

NOAA archive and access services for Jason-2/3 products

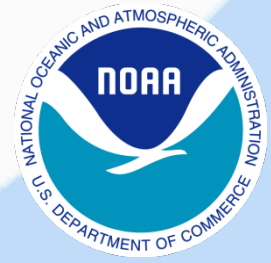
Dr. Deirdre Byrne¹

Dr. Yongsheng Zhang^{1,2}

October 28, 2014

NOAA National Oceanographic Data Center

Cooperative Institute for Climate And Satellites-University of Maryland

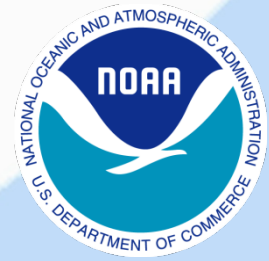


Background

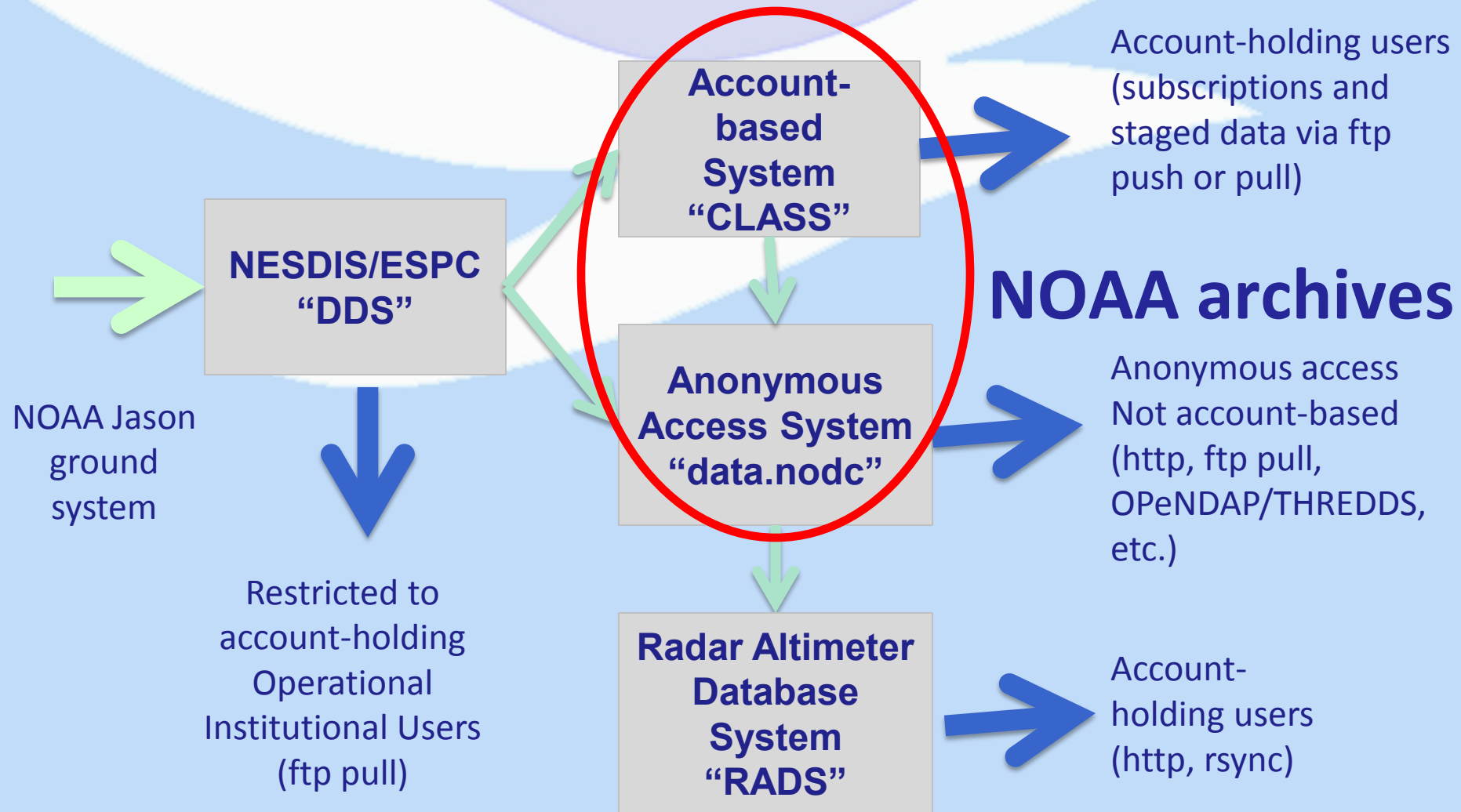
The US National Oceanographic Data Center (NODC) serves as the endpoint for the NOAA Jason Ground System. We provide a number of kinds of user services including long-term archival storage of all mission data, on-line access to the xGDR family of data products, and a helpdesk system. This system is implemented for Jason-2 and services will extend to Jason-3 after launch. In this presentation we will look at:

- Who is using the data
- User needs
- Access methods, volumes, products
- Change over time in data access



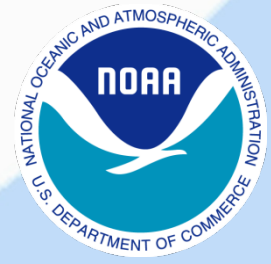


Jason 2/3 Data Flow in NOAA



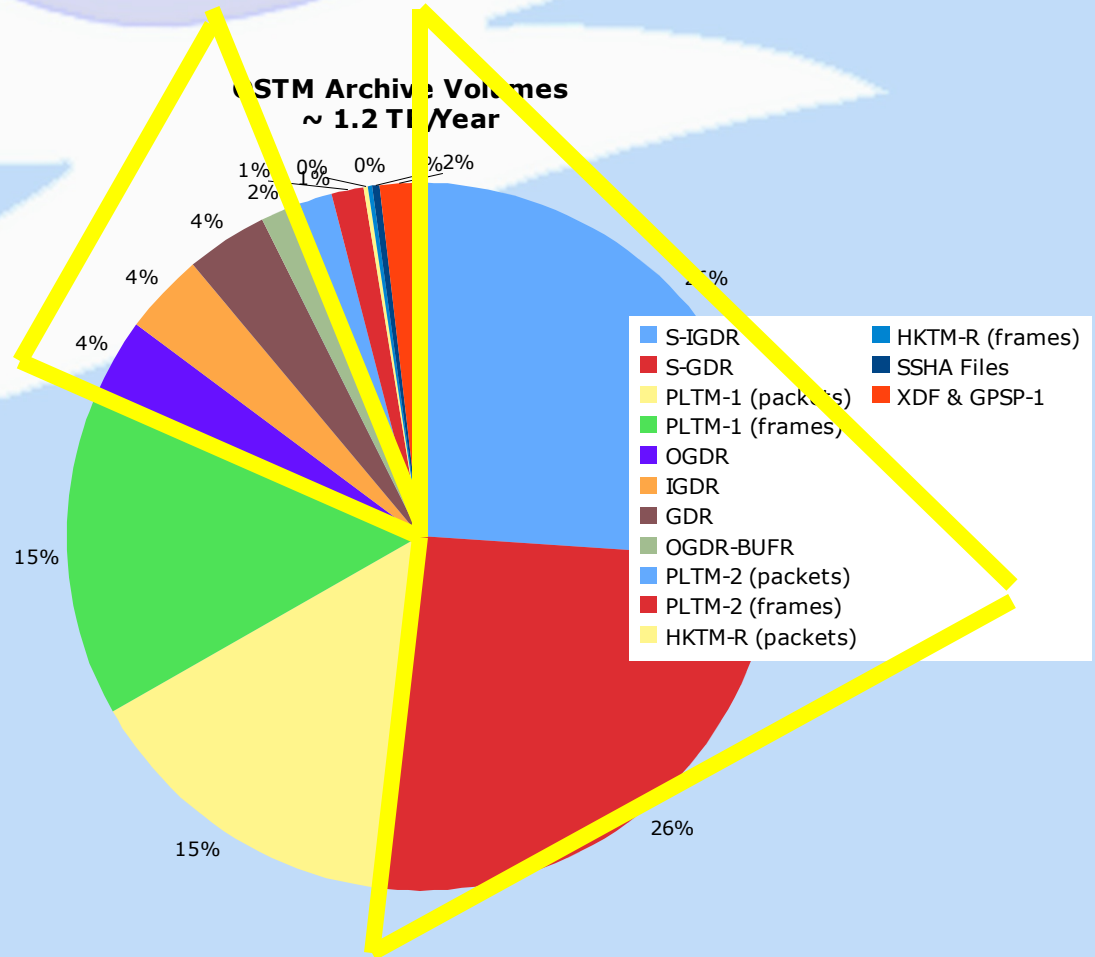


NOAA Jason “archival” distribution systems



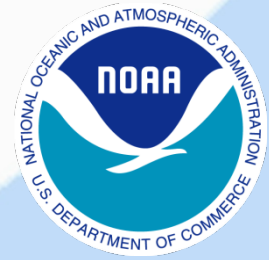
Account-based System
All mission products or 1.2 TB per year.

Anonymous Access System
Level-2 data (xGDR family) amounting to 800GB/year.

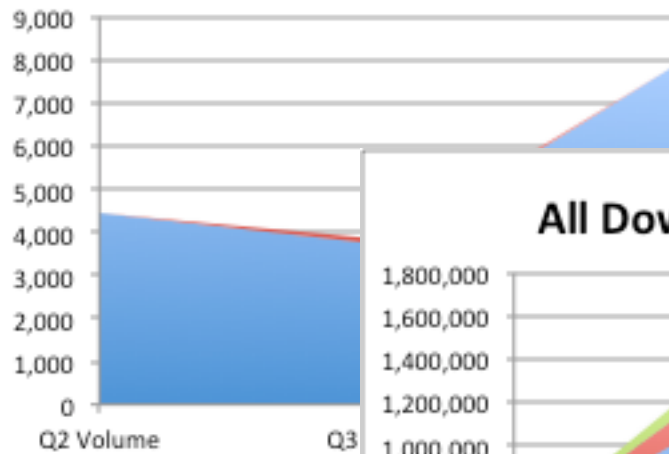




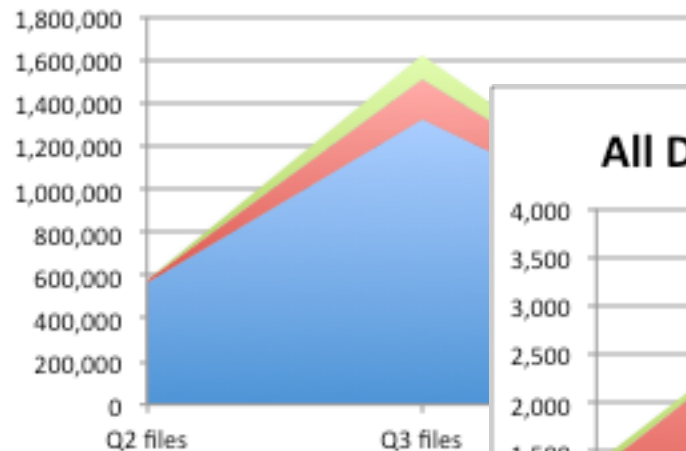
Protocols – 2014 data (Q2:JFM Q3:AMJ Q4:JAS)



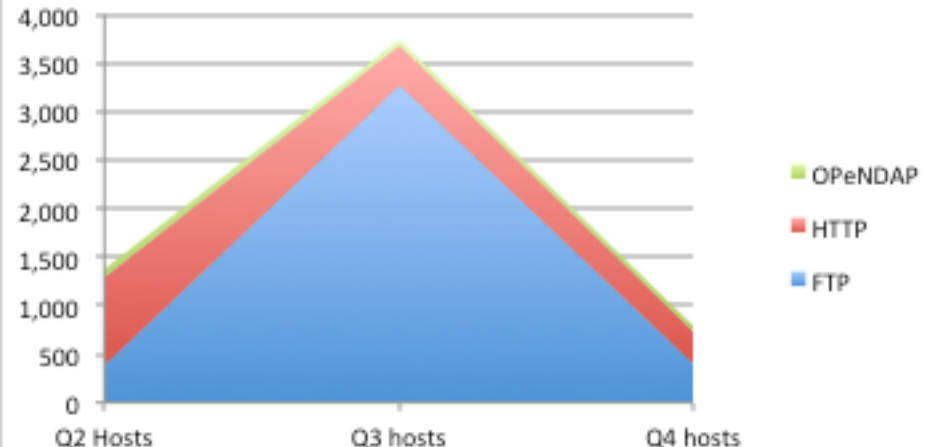
All Downloads - by Volume



All Downloads - by File



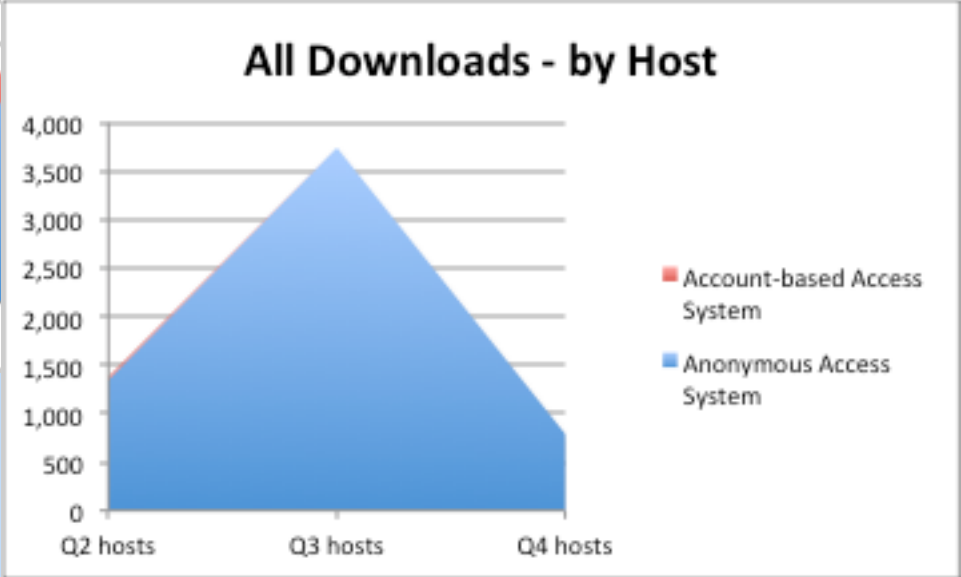
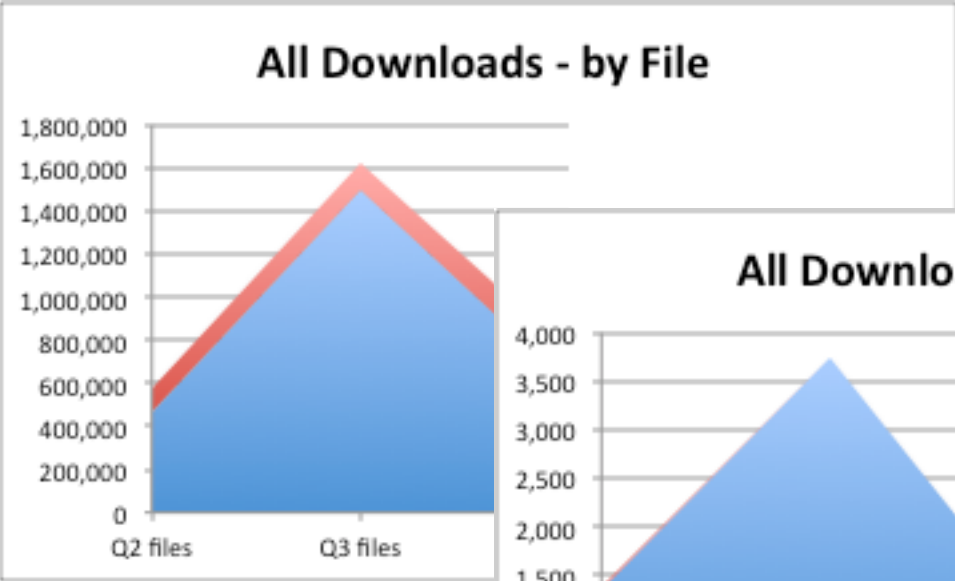
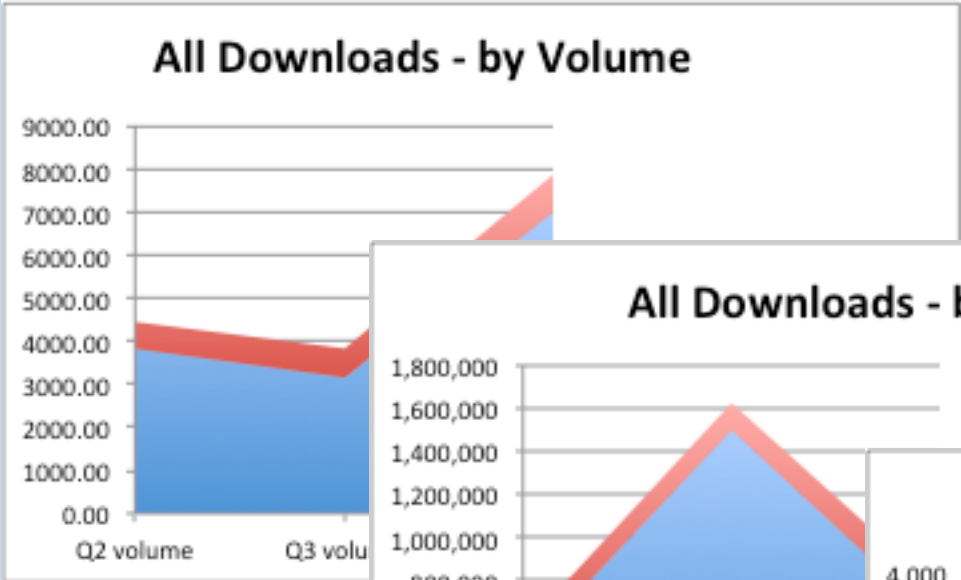
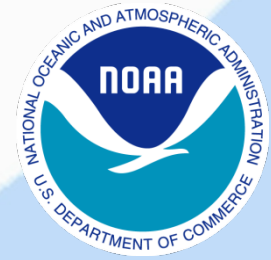
All Downloads - by Host



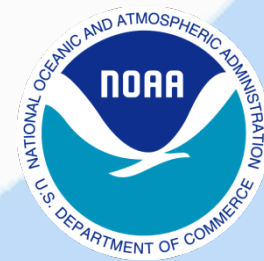
Throughout this talk:
Anonymous access host count = distinct IP addresses. Account-based host count = distinct account names.



Account-based v. Anonymous Access In 2014



- Account-based Access System
- Anonymous Access System



Access by Domain in 2014 – Account-based System

Account-Based System
Q2 Volume 654 GB

- Domain .noaa.gov
- Domain other .gov
- Domain .mil



Account-Based System
Q3 Volume 696 GB

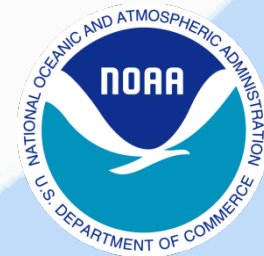
- Domain .noaa.gov
- Domain other .gov



Account-Based System
Q4 Volume 855 GB

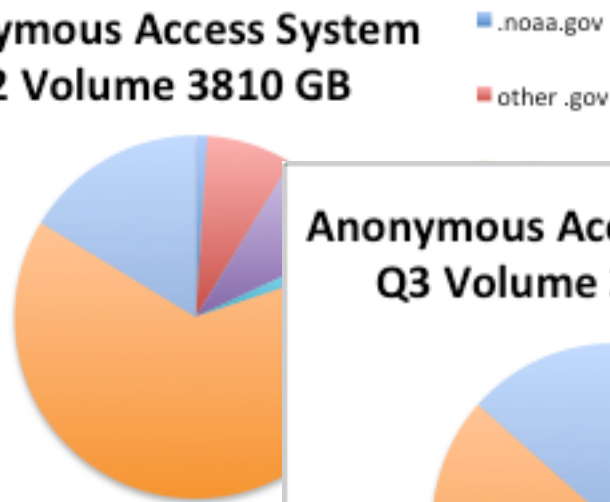
- Domain .noaa.gov
- Domain other .gov
- Domain .mil
- Domain .edu
- Domain .com
- Domain foreign (CA, AU, IT)
- Domain unknown



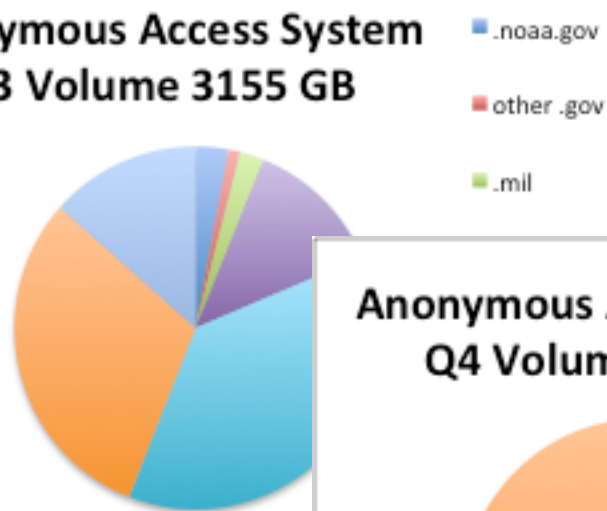


Access by Domain in 2014 – Anonymous Access System

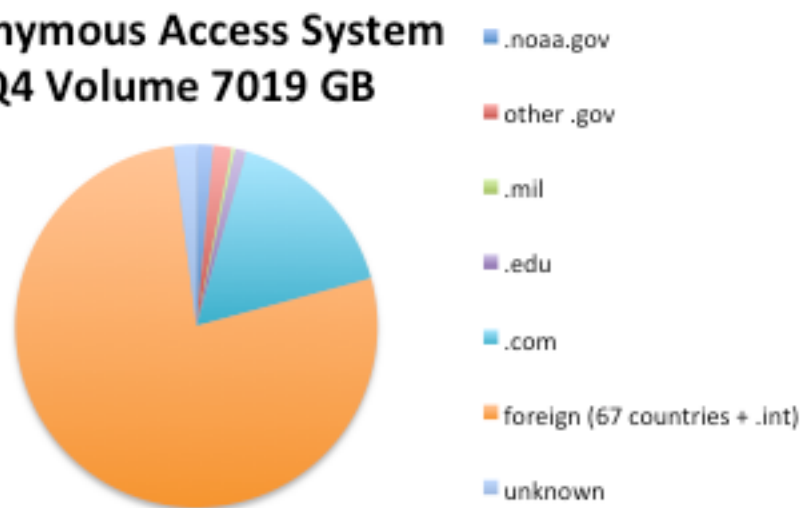
Anonymous Access System
Q2 Volume 3810 GB

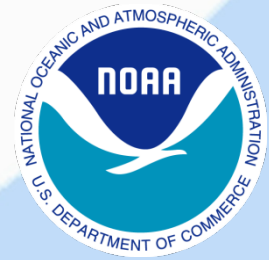


Anonymous Access System
Q3 Volume 3155 GB



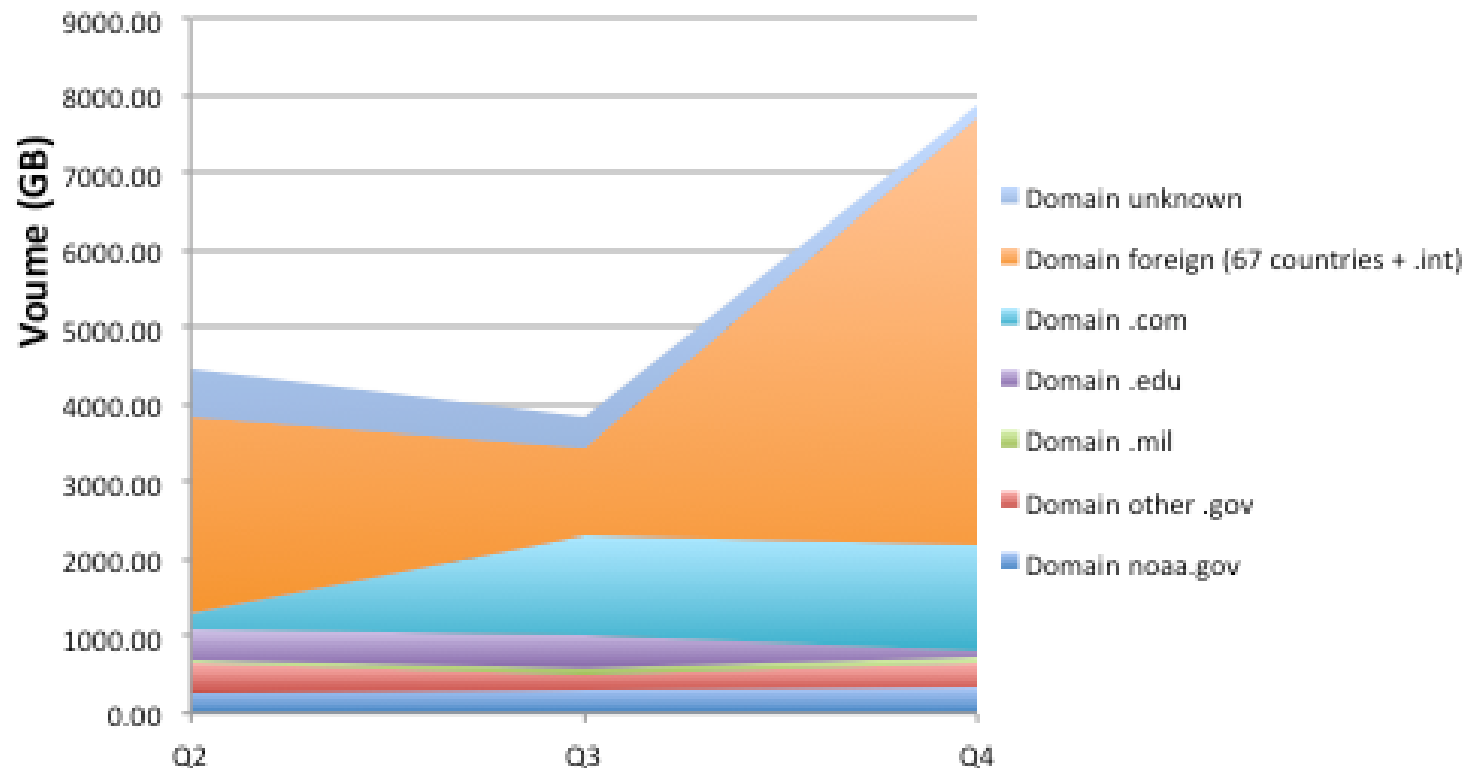
Anonymous Access System
Q4 Volume 7019 GB

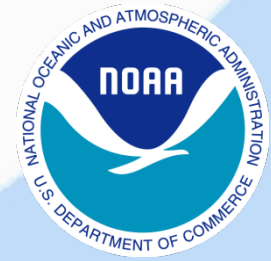




Access by Domain in 2014

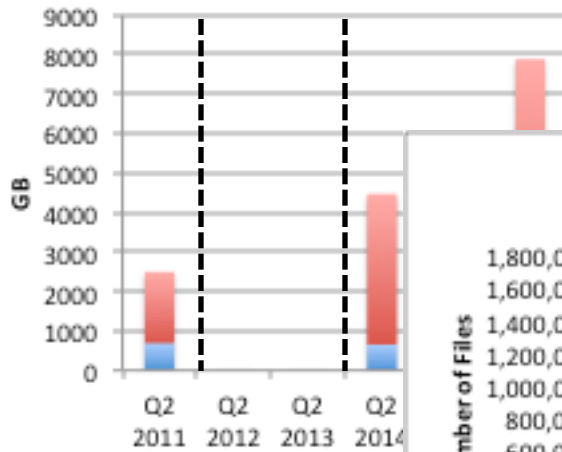
Access by Domain: All Systems



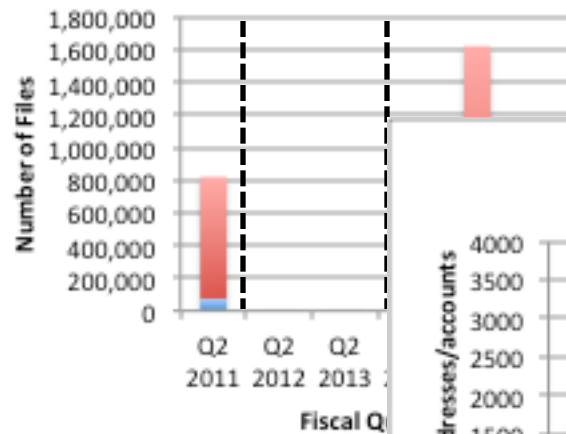


2011 versus 2014

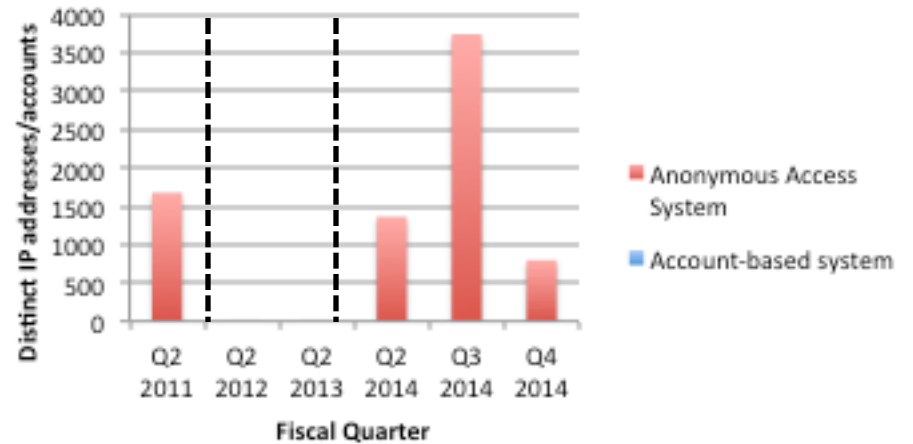
Volume



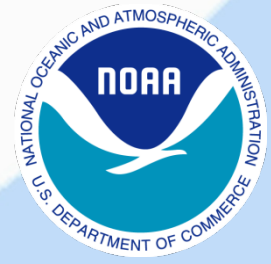
Files



Hosts



- Anonymous Access System
- Account-based system



Jason User Communities Defined

**High volume,
Low
assistance**

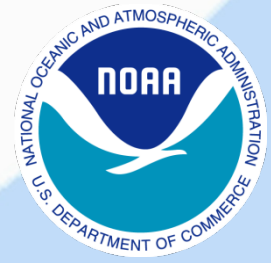
- **Operational Agencies and the Jason-2/3 Science Working Team (SWT):** Level of expertise: Highest. Agencies incorporate altimetry into mesoscale numerical models. SWT members monitor instrument performance, apply calibrations, specify reprocessing requirements and schedule and so on.

**Medium
volume,
medium
assistance**

- **Ocean Surface Topography Science Team (OSTST):** Level of expertise: high. OSTST members are data users but also data producers. They help define the data formats, products and protocols that should be routinely supported.
- **Scientists without subject expertise in altimetry:** most NOAA clients (e.g. NCCOS and NMS), and academics such as ecologists, chemical and biological oceanographers, and some graduate students. Level of expertise: low to intermediate. These users often have a high-level understanding of what altimetry can provide but lack the technical skill or familiarity to manipulate the data. They tend to be primary drivers for routine, value-added products such as gridded fields and derived geostrophic currents.

**Low volume,
high level of
assistance,**

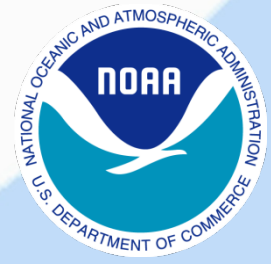
- **Non-specialists** conducting scientific research or planning: undergraduates, geospatial information system (GIS) users. Level of expertise: low. These users are aware of the existence of satellite altimetry and have some idea of the information it might provide, but often need individual help accessing the data or reformulating it into a value-added product such as a time series or georeferenced image.
- **The general public:** Level of expertise: low to non-existent. These users are often seeking general ocean information and have no prior awareness of satellite altimetry. Sample users: individuals and shipping companies wanting wave height statistics.



NOAA Jason Archive User Communities in Practice

Jason-2 data dissemination in brief (estimated from quarterly stats.)	
2011	2014
over 6,500 users	Over 7,500 users
3 million files	4 million files
~7 Tb or 7x the annual volume of new data	~20 Tb or 20x the annual volume of new data

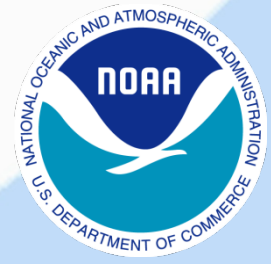
- **Science Working Teams (SWTs) and Ocean Surface Topography Science Team (OSTST):** primarily we hear from this community when service outage or product delivery delay occurs.
- **Scientists without subject expertise in altimetry:** Current examples include foreign graduate students from South America and the Middle East with questions about specific flags/corrections or how to access data in a specific format.
- **Non-specialists:** A significant proportion of user requests are for information/data in GIS-ready format.
- **The general public:** Sample users: individuals wanting wave height statistics. E.g., a wargaming officer from the US Navy who wants sea-state information off the West African coast, or a shipping company wanting SWH along their routes.



Takeaway Points (1 of 2)

- **Ftp** is by far the preferred protocol when choice is made available, accounting for over 95% of data volume. It is also the preferred method used by most (but not all) large-volume users. That said, http accounts for about 15% of the data volume, and OPeNDAP for about 15% of the users.
- **Subscription systems** work well for operational users who have standing requirements. It should be noted that these operational users could just as easily use the open system, many without significant latency but choose not to.
- **OPeNDAP:** shows a small but persistent impact. In at least one month (May 2014) we had over 100,000 files accessed by OPeNDAP (a user from a US academic institution).
- **Anonymous access is spiky** with volume, domain and number of files accessed fluctuating by up to an order of magnitude quarter to quarter and large swings in the dominant categories of user domains.

And finally



Takeaway Points (2 of 2)

- **Anonymous Access** is a success. It is by far the preferred method of access. Any one of the users who downloaded data anonymously was also eligible to create an account with a username and password. But overwhelmingly (99% of) users choose not to do so when given a choice, and that includes some of the highest volume users. Open access has seen a gain in the volumes accessed and number of users. Account-based access has not. **We posit a number of reasons for this pattern.**

In the case of expert users anonymous access is simply more convenient. Many scientists download data only occasionally and then spend a good while analyzing it. Anonymous access obviates the need to recall ones username and password for these accesses. In addition, if a scientist decides to investigate a new product, her workflow is not interrupted by needing to request access to it, justify the need, and then stop work while waiting for permission to access.

For all others - marine scientists, non-specialists and the general public anonymous access is less intimidating. Users can freely download and examine some files and products without having to justify to anyone why they might want to. They are free to make some silly mistakes and get a general sense of what the data look like without having to put their ignorance on display. Meanwhile, non-expert users with a real need are free to contact the User Helpdesk for assistance. By and large these occasional users have one-off needs and are not interested in getting regular updates or newsletters, so the argument of constant contact does not apply.