GHRC

LONG TERM STRATEGIC PLAN

Dr. Rahul Ramachandran

DAAC Manager

NASA/MSFC

rahul.ramachandran@nasa.gov

Presented at the GHRC User Working Group Meeting
October 7, 2015

Motivation

- Motivation
  - User Working Group (UWG) identified mismatch between existing GHRC Mission and its data holdings

- UWG recommendations
  - Evaluation and modification of GHRC mission and objectives
  - Develop a 5 year vision/strategic plan for GHRC

Existing Problem

- GHRC Current Mission Statement
  - "To serve as NASA's Earth science *data stewards* for scientific, educational, commercial and governmental communities, with a focus on data for the global hydrologic cycle
    - Hydrologic Cycle
    - Severe Weather Interactions
    - Lightning
    - Atmospheric Convection"

Actual data holdings have *evolved over time* and are more focused on lightning, field campaigns, selected passive microwave and extreme weather

Strategic Plan Development: Approach Used

- Metadata/Metrics Analysis:
  - Thorough analysis of GHRC and other DAAC data holdings, GHRC metrics and its existing operations
- SWOT Analysis:
  - Strength Weaknesses Opportunities Threats (SWOT) analysis on possible future directions including:
    - *Retain Current Focus*
    - *Field Campaigns*
    - *Atmospheric Phenomena and its governing physical and dynamical processes*

Metadata / Metrics Analysis

...

DAAC's Thematic Alignment

AGRICULTURE (1878)
agricultural aquatic sciences, agricultural chemicals, agricultural engineering, agricultural plant science, animal commodities show all...

ATMOSPHERE (7746)
aerosols, air quality, altitude, atmospheric chemistry, atmospheric electricity show all...

GHRC, GESDISC, ASDC, MODAPS

BIOLOGICAL CLASSIFICATION (4131)
animals/invertebrates, animals/vertebrates, bacteria/archaea, fungi, plants show all...

BIOSPHERE (7028)
aquatic ecosystems, ecological dynamics, terrestrial ecosystems

ORNL, OPBG

CLIMATE INDICATORS (384)
atmospheric/ocean indicators, biospheric indicators, cryospheric indicators, land surface/agriculture indicators, paleoclimate indicators show all...

CRYOSPHERE (2755)
frozen ground, glaciers/ice sheets, sea ice, snow/ice

NSIDC, ASF

HUMAN DIMENSIONS (3830)
boundaries, economic resources, environmental governa... environmental impacts, habitat ... tation show all...

SEDAC

LAND SURFACE (5505)
... frozen ground, ... perature, land use/land

LP-DAAC

OCEANS (9731)
... topography, ... t monitoring,

PODAAC, OPBG

PALEOCLIMATE (1485)
ice core records, land records, ocean/lake records, paleoclimate reconstructions show all...

SOLID EARTH (3013)
earth gase... geodetics, geomagne... processes show all...

CDDIS, ASF

SPECTRAL/ENGINEERING (2747)
gamma ray, infrared wavelengths, lidar, microwave, platform characteristics show all...

SUN-EARTH INTERACTIONS (358)
ionosphere/magnetosphere dynamics, solar activity, solar energetic particle flux, solar energetic particle properties show all...

TERRESTRIAL HYDROSPHERE (3214)
glaciers/ice sheets, ground water, snow/ice, surface water, water quality/water chemistry show all...

*Dataset/Scientific Keywords Underlined – primary DAAC for the theme

Takeaway:
- Six DAACs cover broad thematic focus
  - PODAAC, NSIDC, SEDAC, CDDIS, ORNL, LP
- Four DAACs focusing on Atmosphere
  - GHRC, GESDISC, ASDC, MODAPS

Within Atmosphere Theme

Science Keywords > ATMOSPHERE

AEROSOLS (765)
aerosol backscatter, aerosol extinction, aerosol optical depth/thickne… radiance - show all...

ASDC, GESDISC

AIR QUALITY (865)
carbon …s, lead, nitrogen oxides, particul…

GESDISC

ALTITUDE (859)
barometric altitude, geopotential height, mesopause, planetary boundary layer height, station height - show all...

ATMOSPHERIC CHEMISTRY (1659)
carbon and hydrocarbon compounds, halocarbons and halogens, hydrog… nitrogen compounds, oxygen compounds - sho…

ASDC

ATMOSPHERIC ELECTRICITY (97)
atmospheric conduc… field, lightning, total electron content - show all...

GHR

ATMOSPHERIC PHENOMENA (520)
cyclones, drought, fog, freeze, frost - show all...

GHRC

ATMOSPHERIC PRESSURE (1913)
anticyclones/cyclones, atmospheric pressure measurements, differential pressure, gravity wave, hydrostatic pressure - show all...

ATMOSPHERIC RADIATION (2328)
absorption, airglow, albedo, anisotropy, atmospheric emitted radiation - show all...

ASDC

ATMOSPHERIC TEMPERATURE (3179)
air temperature, atmospheric stability, boundary layer temperature, degree days, deiced temperature - show all...

ATMOSPHERIC WATER VAPOR (2544)
condensation, dew point temperature, evaporation, evapotranspiration, humidity - show all...

ATMOSPHERIC WINDS (2281)
boundary layer winds, convection, convergence/divergence, flight level winds, streamfunctions - show all...

CLOUDS (1401)
cloud droplet distribution, cloud dynamics, cloud microphysics, cloud prop… diative transfer - show all...

ASDC

PRECIPITATION (2015)
acid rain, droplet… s - show all...

GESDISC, GHRC

Takeaway:

- ASDC –focus areas on ATS Radiation/Chemistry
- Overlap between GESDISC, GHRC – Precipitation
- GHRC focus
  - Electricity
  - *Atmospheric Phenomena (Focus on Weather)*

NASA Science Program Alignment

These science questions translate into seven overarching science goals to guide the Earth Science Division's selection of investigations and other programmatic decisions:

- Advance the understanding of changes in the Earth's radiation balance, air quality, and the ozone layer that result from changes in atmospheric composition (*Atmospheric Composition*)
- Improve the capability to predict weather and extreme weather events (*Weather*)
- Detect and predict changes in Earth's ecological and chemical cycles, including land cover, biodiversity, and the global carbon cycle (*Carbon Cycle and Ecosystems*)

Seven Key Earth Science Focus Areas:

- Atmospheric Composition
- ***Weather***
- Carbon Cycle and Ecosystems
- **Water and Energy Cycle**
- Climate Variability and Change
- Earth Surface and Interior
- ***Societal Benefits/Applications***

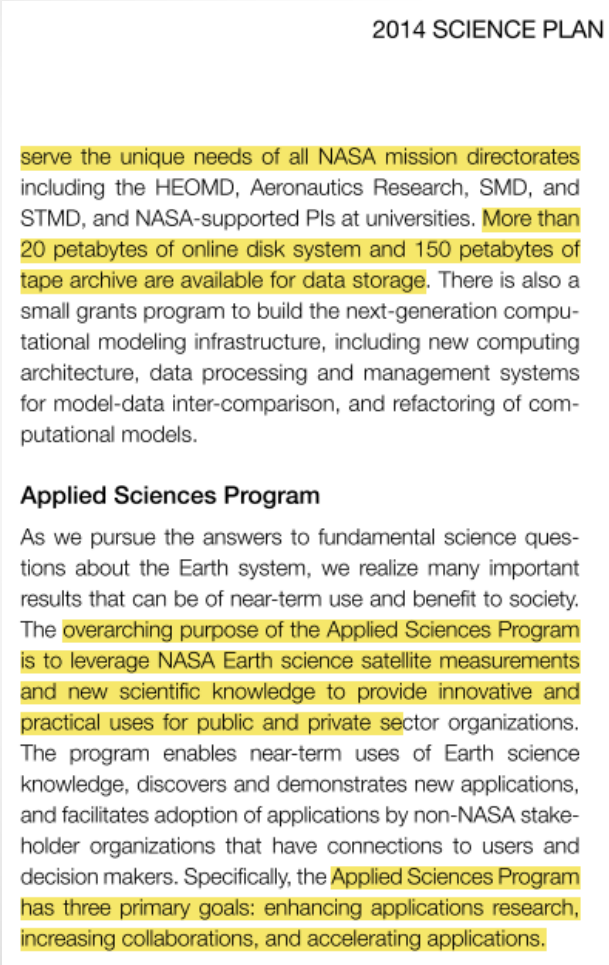
Current Mission Alignment

- Too much overlap with other DAACs

Takeaway: ***Possible New Mission Alignment***

- Can provide unique value by focusing on these two areas

Applied Science Program Alignment



2014 SCIENCE PLAN

serve the unique needs of all NASA mission directorates including the HEOMD, Aeronautics Research, SMD, and STMD, and NASA-supported PIs at universities. More than 20 petabytes of online disk system and 150 petabytes of tape archive are available for data storage. There is also a small grants program to build the next-generation computational modeling infrastructure, including new computing architecture, data processing and management systems for model-data inter-comparison, and refactoring of computational models.

Applied Sciences Program

As we pursue the answers to fundamental science questions about the Earth system, we realize many important results that can be of near-term use and benefit to society. The overarching purpose of the Applied Sciences Program is to leverage NASA Earth science satellite measurements and new scientific knowledge to provide innovative and practical uses for public and private sector organizations. The program enables near-term uses of Earth science knowledge, discovers and demonstrates new applications, and facilitates adoption of applications by non-NASA stakeholder organizations that have connections to users and decision makers. Specifically, the Applied Sciences Program has three primary goals: enhancing applications research, increasing collaborations, and accelerating applications.

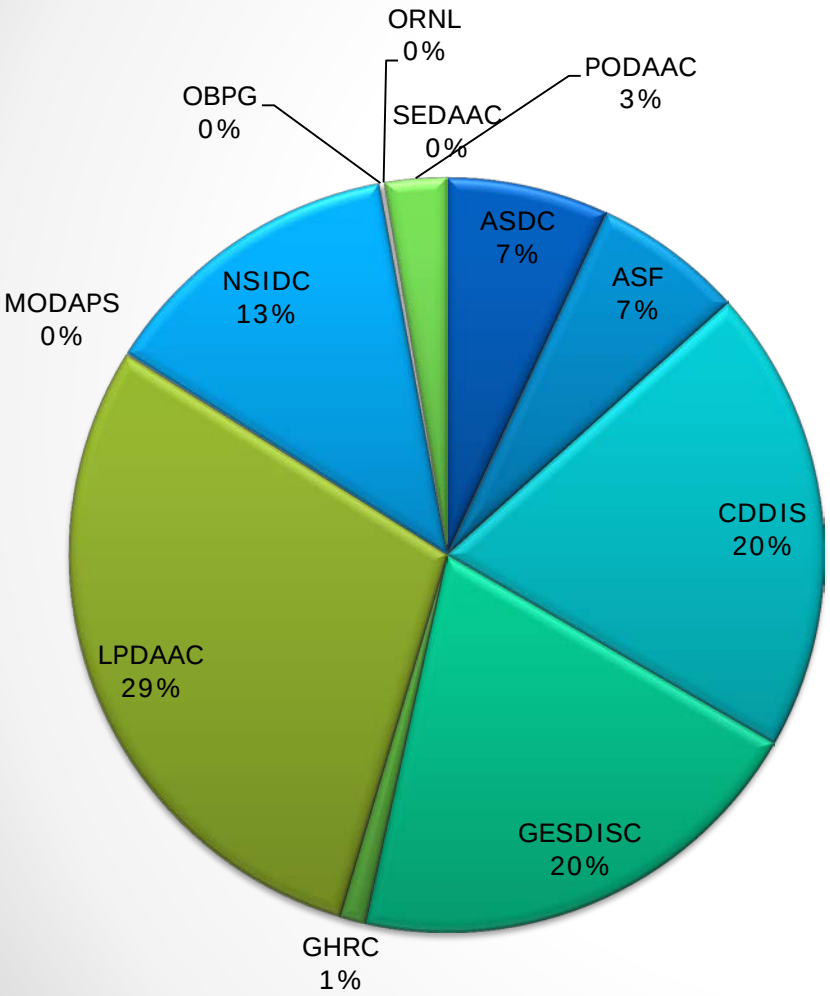
Applied Science Program: Application Area

- ***Disasters***
- Ecological Forecasting
- Health and Air Quality
- Water Resources

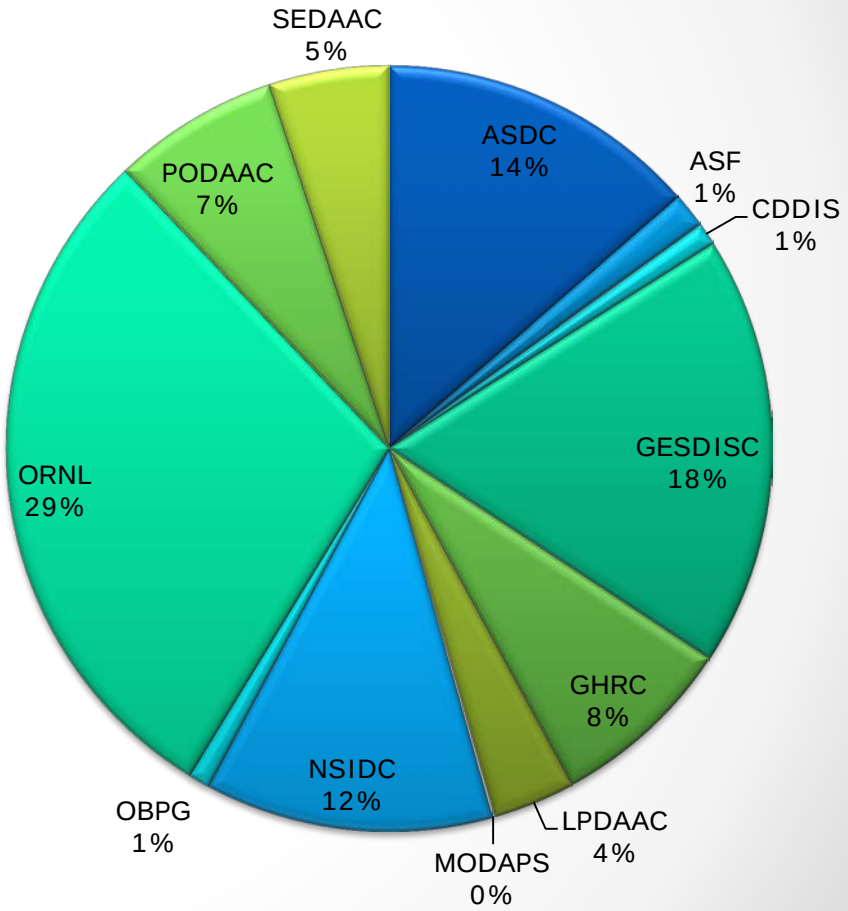
Takeaway:

- *Focus on a subset of disasters that fall within our new Atmospheric Phenomena (Weather) scope*
- *Archive reusable products developed from NASA Applied Science research projects*
- *Support NASA's disaster activities by providing "curated data albums" for events on demand*

Data Portfolio Across DAACs



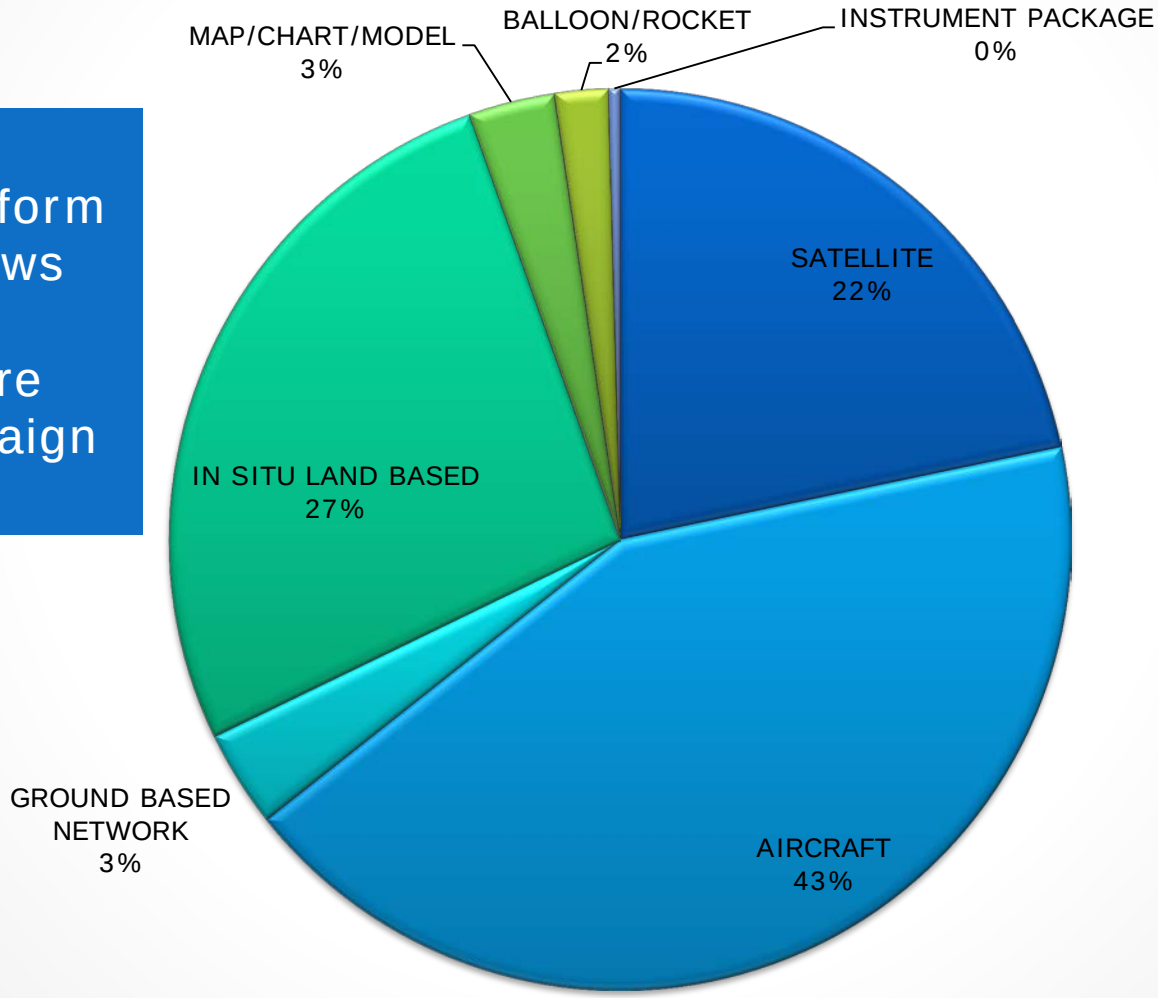
Data Granules



Data Collection

GHRC Platform Profile

Takeaway:
GHRC platform profile shows 78% of holdings are field campaign data



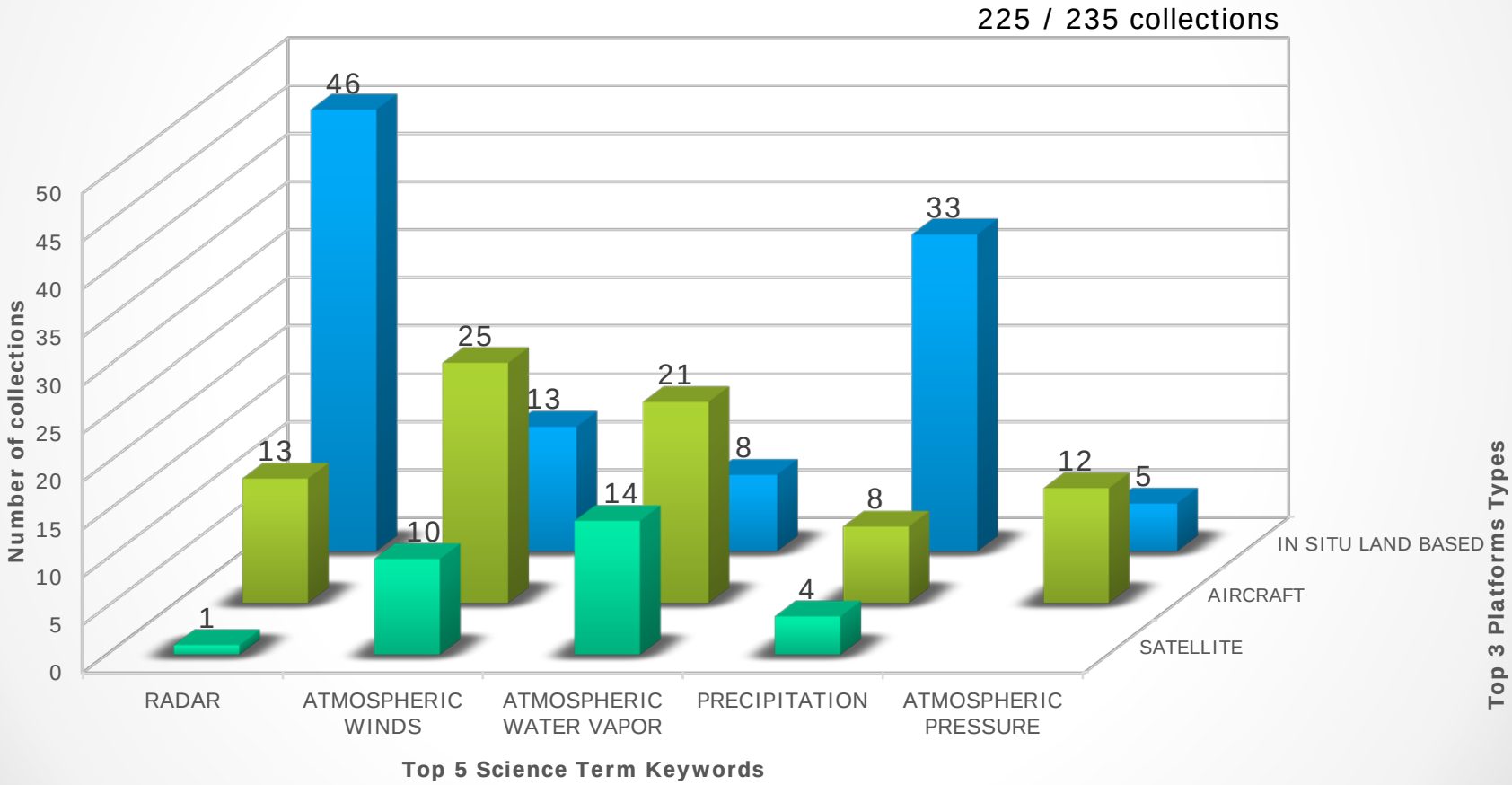
Current Holdings: By the Numbers

Holdings	Counts
Datasets	235
Granules (i.e., data files)	721,498
Field Campaigns	21
Variables measured	82
Science Keywords	29
Instruments	109
Platforms	46
Platform Types	8

Source: GHRC Dashboard

Slide source: Patrick Gatlin

Current Holdings: Top 5



Slide source: Patrick Gatlin

Source: GHRC Dashboard

Current Holdings: Strategic Areas

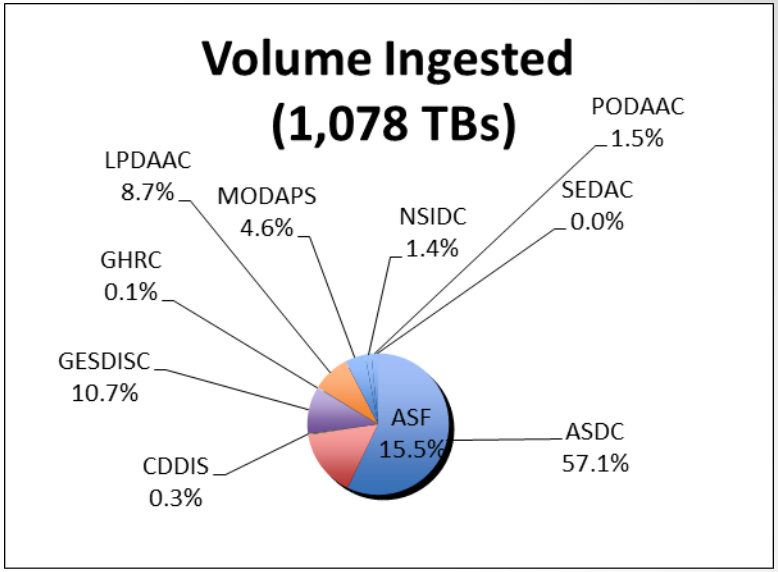
Atmospheric Phenomena / Hazards	Field Campaigns
Hurricanes	CAMEX-3, CAMEX-4, TCSP, NAMMA, GRIP, HS3
Lightning	ACES
Flooding	Precipitation Science: GPM GV (LPVEx, MC3E, GCPEx, IFloodS, IPHEx, OLYMPEx)
	Tropical Composition: TC4

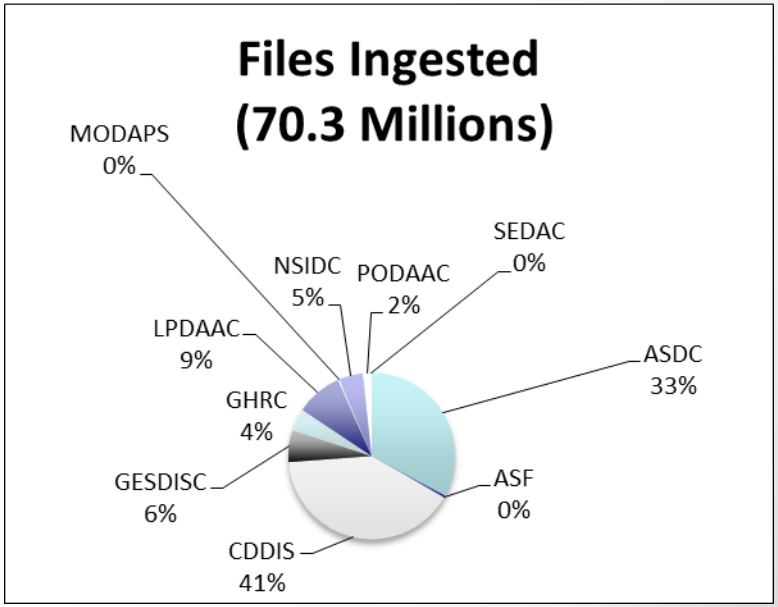
Slide source: Patrick Gatlin

Ingest

Ingest is the amount of data coming into a DAAC over a period of time and includes all product levels. For this report, the data is presented as the amount of data entered into each DAAC during FY2014. The sum of all data centers is the total ingest for EOSDIS. OBPG and ORNL are unable to provide ingest metrics at this time.

DAAC	Volume (TBs)	Files (Millions)
ASDC	615.5	23.1
ASF	166.6	0.3
CDDIS	2.7	28.5
GESDISC	115.8	4.3
GHRC	1.6	3.1
LPDAAC	93.7	6.3
MODAPS	49.9	0.3
NSIDC	15.4	3.2
PODAAC	16.7	1.2
SEDAC	0.0	0.0
Total	1,077.9	70.3





Takeaway: GHRC has small ingest rate partly because of the type of data (low volume/ high number of files)

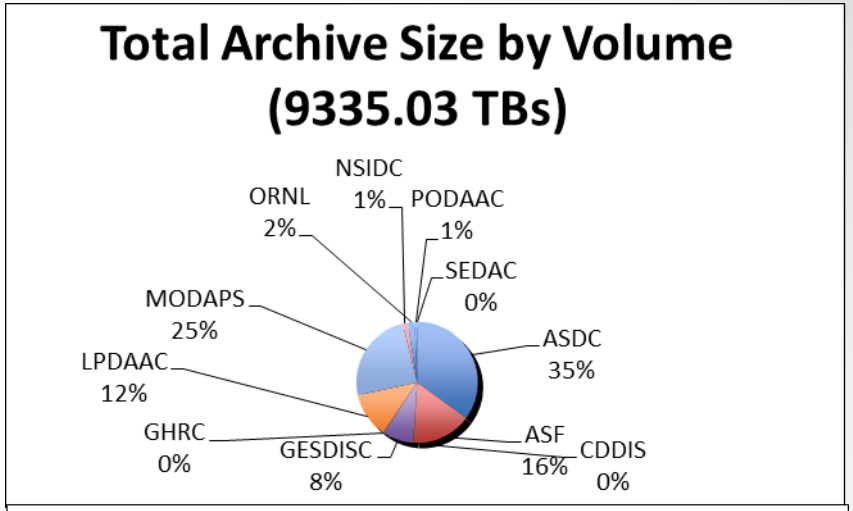
Source: EOSDIS FY2014 Annual Metrics Report, Wanchoo et al, January 2015

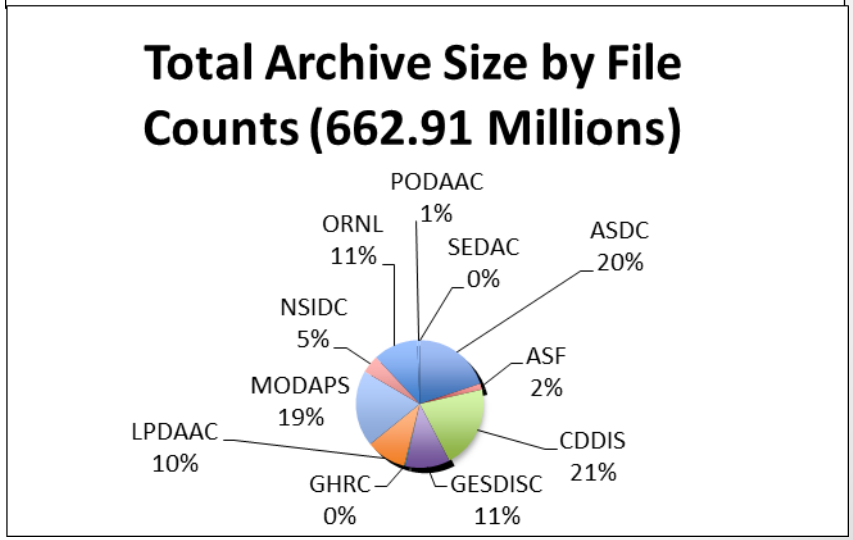
Total Archive Size (End of FY2014

The Total Archive Size describes the EOSDIS archive at the end of FY2014.

DAAC	Volume (TBs)	Files (Millions)
ASDC	3,268.56	131.31
ASF	1,486.06	10.26
CDDIS	11.42	139.14
GESDISC	775.31	74.87
GHRC	8.81	2.00
LPDAAC	1,131.11	67.71
MODAPS	2,298.74	128.11
NSIDC	121.29	30.06
ORNL	175.49	75.16
PODAAC	54.88	4.30
SEDAC	3.36	0.00
Total	9,335.03	662.91

Takeaway: GHRC is one of the smallest DAAC (by data volume/number of files)



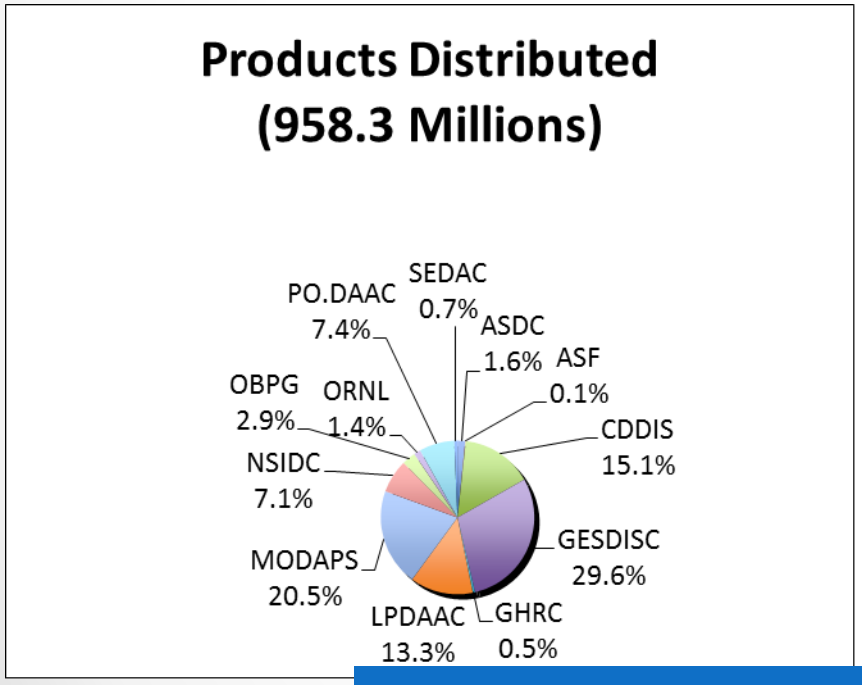


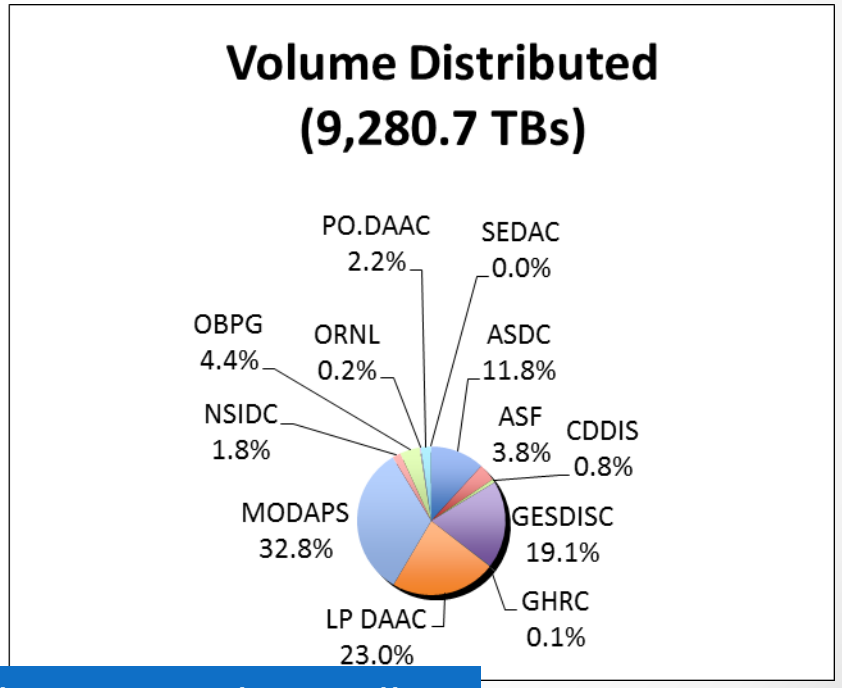
Source: EOSDIS FY2014 Annual Metrics Report, Wanchoo et al, January 2015

Distribution By DAAC

Distribution presents the amount of data successfully distributed to Public Users.

Distribution by DAAC	ASDC	ASF	CDDIS	GESDISC	GHRC	LPDAAC	MODAPS	NSIDC	OBPG	ORNL	PO.DAAC	SEDAC	Total
Products (Millions)	15.47	1.11	144.29	283.26	4.50	127.08	196.14	67.73	27.46	13.61	71.26	6.39	958.31
Total Volume (TBs)	1,094.30	349.97	72.34	1,770.65	8.98	2,132.95	3,045.81	169.88	404.46	21.51	207.11	2.72	9,280.67





Takeaway: GHRC distributes mostly small sized data files

Source: EOSDIS FY2014 Annual Metrics Report, Wanchoo et al, January 2015

Active Ratio

Active Ratio = Volume/Files Distributed/ Archive Volume/Distributed

- Measure of useful data sets in the archive wanted by the user community

DAAC	Volume (TBs)	Files (Millions)	Distribution (Vol)	Distribution (Files)	Active Ratio (Vol)	Active Ratio (Files)
ASDC	3,268.56	131.31	1094.3	15.47	33.48%	11.78%
ASF	1,486.06	10.26	349.97	1.11	23.55%	10.82%
CDDIS	11.42	139.14	72.34	144.29	633.34%	103.70%
GESDISC	775.31	74.87	1770.65	283.26	228.38%	378.35%
GHRC	**8.81**	**2.00**	**8.98**	**4.5**	**101.95%**	**225.40%**
LPDAAC	1,131.11	67.71	2131.92	127.08	188.48%	187.68%
MODAPS	2,298.74	128.11	3045.81	196.14	132.50%	153.11%
NSIDC	121.29	30.06	169.88	67.73	140.07%	225.31%
ORNL	175.49	75.16	21.51	13.61	12.26%	18.11%
PODAAC	54.88	4.30	207.11	71.26	377.39%	1656.85%
SEDAC	3.36	0.00	2.72	6.939	80.97%	0.00%

Takeaway: GHRC archives and distributes useful data

Active Ratio by Volume: GHRC 7th/11

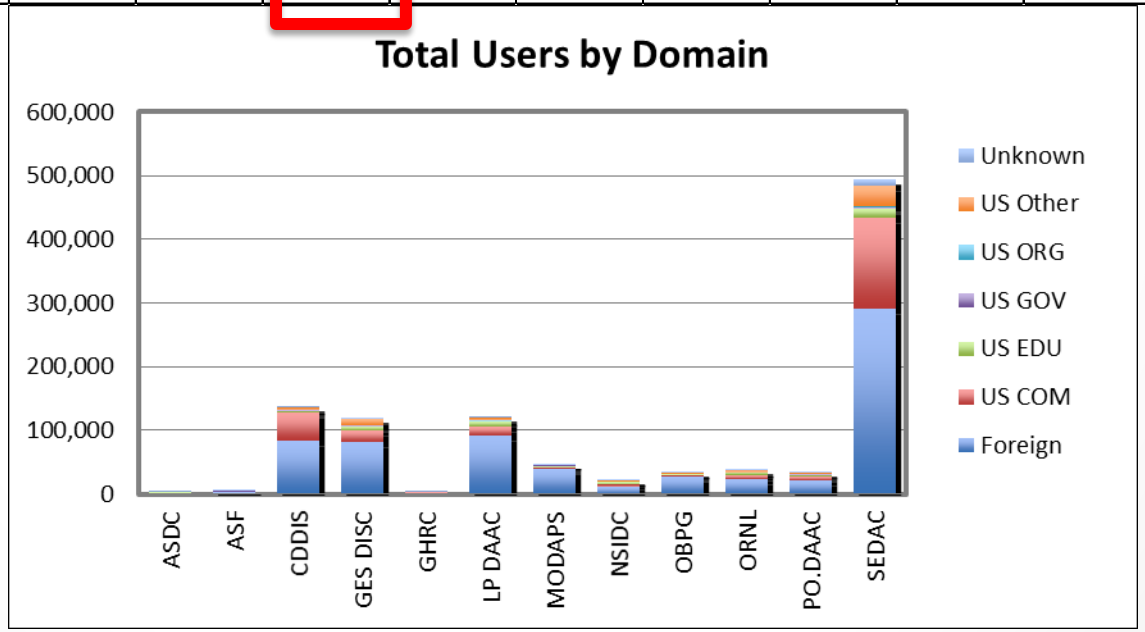
Active Ratio by Files: GHRC 3rd/11

Source: EOSDIS FY2014 Annual Metrics Report, Wanchoo et al, January 2015

Distinct Data Users

Distinct Data Users presents the number of distinct public users who received data product files.

Distinct Data Users By	ASDC	ASF	CDDIS	GES DISC	GHRC	LP DAAC	MODAPS	NSIDC	OBPG	ORNL	PO.DAAC	SEDAC	Total
Foreign	1,411	2,175	82,940	81,987	2,069	92,265	39,086	11,707	28,041	24,029	20,720	291,368	677,798
US COM	556	495	45,386	16,942	467	13,653	2,509	4,231	1,180	6,011	5,794	141,528	238,752
US EDU	832	1,258	2,200	6,822	484	7,251	1,909	2,387	1,167	3,447	2,938	14,302	44,997
US GOV	851	216	2,145	2,546	282	1,501	979	628	237	604	2,538	3,554	16,081
US ORG	14	59	154	314	3	393	26	61	66	69	66	1,279	2,504
US Other	33	402	3,877	9,105	258	5,046	1,438	1,304	2,384	2,176	1,979	33,023	61,025
Unknown	10	81	1,148	1,048	46	1,299	356	361	254	678	524	9,182	14,987
Total Distinct Data Users	3,707	4,686	137,850	118,764	3,609	121,408	46,303	20,679	33,329	37,014	34,559	494,236	1,056,144



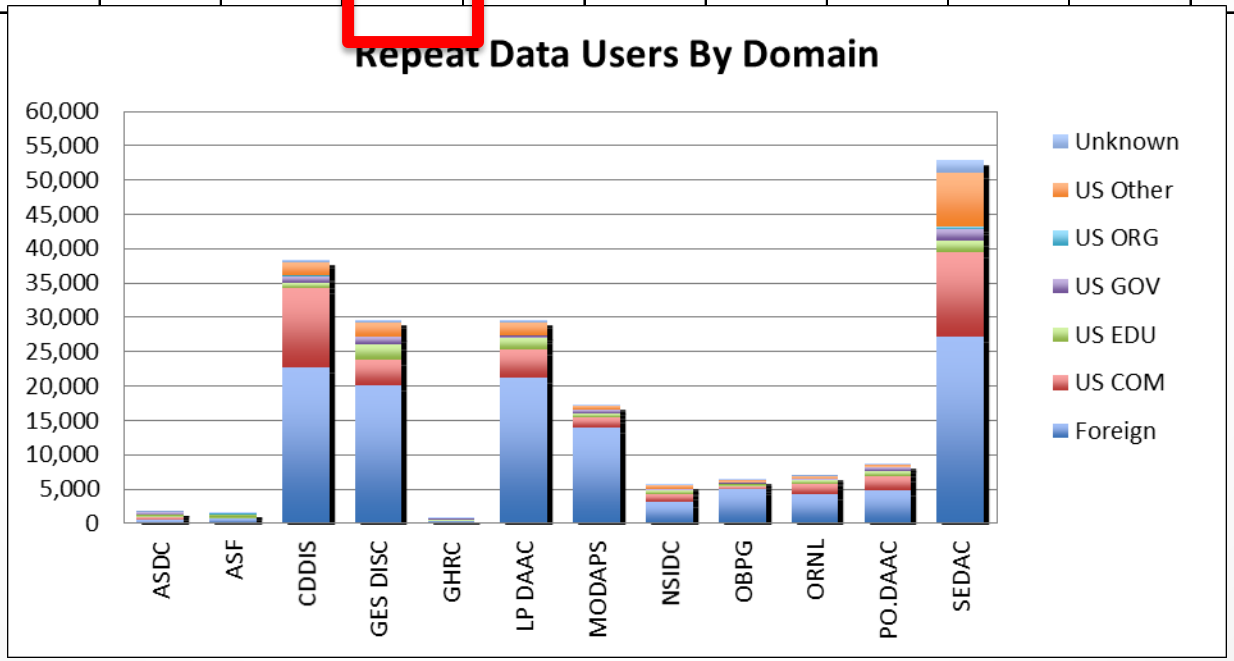
Takeaway: Lowest user numbers amongst all DAACs. GHRC does a poor job of reaching out to a wider user community

Source: EOSDIS FY2014 Annual Metrics Report, Wanchoo et al, January 2015

Repeat Data Users

Repeat users are distinct Public users who received data on more than one day in the fiscal year.

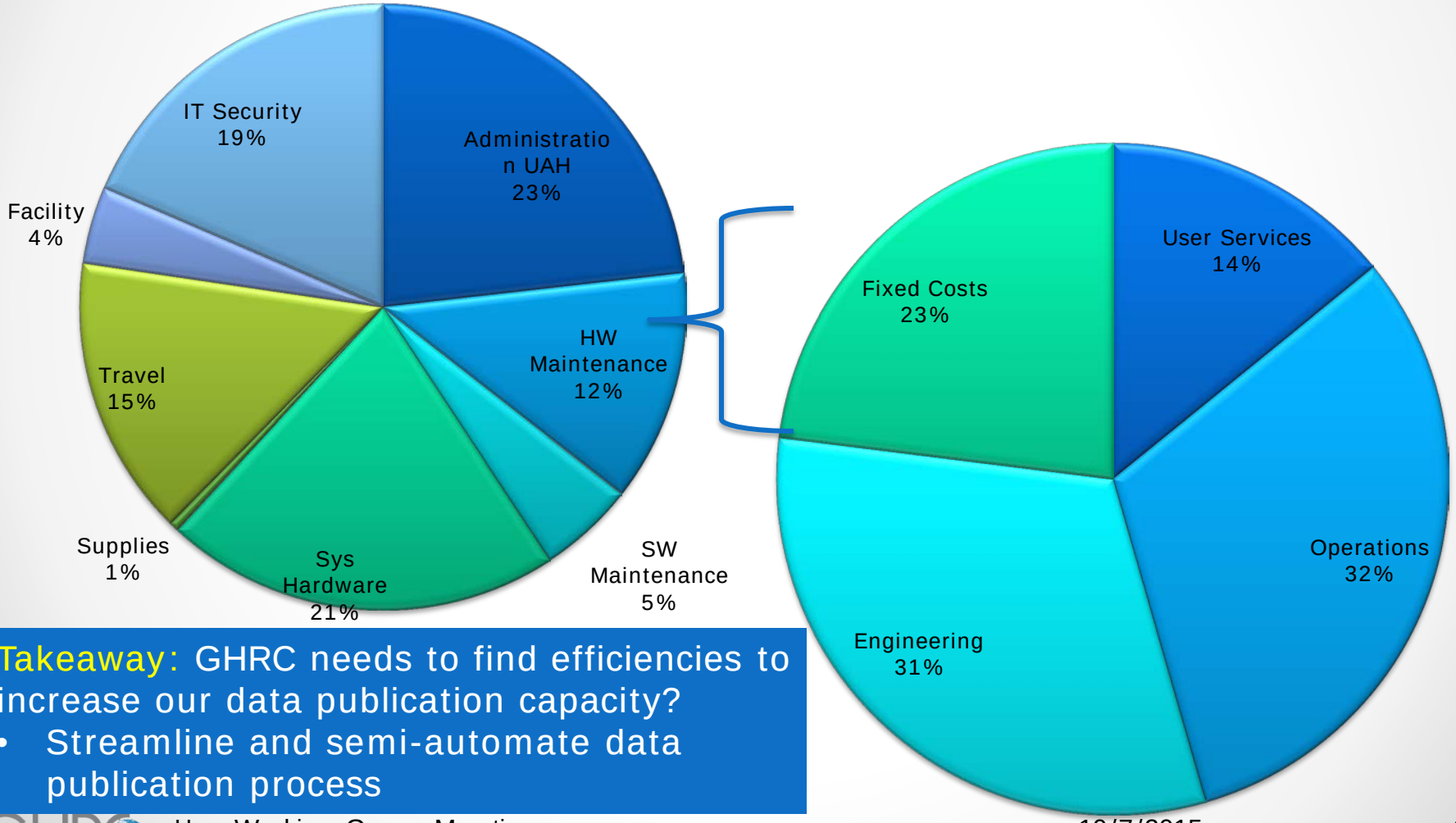
Repeat Data Users By Domain	ASDC	ASF	CDDIS	GES DISC	GHRC	LP DAAC	MODAPS	NSIDC	OBPG	ORNL	PO.DAAC	SEDAC	Total
Foreign	556	631	22,639	20,044	349	21,312	14,041	3,054	5,037	4,215	4,860	27,116	123,854
US COM	268	151	11,702	3,716	68	3,947	1,342	1,146	321	1,530	1,955	12,368	38,514
US EDU	365	481	641	2,327	152	1,717	664	607	451	568	808	1,776	10,557
US GOV	381	89	1,030	1,122	73	370	451	249	100	119	587	1,578	6,149
US ORG	5	31	99	65	0	111	7	26	34	16	12	409	815
US Other	8	151	1,935	2,039	48	1,771	566	403	425	353	347	7,751	15,797
Unknown	3	18	289	317	19	301	184	123	82	141	206	1,843	3,526
Total Repeat Data Users	1,586	1,552	38,335	29,630	709	29,529	17,255	5,608	6,450	6,942	8,775	52,841	199,212



Takeaway: Lowest repeat users amongst all DAACs. GHRC does a poor job of reaching out to a wider user community

Source: EOSDIS FY2014 Annual Metrics Report, Wanchoo et al, January 2015

Core Funding Profile



Takeaway: GHRC needs to find efficiencies to increase our data publication capacity?

- Streamline and semi-automate data publication process

Metadata/Metrics Analysis Summary

- GHRC is one of the smallest DAAC (by data volume/number of files)
- Atmosphere theme contains GHRC, GESDISC, ASDC, MODAPS with significant overlap between GHRC and GESDISC
- GHRC platform profile shows 78% of holdings are field campaign
- GHRC is active (3rd by Distribution/Archive ratio)
- GHRC user numbers are low and need to be improved
- GHRC Operational efficiency needs to be improved by streamlining internal processes

SWOT Analysis

Future Directions:

- Retain Current Focus
- Atmospheric Phenomena (Weather/Hazards)
- Field Campaigns

Direction: Retain Current Focus

Strengths

- Small both in size of data and personnel, can adapt quickly
- Technical skills to handle data at all production levels
- AMPR/LIS/HIRAD instrument at MSFC

Weaknesses

- Lack a perceived science identity
- Inability to provide quantifiable ROI
- Lack of a strategic plan with the current focus

☒ Opportunities

- Limited, with no foreseeable missions or new data sources

Threats

- Can be perceived as not providing enough value to the stakeholder community

Takeaway: Not a viable option.
Data holdings mismatch with existing mission statement can lead to wrong perception within the science community

Direction: Atmospheric Phenomena (and Hazards)

Strengths

- Already hold Lightning and Hurricane data sets
- Collocation with the HSV NWS , NASA SPoRT and NASA SERVIR
- Collocation with UAH ATS/ESSC Severe Weather researchers

Weaknesses

- GHRC is a 8x5 operation; disasters support may require 24x7 operations.
- May need new “services” portfolio needed

☑ Opportunities

- Serve as the interface to other agencies wanting NASA Data for specific hazards
- Provide data stewardship services to Applications Science program funded research
- Create Virtual Collections using data from other DAAC with value added
- User base expands from “researchers” to federal, state and local agencies
- Build GHRC’s GIS expertise leveraging UAH ESSC expertise

Threats

- Just focusing on hazards does not align with other DAACs science themes
- Science missions are rarely focused on Hazards
- Could be perceived as overlapping with mission of NOAA and other agencies

Takeaway: Viable option – but needs to support both NASA Research and Applications Program

Direction: Field Campaign

Strengths

- Extensive field campaign data holdings
- Strong working relationships with Hurricane Science and GPM Ground Validation teams
- Good relationship with NASA ESPO (who manage airborne campaigns)
- GHRC's Field Campaign collaboration tools are well used

Weaknesses

- Data publication process is cumbersome and manual
- Need better coordination with ESPO to link to their holdings
- Can't become "the field campaign DAAC" as other DAACs have their own field campaign heritage and expertise.

☑ Opportunities

- Build state of the art infrastructure on the new HS3 information system currently being developed
- Possible future Earth Venture missions, especially if they have Hurricane Science focus
- Possible future satellite GV missions

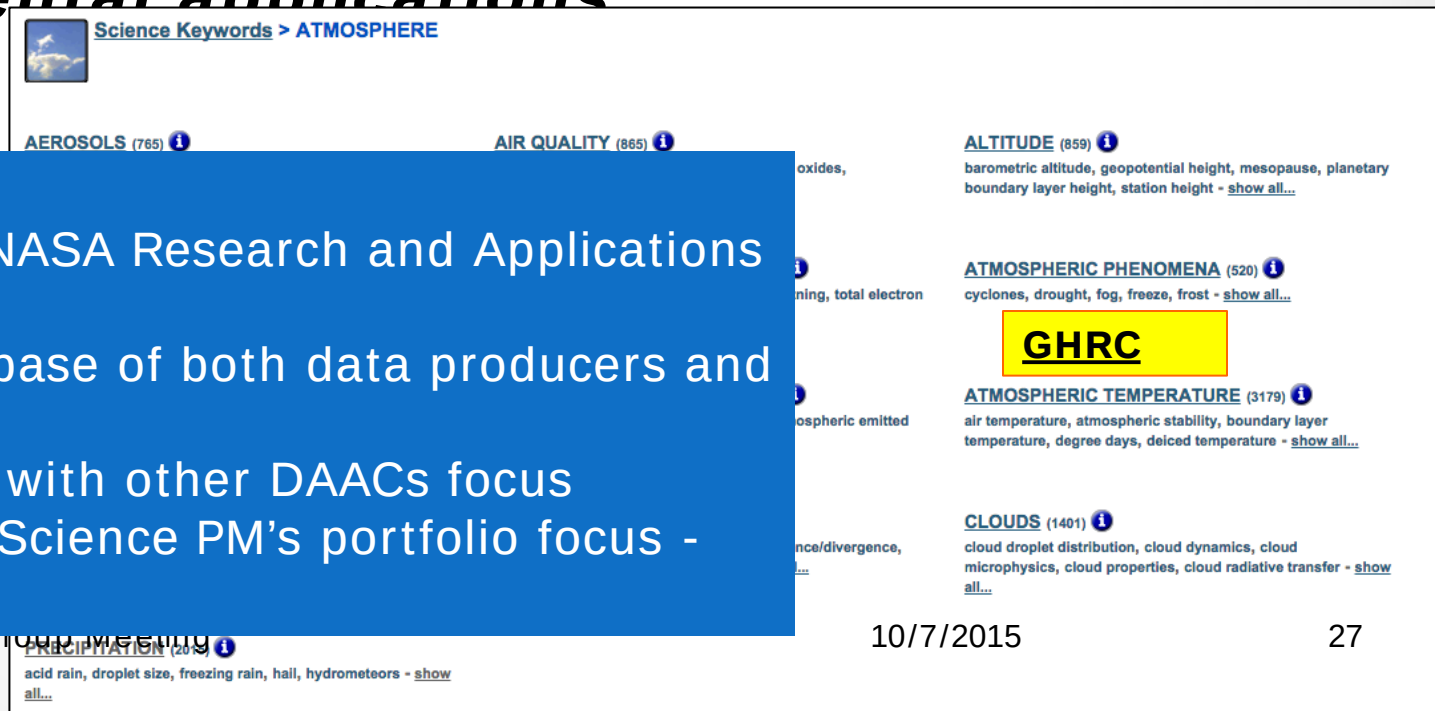
Threats

- Field campaign data are expensive and hard to manage

Takeaway: Cannot be the primary focus but can be a secondary expertise within the selected science theme

New Mission Statement

The mission of the GHRC is to provide a comprehensive archive of both data and knowledge augmentation services with a focus on ***atmospheric phenomena***, *its governing dynamical and physical processes and associated* ***environmental applications***



Takeaway:

- Supports both NASA Research and Applications Program
- Expands GHRC base of both data producers and data users
- Doesn't overlap with other DAACs focus
- Still aligns with Science PM's portfolio focus - *Weather*

New Specializations

- Provide data stewardship services to both NASA's Research and *Applied Science* funded programs focusing on the new theme
- Hazard/Disaster specialization (environmental applications) to help accelerate operational use of NASA data
  - Serve as the interface to other agencies needing NASA data for specific hazards, past and/or present
- Provide state of the art data and information system infrastructure to support all stages of *Field Campaigns* from mission planning, coordination to archive

Approach to Move Forward

- Expect ***no or minimal change*** to existing GHRC budget profile
- Find operational efficiencies for cost savings and apply them towards moving in the new direction
- Use programs such as ACCESS, AIST to incrementally build GHRC's capabilities
- Utilize existing HS3 data system (currently in design phase) to support the Field Campaign specialization

Data Stewardship Levels of Service Matrix: *Blueprint for the Future*

• • •

- Matrix will be used to inform and create future GHRC Annual Work Plans and Research and Development Focus
- Goal is to incrementally move in this new direction and add new functionality over next 5-10years

Data Stewardship Levels : Peng, Ge, Jeffrey L Privette, Edward J Kearns, Nancy A Ritchey, and Steve Ansari. 2015. “A UNIFIED FRAMEWORK FOR MEASURING STEWARDSHIP PRACTICES APPLIED TO DIGITAL ENVIRONMENTAL DATASETS” 13 (February): 231–253.

Stewardship: Levels of Service

Functional Levels of Services	Current Core Services	Future Core Services	Hazard Specialization	Field Campaign Specialization
Accessibility (L4) - users can easily find and access data	• Data publicly available • Direct file downloads • Data and other research resources searchable online (Collection and Granule Level) • Standard data service (machine api) (OpenDAP) • Data search metrics compiled monthly (EMS)	• *Metrics dashboard (internal and online)*	• *Serve as the interface to other agencies wanting NASA Data for specific hazards, past and/or present* • *Provide a systematic recipe based curation process to create Disaster specific Virtual Collections to support NASA's Disaster response data effort*	• *Vector/tabular data stored in geospatial database* • *Decompose reports (ancillary documentation) and make the searchable* • *Coincidence search on data sets* • *Event based search - ex-intensification, genesis, landfall etc.* • *Visual search (by browse imagery)* • *FC metadata analytics*

Stewardship: Levels of Service

Functional Levels of Services	Current Core Services	Future Core Services	Hazard Specialization	Field Campaign Specialization
Usability (L3+) - users are able to use the data and learn whether the data are suitable for their own data requirements	• Standard based interoperability data format and metadata (HDF, netCDF) • Rich documentation - (user guides) • Core data preprocessing services - Subsetting/aggregation via OpenDAP • Limited visualization in multiple data formats	• *Data recipe code repository* • *Data usage reports to data producers* • *Machine readable documentation* • *Searches on Documents/Report* • *Visualization (Federated Giovanni instance)*	• *Provide a data recipes to allow easy ingest of NASA data into Decision Support Tools*	• *PI to query data (value based search, subset, analysis)??* • *Build validation tools* • *Information fusion : synthesize all information and data captured during field campaign* • *Visualization to recreate flights to provide context* • *GDAL - projections, statistical aggregation and format transformations* • *Support virtual collections (bundles)* • *Links to publication*

Stewardship: Levels of Service

Functional Levels of Services	Current Core Services	Future Core Services	Hazard Specialization	Field Campaign Specialization
Preservability (L3) - practices associated with data storage for resilience and use of community accepted best practices	• Designated repository for datasets • Redundant storage • Use of community standard archiving metadata	• *Archiving process performance controlled, measured and audited*	• *Support data publication and archive for individual PIs from Applied Science Program* • *Virtual Collections to support post-event analysis*	• *Complete archive package containing core and ancillary data products and reports*
Data Quality Assurance (L2) - procedures used to minimize defects in data during life cycle (missing data, outliers, redundant records) **Assessment** (L1) - to ensure the products are scientifically sound (beta, provisional, validated (stages 1-4)	• Limited quality assurance checks	• *Data quality procedure well documented, implemented* • *Research products assessed* • *Statistical QA/QC (mean, variance, time gaps etc)*		

Operational Improvements

- Processes
  - Create a new data life cycle document (check list) that covers all stages starting from data acquisition to publication and archive *(Helen)*
  - Minimize information collection redundancy and wasted interactions with data producers (*Helen*)
  - **Actively** engage user community – DAAC Scientist?
    - Attend Science Team Meetings, AMS and AGU science session to track emerging science questions
- Infrastructure components
  - Upgrade custom software with off the shelf components
  - Automate data ingestion, metrics reporting process
- Personnel skill mix
  - Need domain science expertise
  - May need data science skill set focused on Applied Science users

Discussion

- Mission Statement
  - Suggestions for improving the words on slides 27/28?
- Blueprint to influence our annual work plan
  - Additional capabilities?