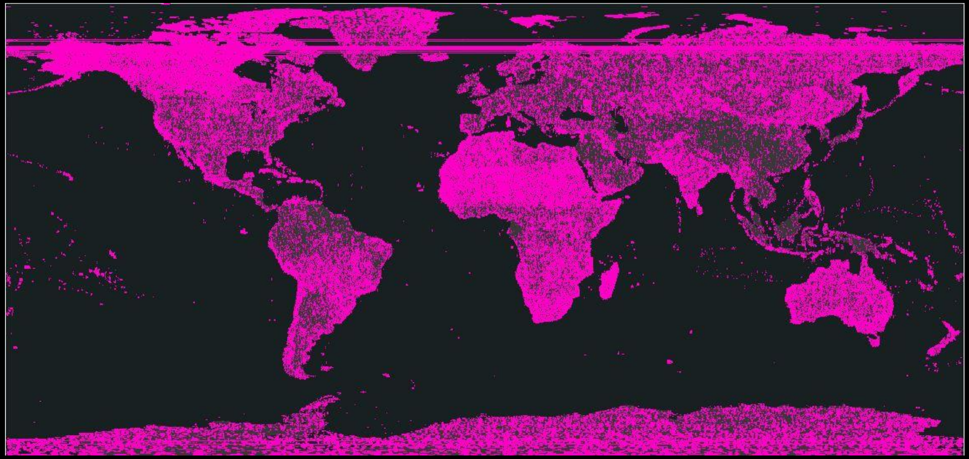


Questions?

Contact Information	
NCCS Cloud Computing and Data Services (CCDS) Lead	Jim Shute (james.k.shute@nasa.gov)
NCCS GIS Manager	Ryan Forbes (ryan.s.forbes@nasa.gov)
NCCS Website	https://www.nccs.nasa.gov
NCCS Spatial Analytics Platform	https://maps.nccs.nasa.gov
NASA Disasters GIS Platform	https://maps.disasters.nasa.gov

How can ABoVE use NCCS Systems?

- ABoVE Science Cloud is part of ADAPT/Explore
- ADAPT has both Linux & Windows Virtual Machines (VMs)
- Incorporates storage, compute, and cloud computing capabilities
- Designed for large-scale data analytics



>12 PB of Maxar Imagery Available on ADAPT

Info on ADAPT:
https://www.nccs.nasa.gov/systems/ADAPT

To access ADAPT & Maxar Imagery: https://above.nasa.gov/sciencecloud.html

Data Collections on Explore/ADAPT

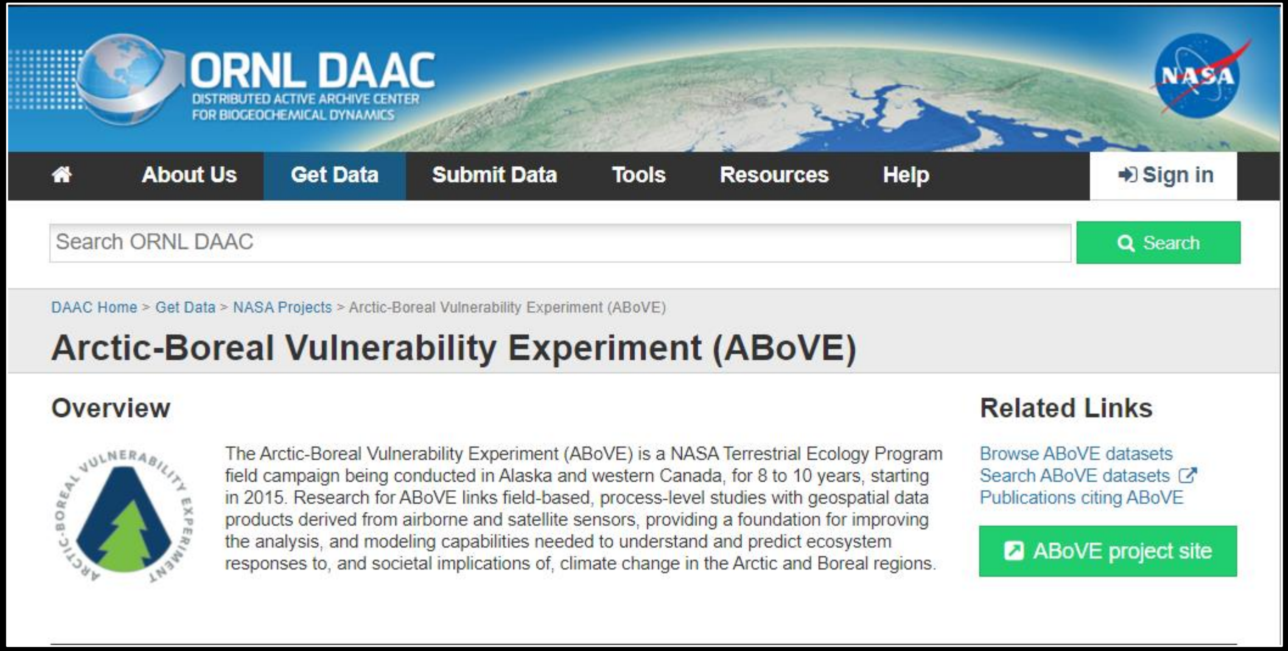
ABoVE Science Cloud

- ABoVE: 80 TB
  - Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) datasets, including over 130 archived as part of ABoVE.
  - National Snow and Ice Data Center (NSIDC) DAAC datasets, including LVIS products archived in support of ABoVE.
  - Alaska Satellite Facility (ASF) datasets related to L-band SAR.
- AMSR-2: 5 TB
- AVHRR/Polar: 40 TB on ADAPT and 10 GB on CSS
- CFHA: 250 TB
- CMIP5: 105 TB
- CREATE-IP: 79 TB
- CSDA-Spire: 30 TB
- DSCOVR: 72 TB (EPIC O3SO2AI)
- DSCOVR: 72 TB (L1B)
- DSCOVR: 72 TB (L2_CLOUD_03)
- FLDAS: 40 TB
- GeoMIP: 14 TB
- Geostationary (GOES): Ingest starting now, planning for 1 PB
- GEOS-5 Nature Runs (g5nr): 5 PB

- HIMAT Snow Reanalysis: 5 TB
- ICEBridge: 2 TB
- ICESat: 8 TB
- ICESat-2: 161 TB
- IMERG: 15 TB
- Landsat: 186 TB
- MAIAC: 107 TB
- MERRA: 86 TB
- MERRA2: 320 TB
- Selected MODIS data: 679 TB
- NEX GDDP: 11 TB
- NEX DCP30: 11 TB
- NGA: 12000 TB (~ 12 PB)

80 TB of ABoVE-Archived Data are on ADAPT

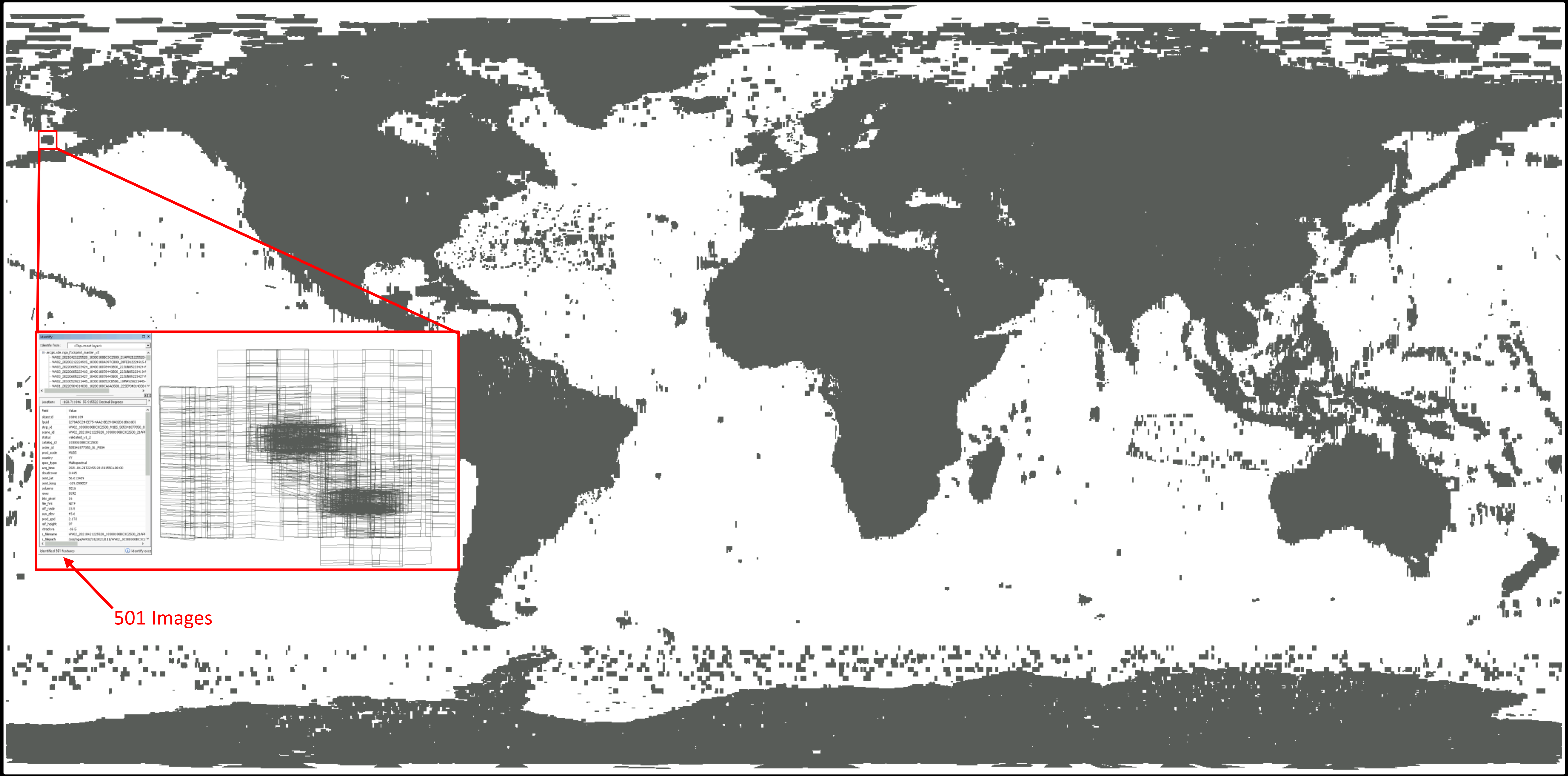
ADAPT brings together all ABoVE datasets in a single location for users.



NGA/Maxar Data Statistics

- 29.2 million images
- 12.8 petabytes
- Sensors
  - GE01 (1,948,067)
  - IK01 (322,417)
  - OV03 (227)
  - QB02 (3,470,713)
  - WV01 (6,362,592)
  - WV02 (11,796,075)
  - WV03 (5,234,308)
  - WV04 (13,505)
- Years
  - 1999 (10), 2000 (25,485), 2001 (34,174), 2002 (176,934), 2003 (260,623), 2004 (274,672), 2005 (252,903), 2006 (309,937), 2007 (482,251), 2008 (740,694), 2009 (802,602), 2010 (1,197,106), 2011 (1,772,616), 2012 (1,990,030), 2013 (1,914,354), 2014 (1,753,856), 2015 (1,886,852), 2016 (2,150,556), 2017 (2,297,327), 2018 (1,923,486), 2019 (1,580,714), 2020 (2,733,255), 2021 (2,560,723), 2022 (2,021,130), 2023 (5,615)

NGA/Maxar Data Coverage



Analysis Ready Data from the Innovation Lab

NCCS ABoVE GLAD ARD Product

- Landsat Analysis Ready Data (ARD) tiles are produced from Landsat Collection 2 data by the Global Land Analysis and Discovery Lab (GLAD) at the University of Maryland.
- The dataset is distributed as 16-day interval composites, with 23 intervals for each year (i.e., Interval ID 5 corresponds to DOY 65 to 80).
- We have downloaded and made available the entirety of this dataset through October 2022 across the ABoVE Study domain, including the Extended Region. There is currently a total of 18TB, with more data to come.
- The GLAD Landsat ARD C2 product has been gridded to the ABoVE Reference Grid at moderate resolution (B) and reprojected to Canada Albers Equal Area Conic (ESRI: 102001).

/css/above/glad.umd.edu/Collection2/GLAD_ARD/ABoVE_Grid

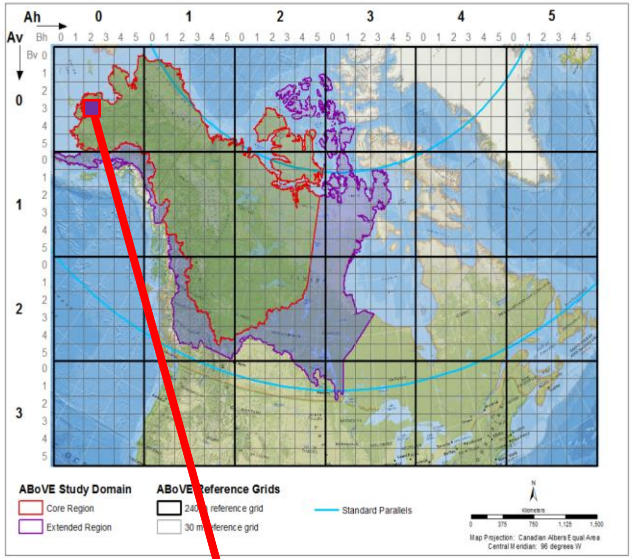


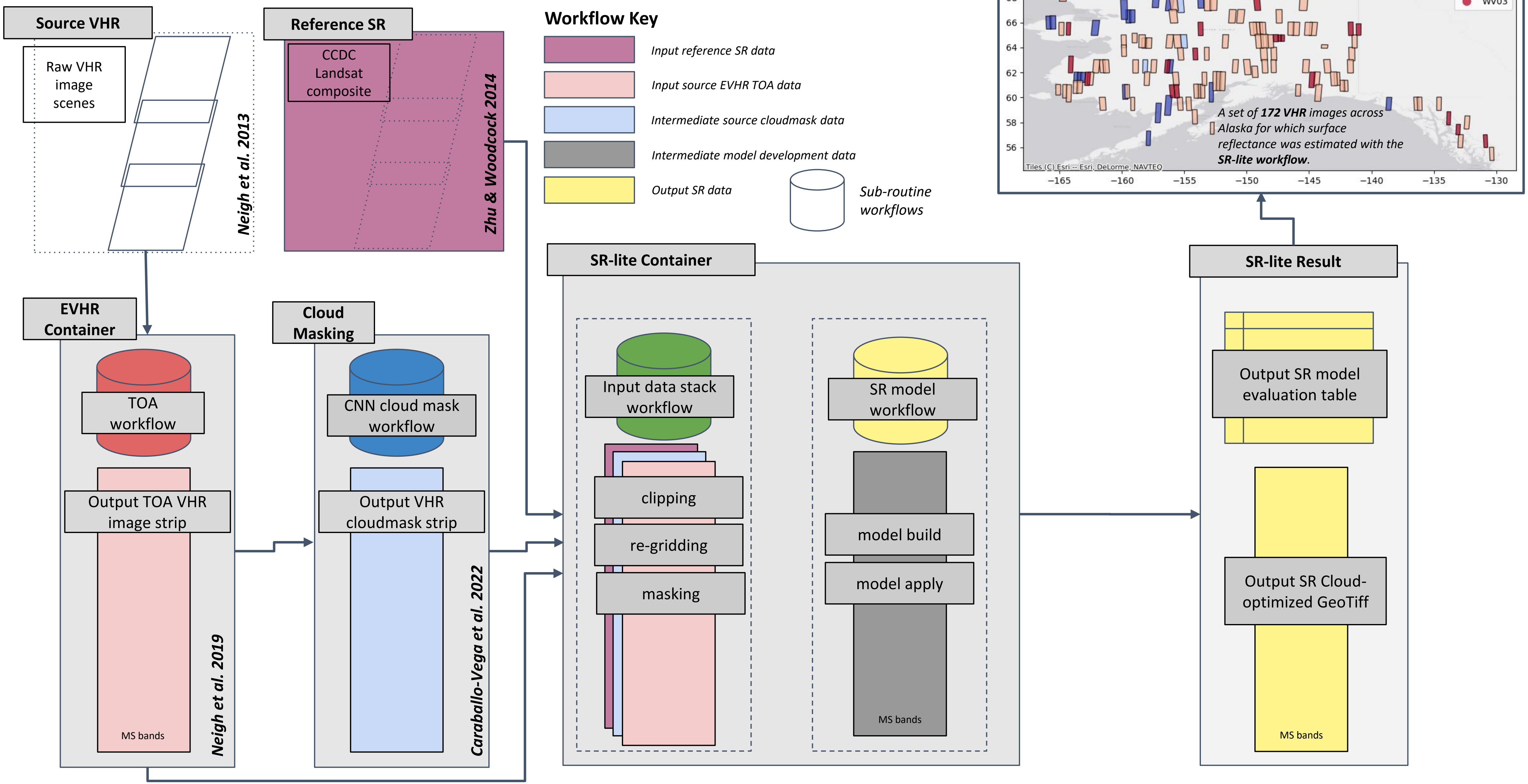
Table 1. Available bands

ID	Band
1	Blue band
2	Green band
3	Red band
4	NIR band
5	SWIR1 band
6	SWIR2 band
7	Normalized brightness temperature
8	Observation quality flag (QF)

ABoVE.GladARD.200511.Bh002v003.001.20220919.tif	
200511	data reference date, given by year and interval ID
Bh002v003	position within the ABoVE reference B grid
001	product version
20220919	production date

Potapov, P., Hansen, M.C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B., Tyukavina A., and Ying, Q., 2020. Landsat analysis ready data for global land cover and land cover change mapping. *Remote Sensing 12*, *426*; doi:10.3390/rs12030426

VHR Analysis Ready Data for Alaska



Montesano, Carrol, Neigh, Macander, Caraballo-Vega, Tamkin in prep. 2023

ABoVE Publications and Products

- 13 publications list the use of ADAPT in their acknowledgements (more are in-process)
- 39 data products using/used EXPLORE in their development (more to come)
  - 30 archived datasets at a NASA DAAC
  - 7 planned products
  - 2 public products

ELSEVIER

ISPRS Journal of Photogrammetry and Remote Sensing

Volume 159, January 2020, Pages 63-77

A systematic evaluation of influence of image selection process on remote sensing-based burn severity indices in North American boreal forest and tundra ecosystems

Dong Chen, Tatiana V. Loboda, Joanne V. Hall

ENVIRONMENTAL RESEARCH LETTERS

LETTER

Time-series maps reveal widespread change in plant functional type cover across Arctic and boreal Alaska and Yukon

Matthew J Macander[1,*], Peter R Nelson[2], Timm W Nawrocki[3], Gerald V Frost[1], Kathleen M Orndahl[4], Eric C Palm[5], Aaron F Wells[6] and Scott J Goetz[4]

Global Change Biology

PRIMARY RESEARCH ARTICLE

Extensive land cover change across Arctic–Boreal Northwestern North America from disturbance and climate forcing

Jonathan A. Wang, Damien Sulla-Menashe, Curtis E. Woodcock, Oliver Sonnentag, Ralph F. Keeling, Mark A. Friedl

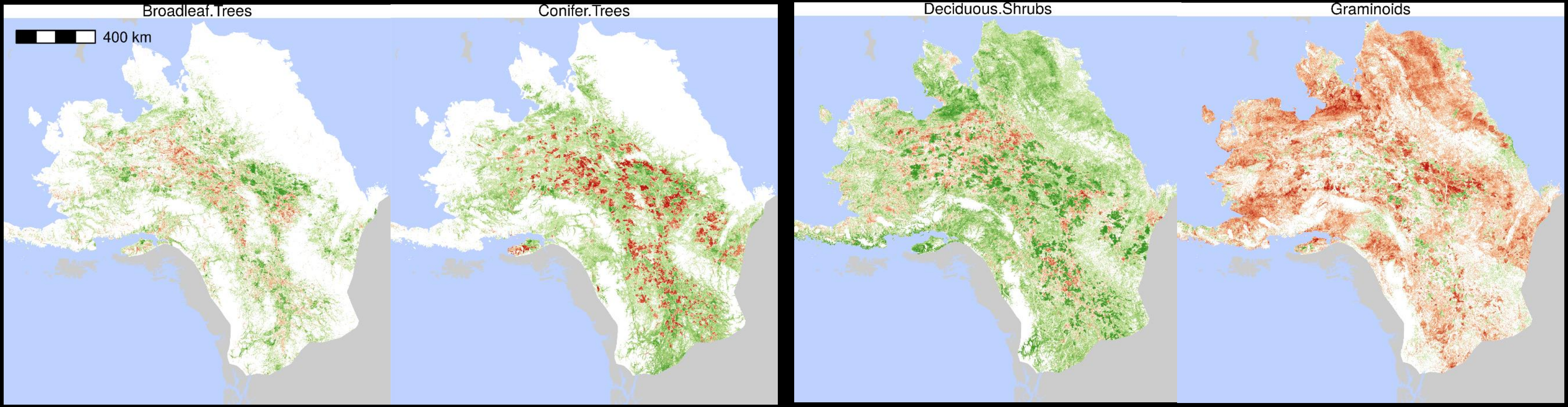
First published: 22 August 2019 | https://doi.org/10.1111/gcb.14804 | Citations: 26

ABR inc. environmental research & services

Time-series maps reveal widespread change in plant functional type cover across Arctic and boreal Alaska and Yukon

Macander et al. (2022) *Environmental Research Letters.*

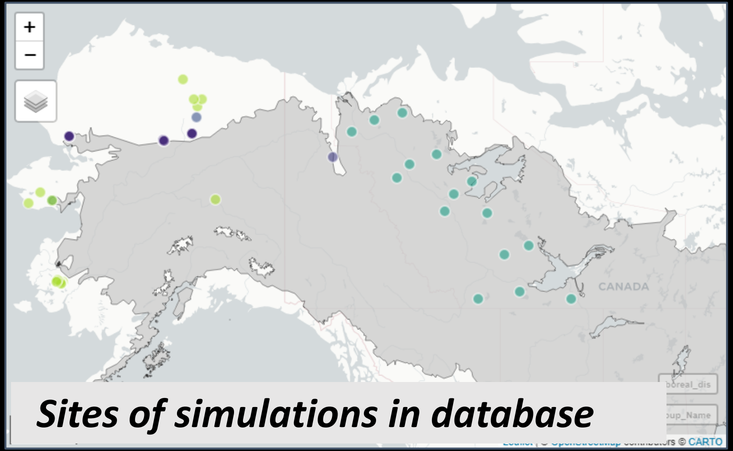
- Documented increases in Deciduous and Evergreen Shrub, Conifer and Broadleaf Tree top cover.
- Associated decreases in Graminoid and Lichen top cover with fire disturbance and shrub increases
- These changes are highly relevant to resource management applications, including wildlife habitat



A Database of Simulated Vegetation Change at Sites Across the Taiga-Tundra Ecotone

P. Montesano, B. Osmanoglu, H. Epstein, E. Heffernan, B. Gay

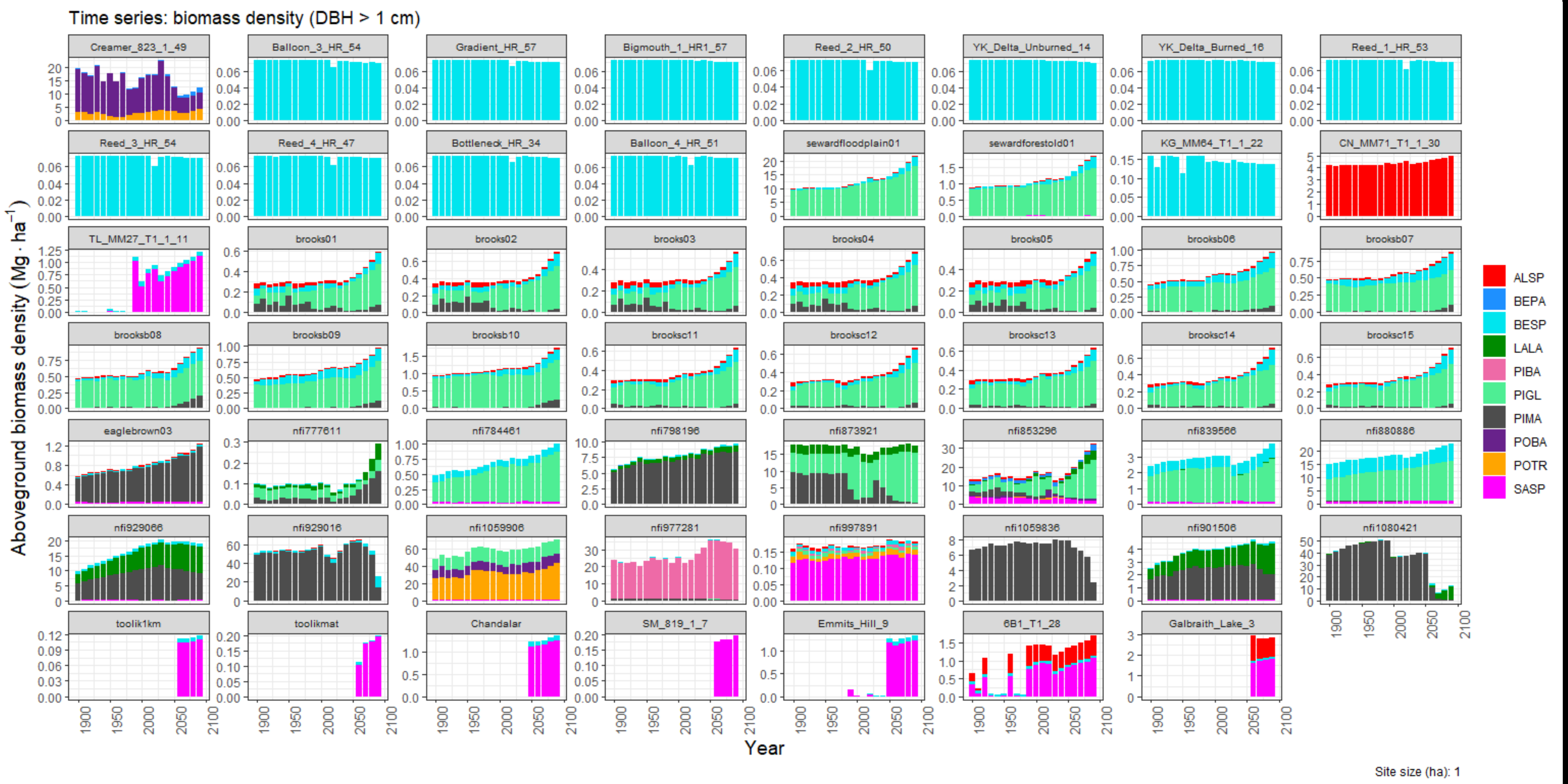
*The forest growth model **SIBBORK-TTE** is run on ADAPT to track the growth of individual trees and shrubs through time (*1900-2100) and across bioclimatic gradients using spatial inputs from ArcticDEM, SoilGRIDS, MERRA2, CMIP6, & NASA cloud cover



Sites of simulations in database

These simulations are assembled in a database to explore the variation in expected changes in vegetation structure and composition across the taiga-tundra ecotone in North America.

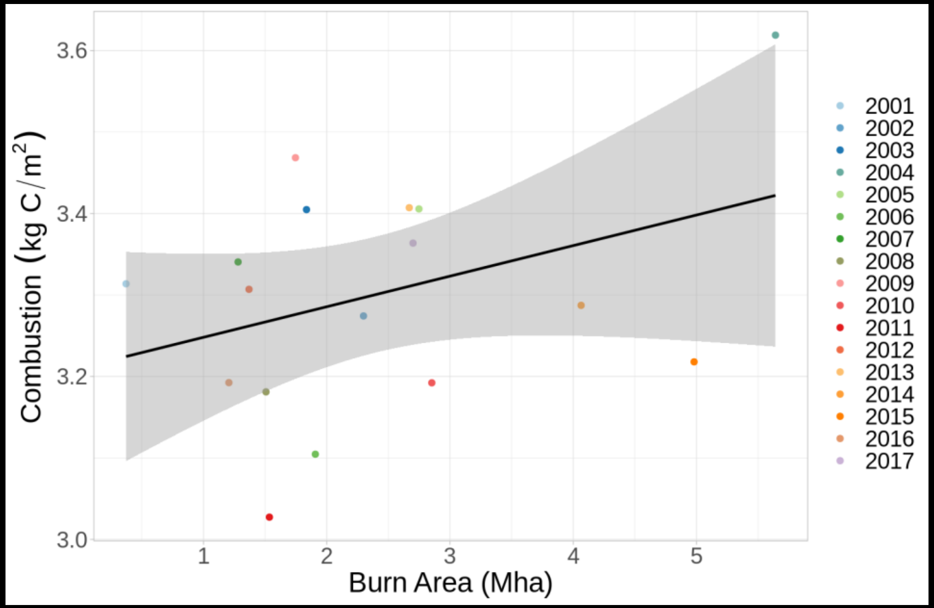
200 years of change in vegetation structure & composition at sites in the taiga-tundra ecotone

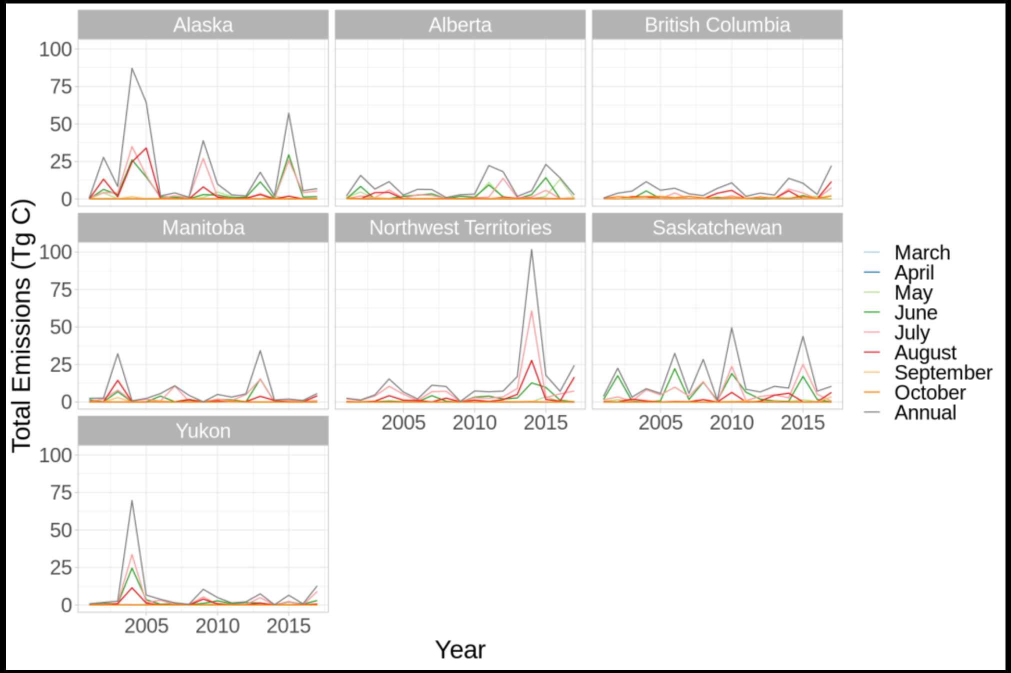


Modeling Emissions and Analyzing Variability in Burned Areas

S. Potter, B. Rogers, et al.

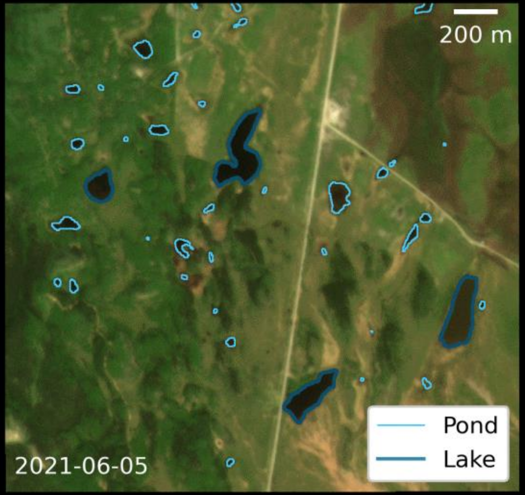
- Over 8,000 jobs submitted on ADAPT
- Using ORNL DAAC datasets to build a statistical model of emissions
- Using MODIS and Landsat data to identify burned areas & model carbon emissions
- Applying a combustion model to estimate carbon emissions from belowground and aboveground sources (for 2001-2019)

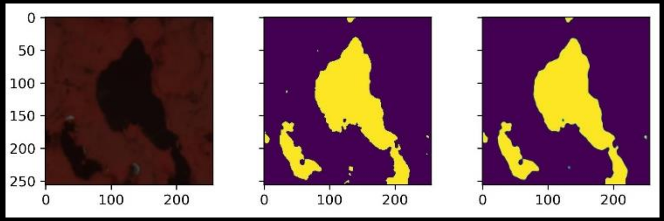




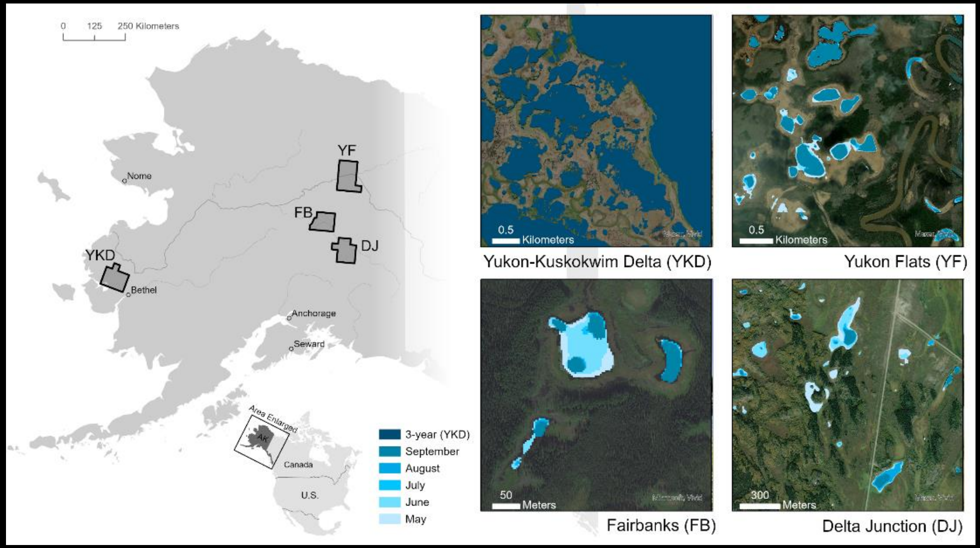
Using High-resolution Satellite Imagery and Deep Learning to Track Dynamic Seasonality in Small Water Bodies

Mullen et al., 2023 (in review)

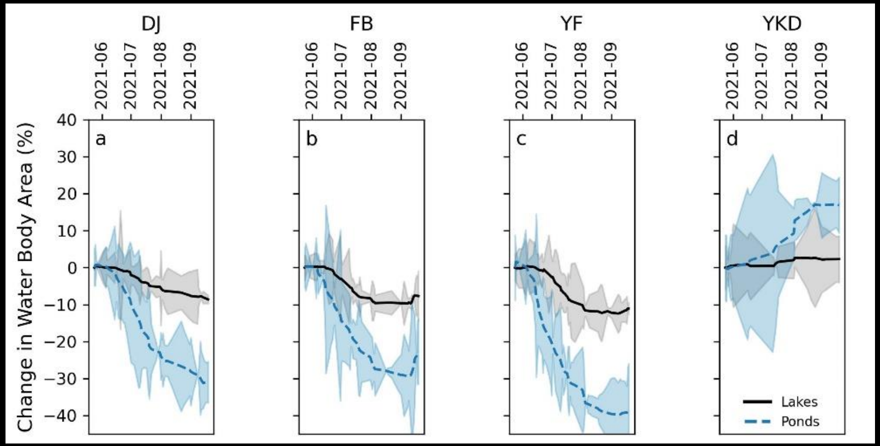




Trained model with >13,000 hand-delineated water bodies on PRISM GPU cluster. Training data development relied heavily on 2 m Maxar WorldView imagery from NGA archive.



Deployed model on over 7,000 PlanetScope images (area ~ 3x Alaska)



Tracked small changes in lake and pond surface area at near-daily temporal resolution

Acknowledging ADAPT in Publications

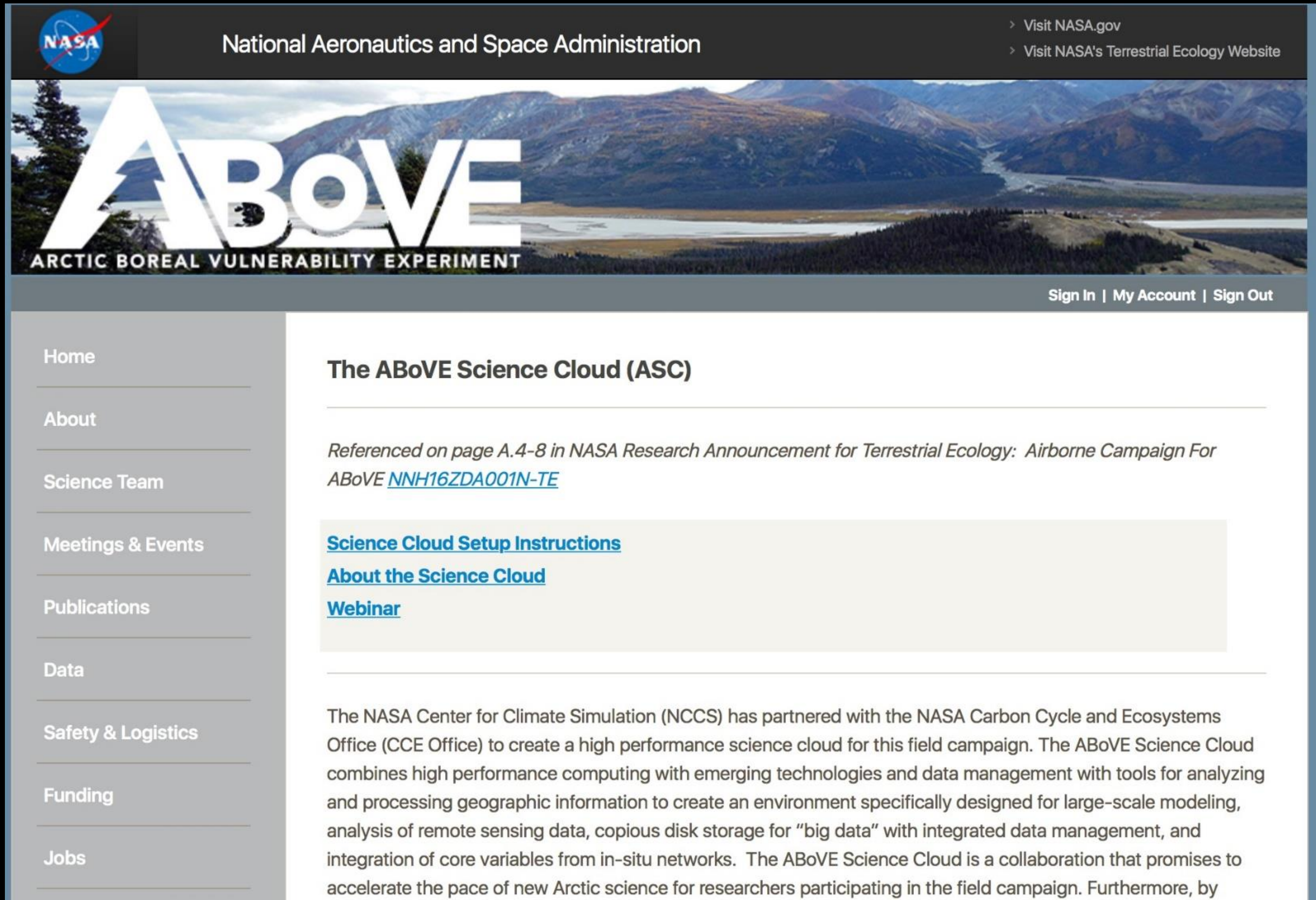
When researchers publish research based on the use of NCCS resources, please acknowledge these NASA-provided resources with language similar to this example:

“Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center.”

Logistics

Gaining Access – ABoVE Website:

- Instructions under “Data”, “ABoVE Science Cloud”, link to Science Cloud Setup Instructions
- Need NASA identity, IT Security training, RSA soft token, process takes a while
- Optional: signed NGA paperwork, new version in process
- Links to monthly webinars, other instructional videos



http://above.nasa.gov/sciencecloud.html

Logistics

NCCS Website: https://www.nccs.nasa.gov/systems/adapt

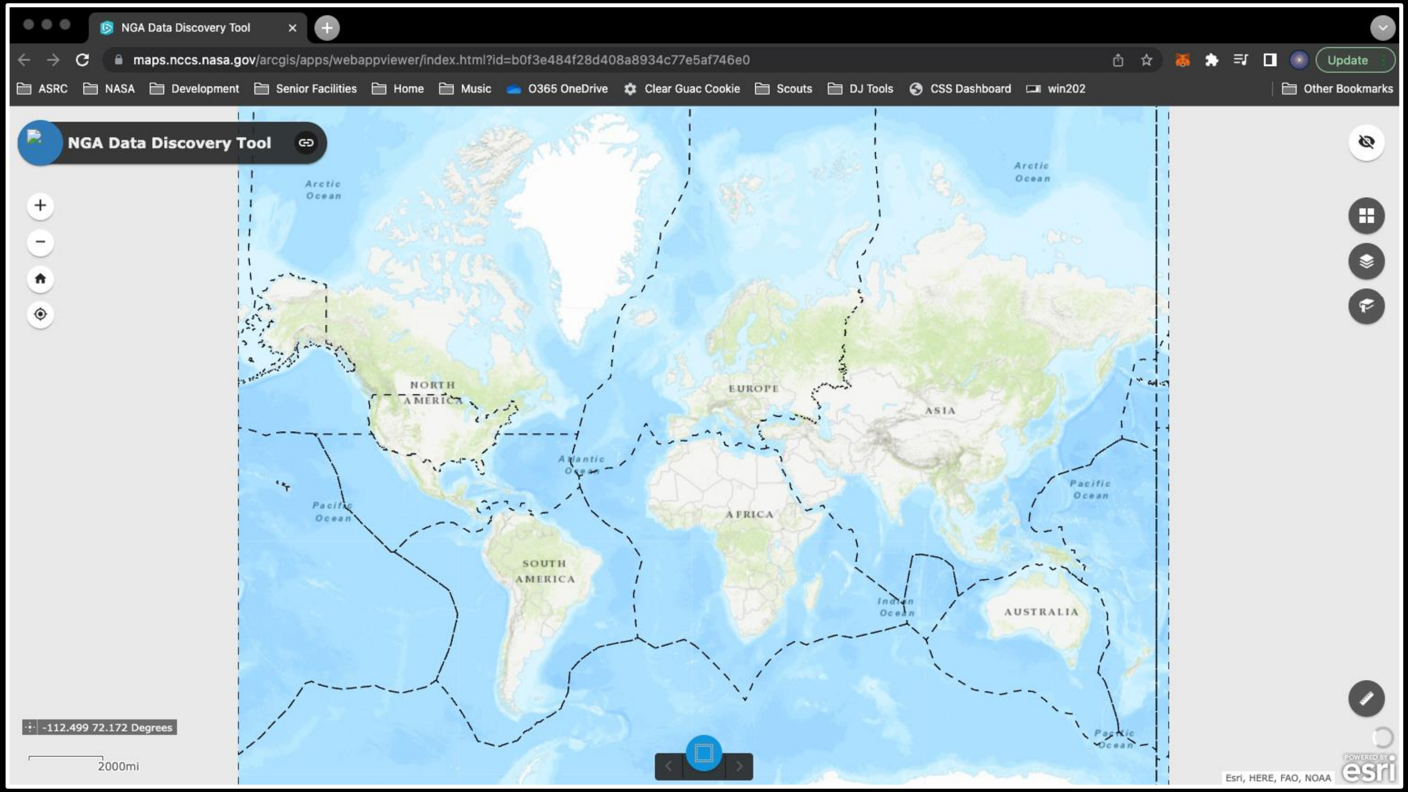
- How to login, data locations, Windows FAQ, ABoVE FAQ, including orthorectification instructions
- Instructional Video Collection: https://www.nccs.nasa.gov/nccs-users/instructional/instructional-videos

- Account Setup Questions – elizabeth.hoy@nasa.gov

- Questions/Issues Using NCCS Systems – support@nccs.nasa.gov

Demos

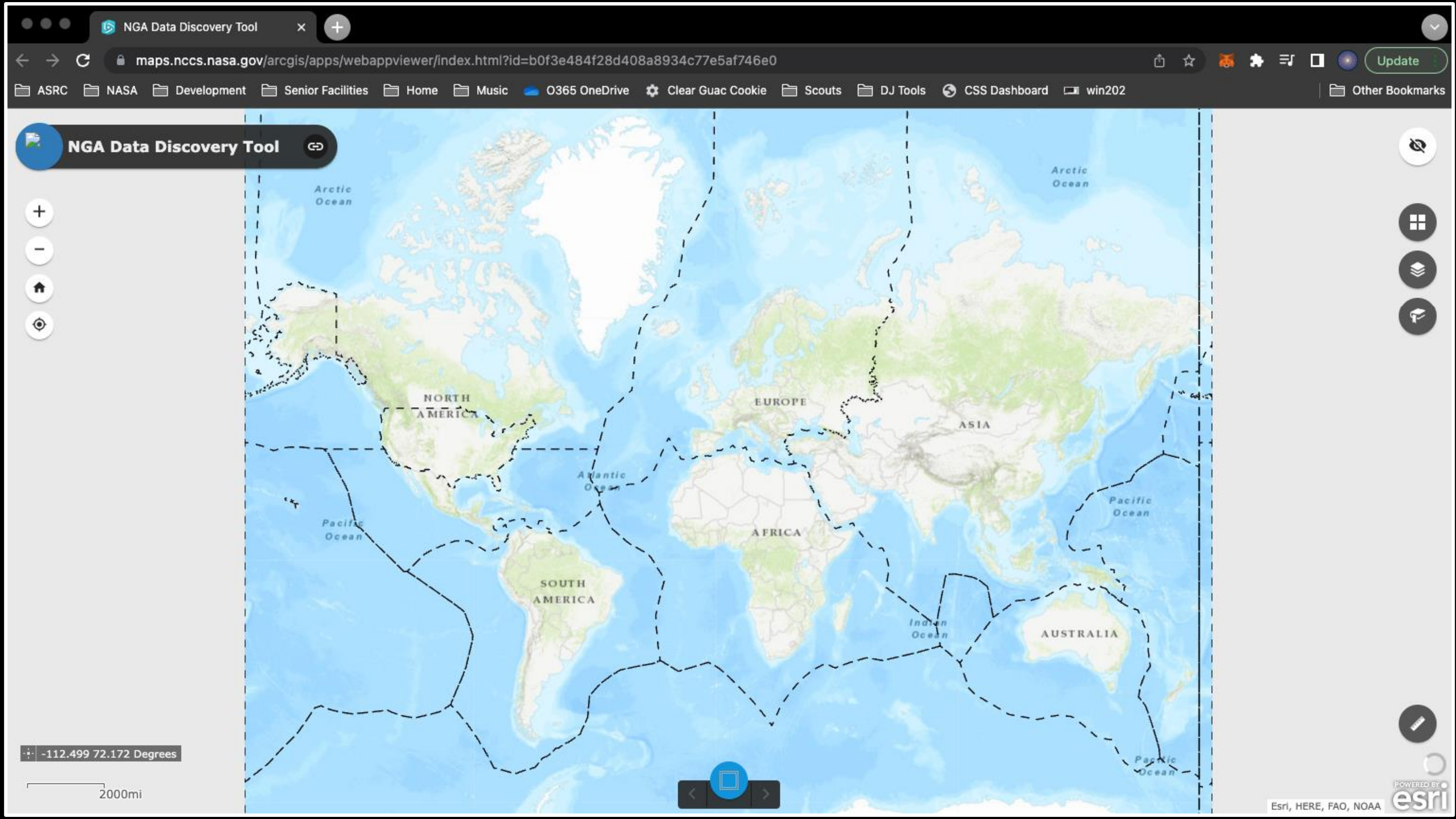
- Accessing ADAPT
  - Linux VM
  - Windows VM
- Jupyter Hub
- Data Discovery Tool



Backup Slides - Demos

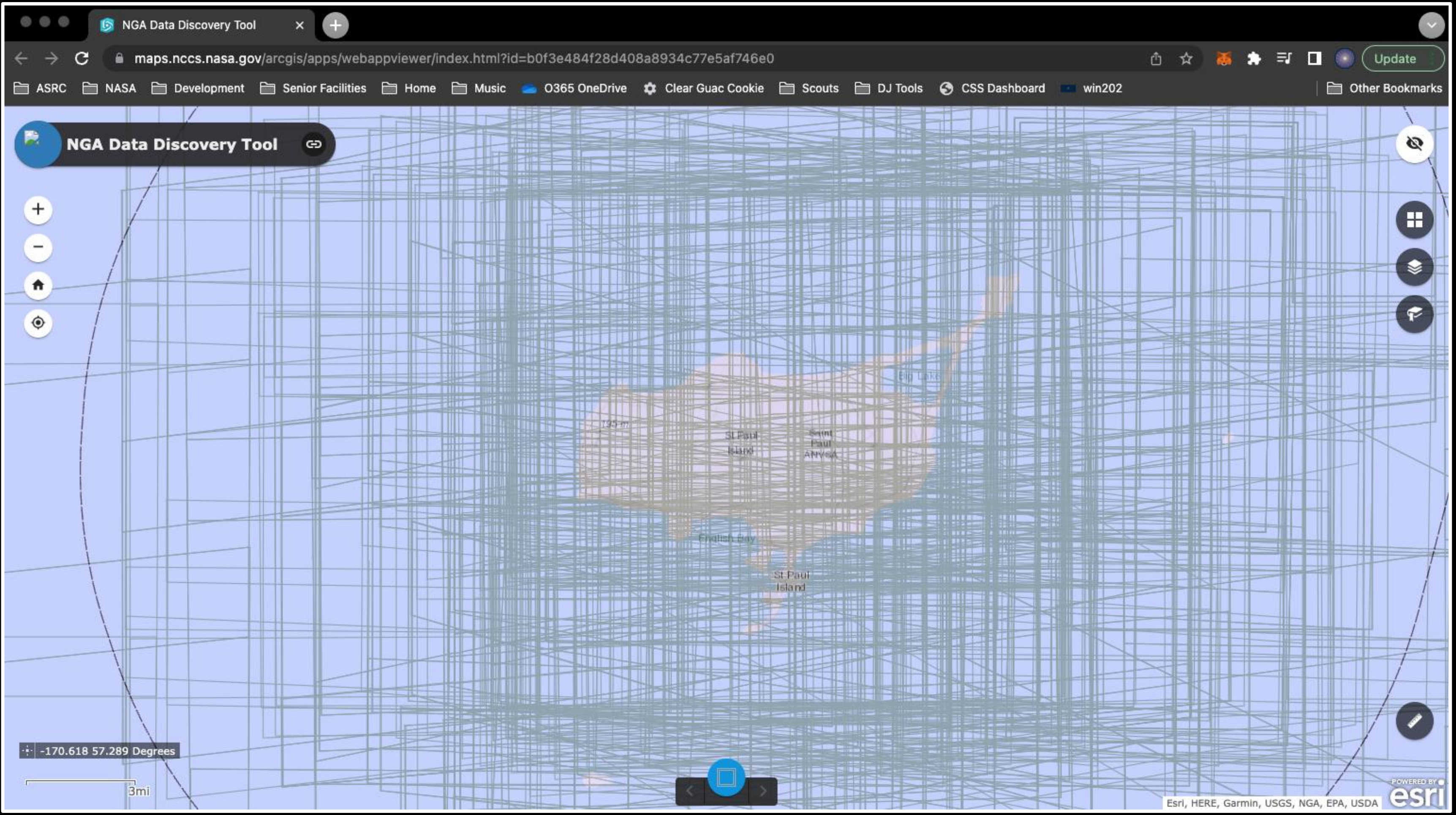
Data Discovery Tool

- Global View (polygons are hidden at this scale)



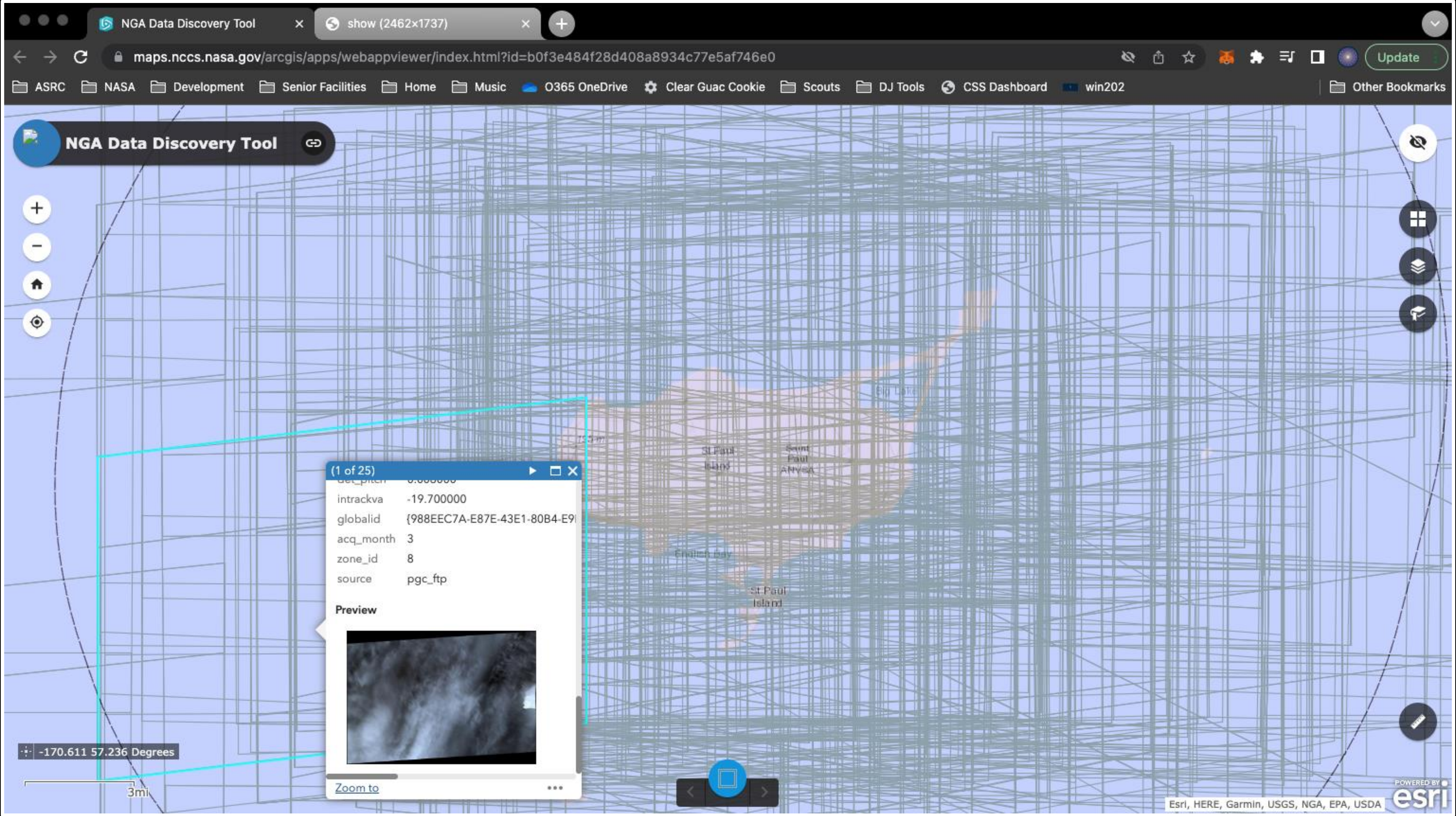
Data Discovery Tool

- Zoom to area of interest (polygons enabled; reduces server impact)



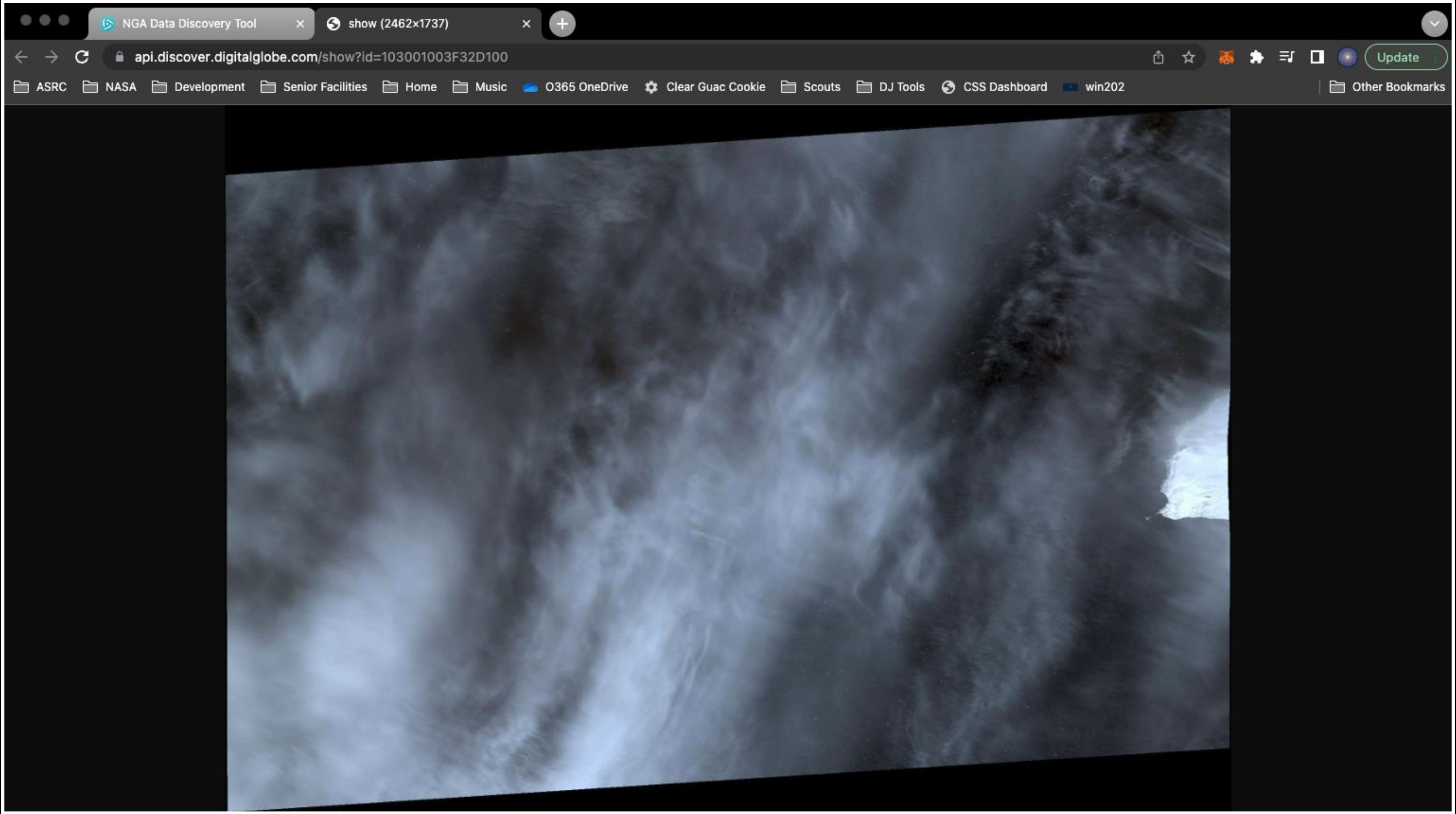
Data Discovery Tool

- Click a polygon to see the attributes, to include image thumbnail



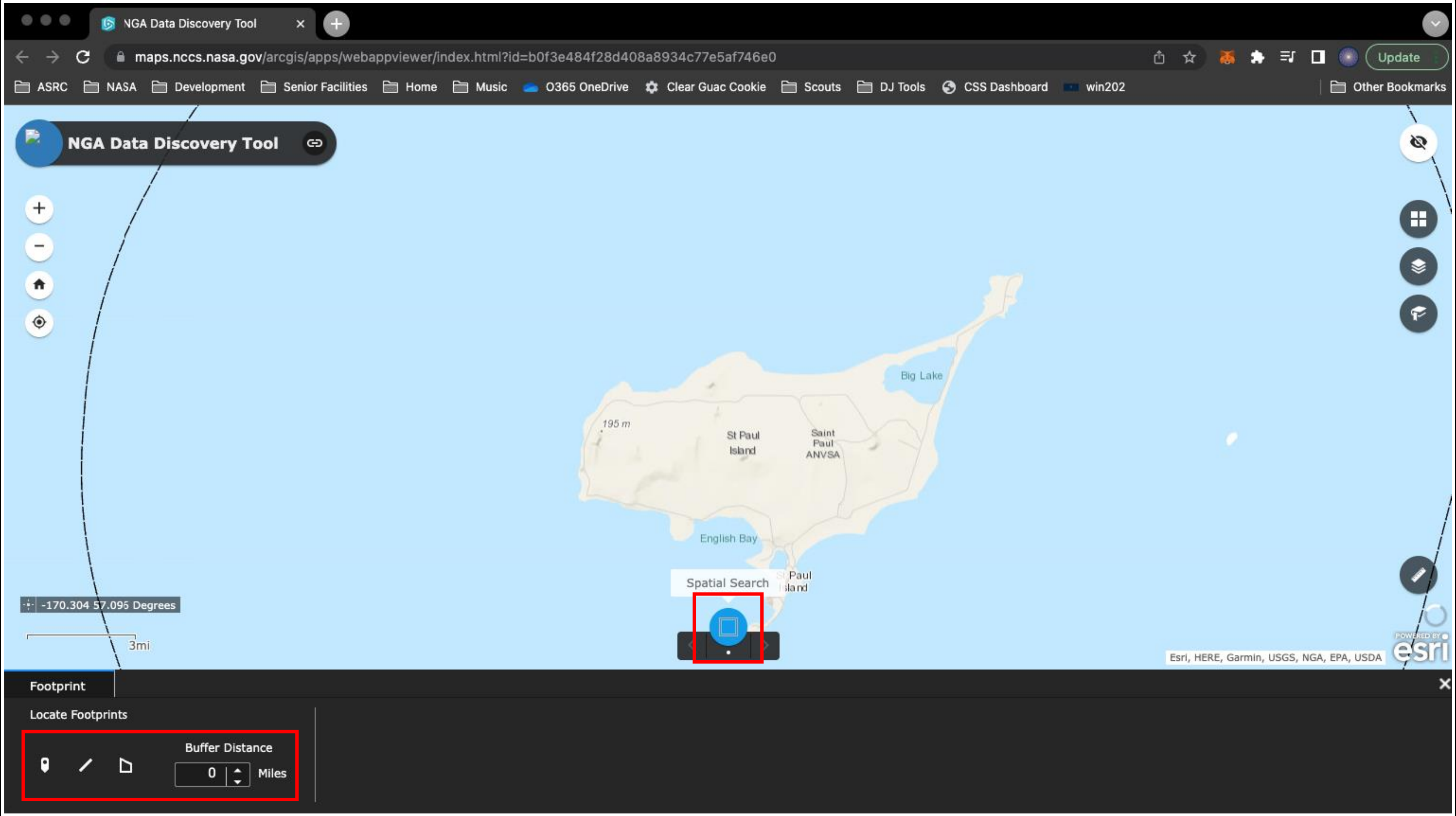
Data Discovery Tool

- Click the previewurl “More info” link or the thumbnail to see the image in more detail



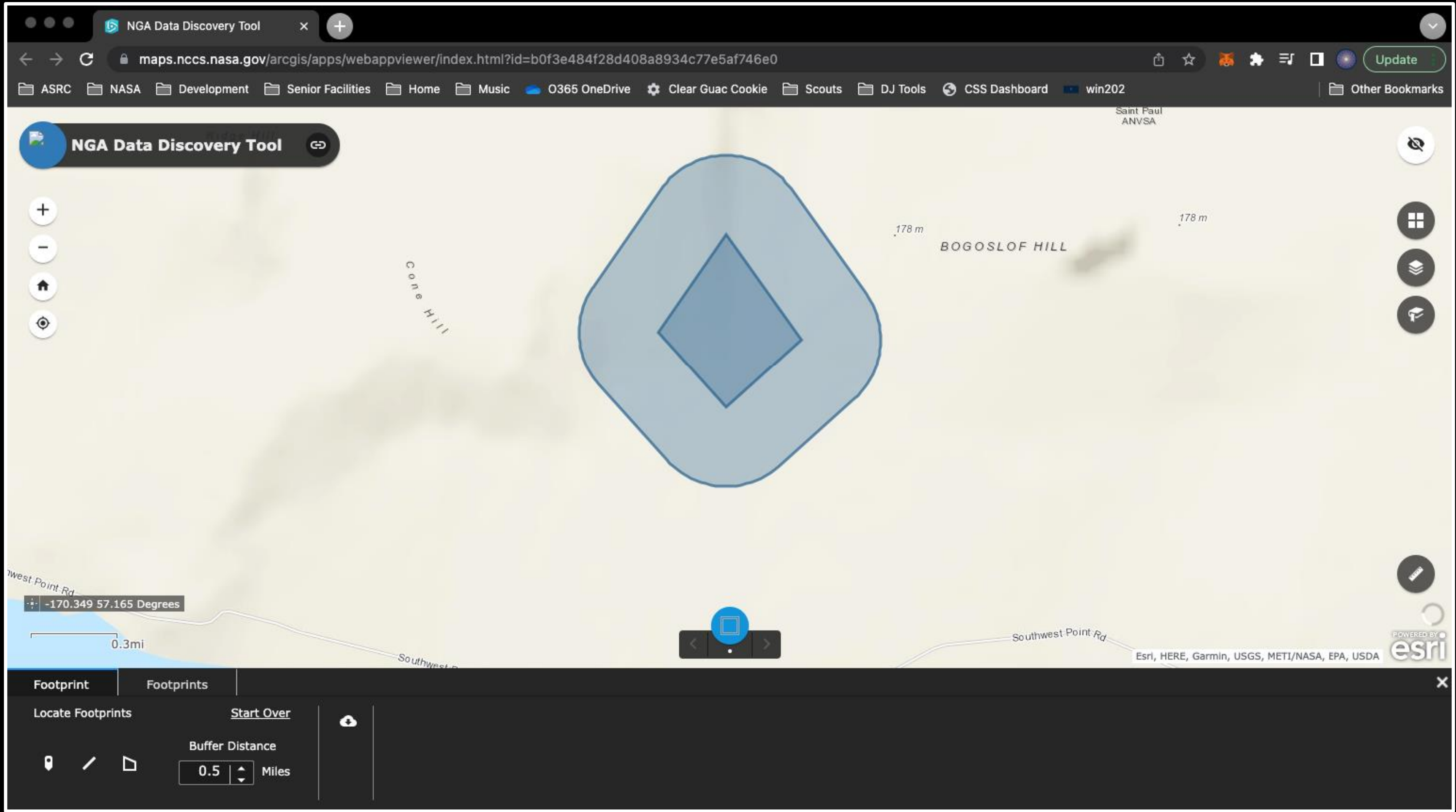
Data Discovery Tool

- Click Spatial Search – select point/line/polygon, set buffer distance



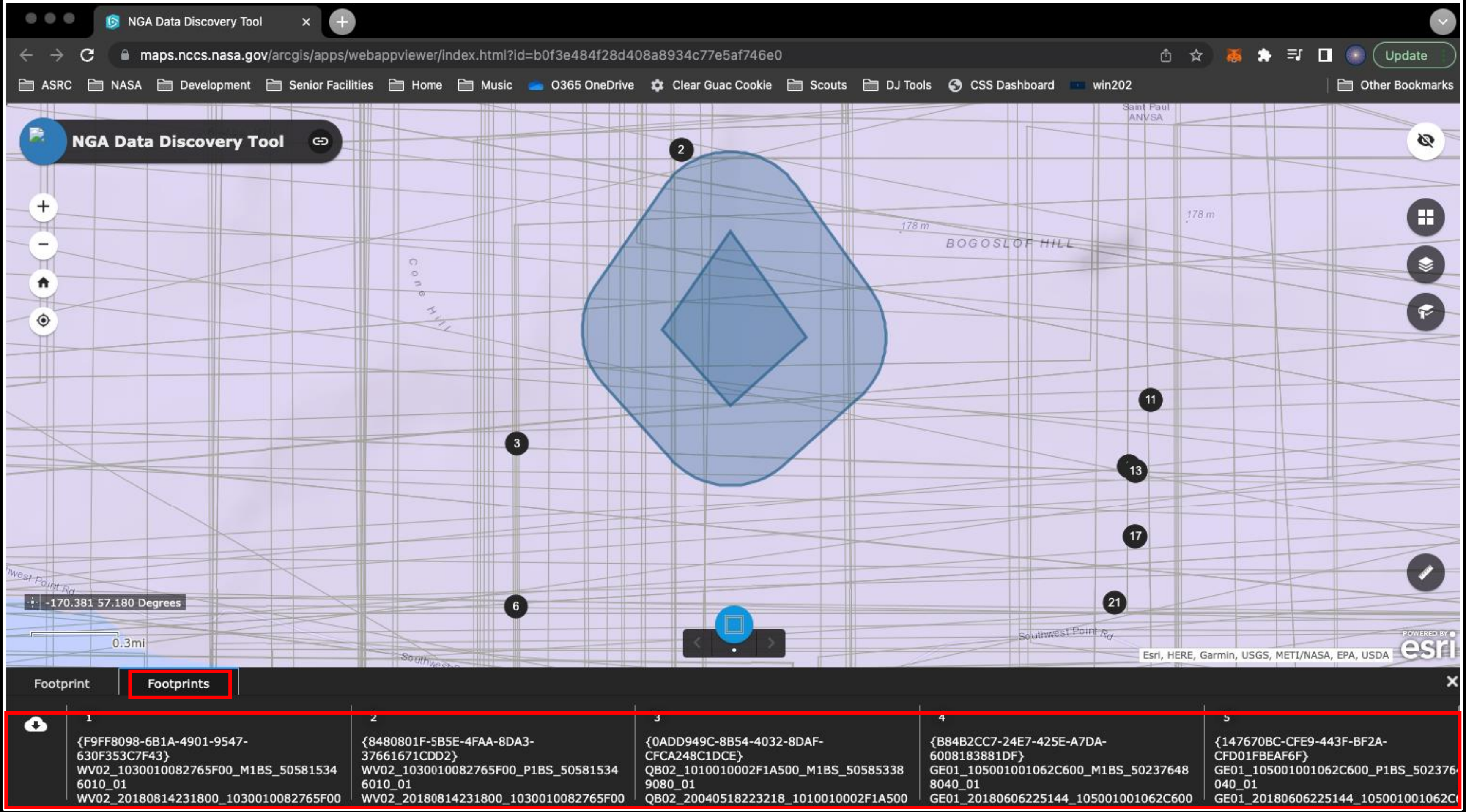
Data Discovery Tool

- After selection, system will generate buffer and zoom to location



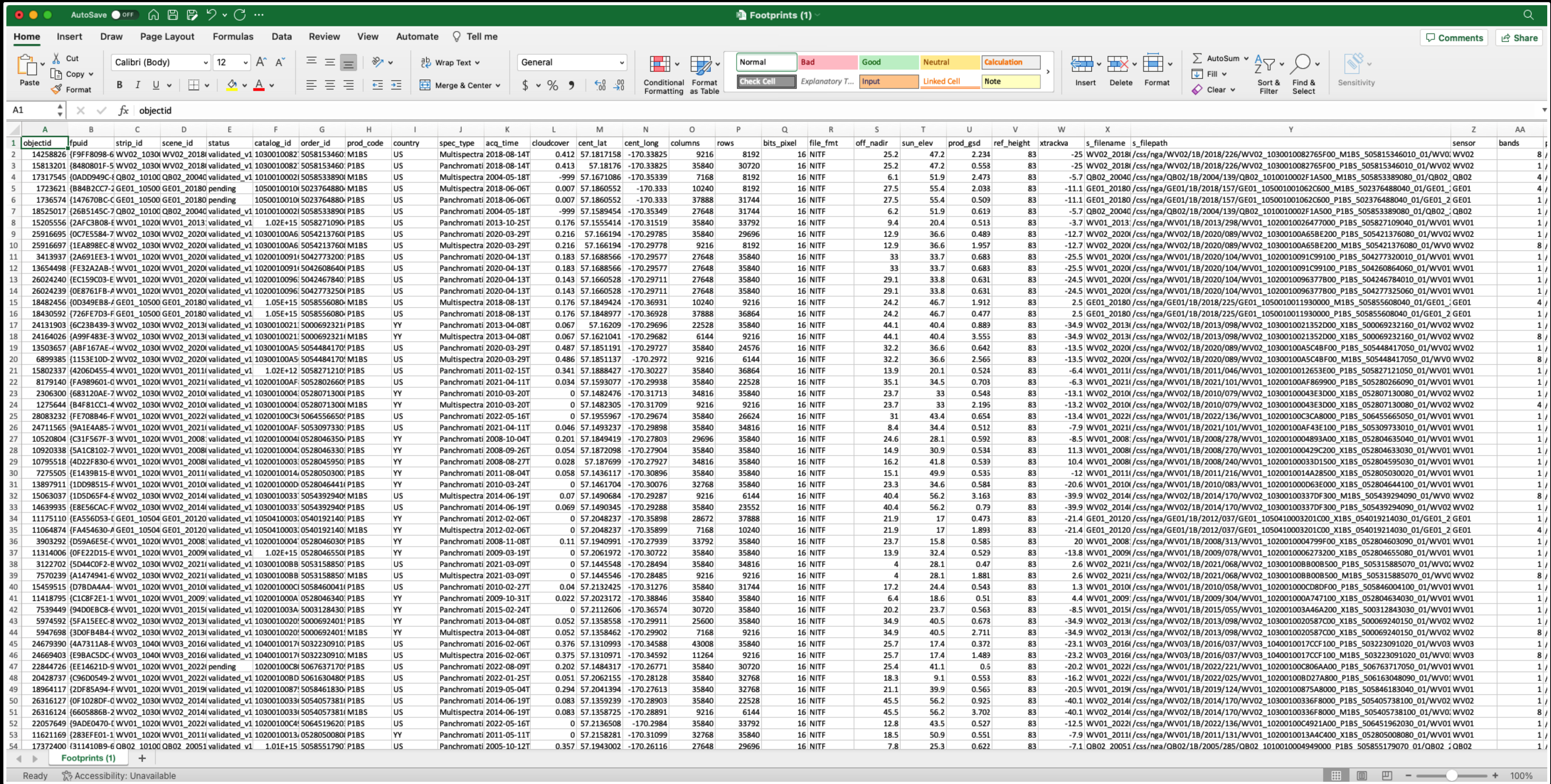
Data Discovery Tool

- Click Footprints to see the images within that buffered location



Data Discovery Tool

- Click the Download button ( ) to save the results as a csv file



Advanced Data Discovery Tool

- With Advanced Search, specify search parameters (sensor, cloudcover, etc.) and export the resulting footprint feature class