Insights from 40 years of data services at NCAR's Research Data Archive

Douglas Schuster (schuster@ucar.edu) and Steven Worley (worley@ucar.edu)

720-488-6817

Highlights

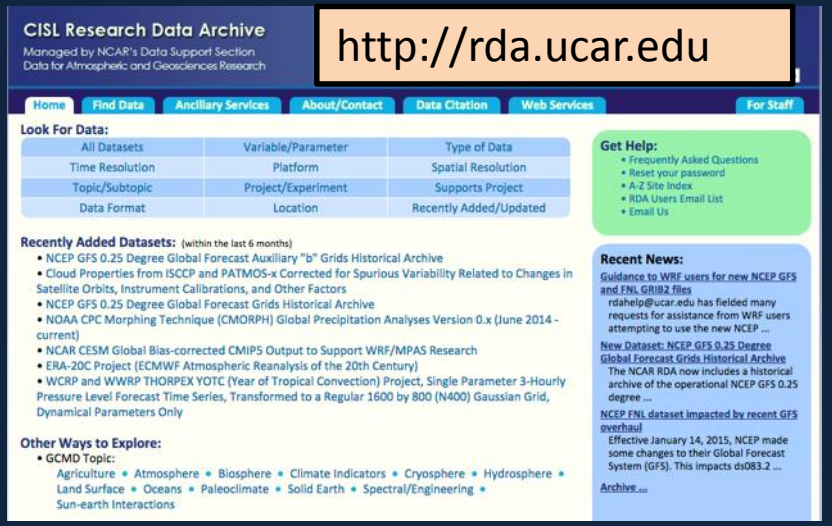
- What is the RDA?
- Evolution of RDA Services
- User Registration Policies
- Future Directions

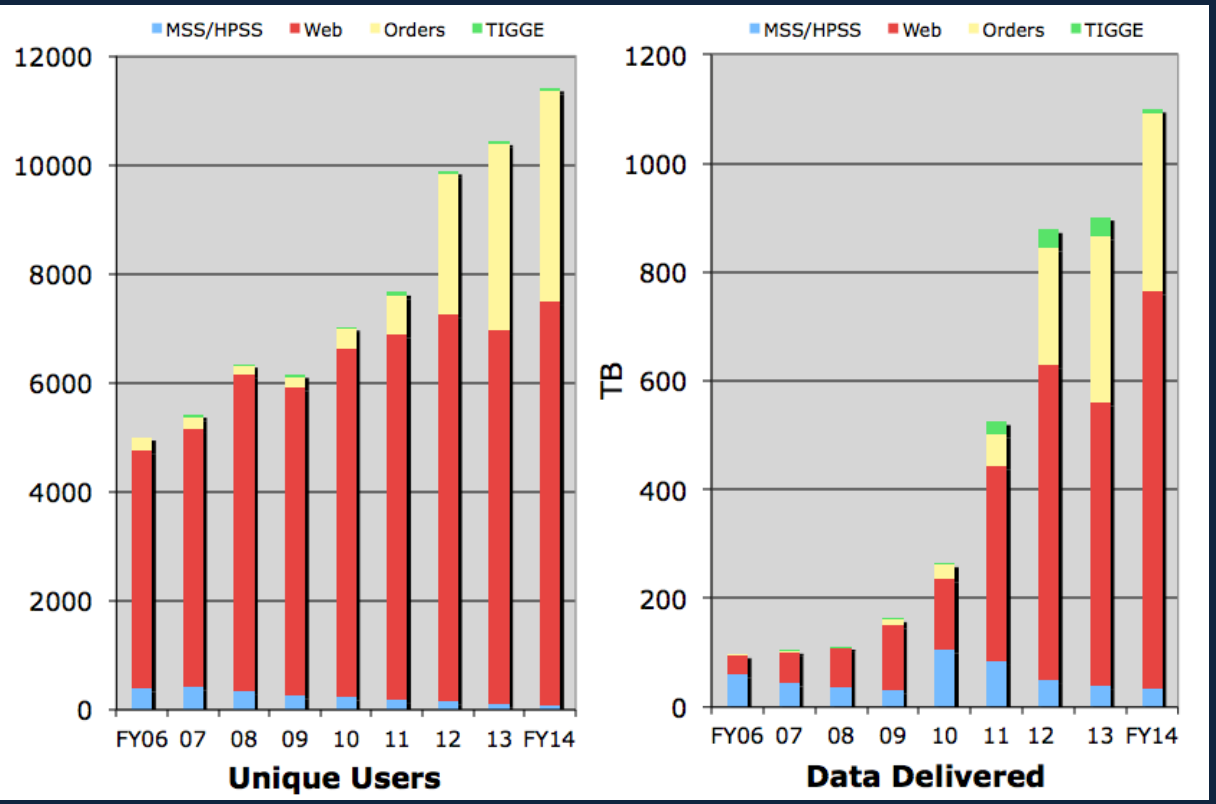
What is the RDA?

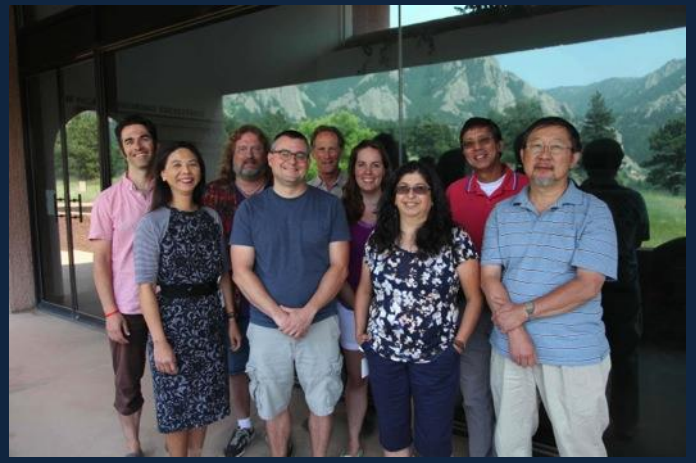
- Purposes
  - Support climate & weather research at NCAR and UCAR Universities
- Collections
  - Ocean & atmosphere observations, analyses, reanalyses, operational NWP outputs
- Basic Metrics
  - Established in 1960s
  - 600+ datasets, 8M files, 1.8 PB
  - +70 datasets growing daily – monthly

What is the RDA?

- Science educated staff
  - Expert consultants and data engineers
- Free and open access







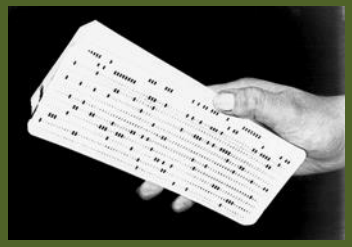
Evolution of RDA services

- General Goals
  - Securely maintain community reference information for a long period
  - Facilitate “easy” discovery of holdings
  - Enable “simple” access to holdings
  - Provide the community with the capability to build upon or reproduce the research of others

Evolution of RDA services

- 1960s to early 1990s –Data Specialist Driven

Data/Metadata Storage



Data Discovery

Data Delivery

Evolution of RDA services

- Mid 1990s to Early 2000s –Data Specialist Driven

Data/Metadata Storage

Data Discovery

Data Delivery

Evolution of RDA services

- Mid 2000s to Present –Metadata Driven

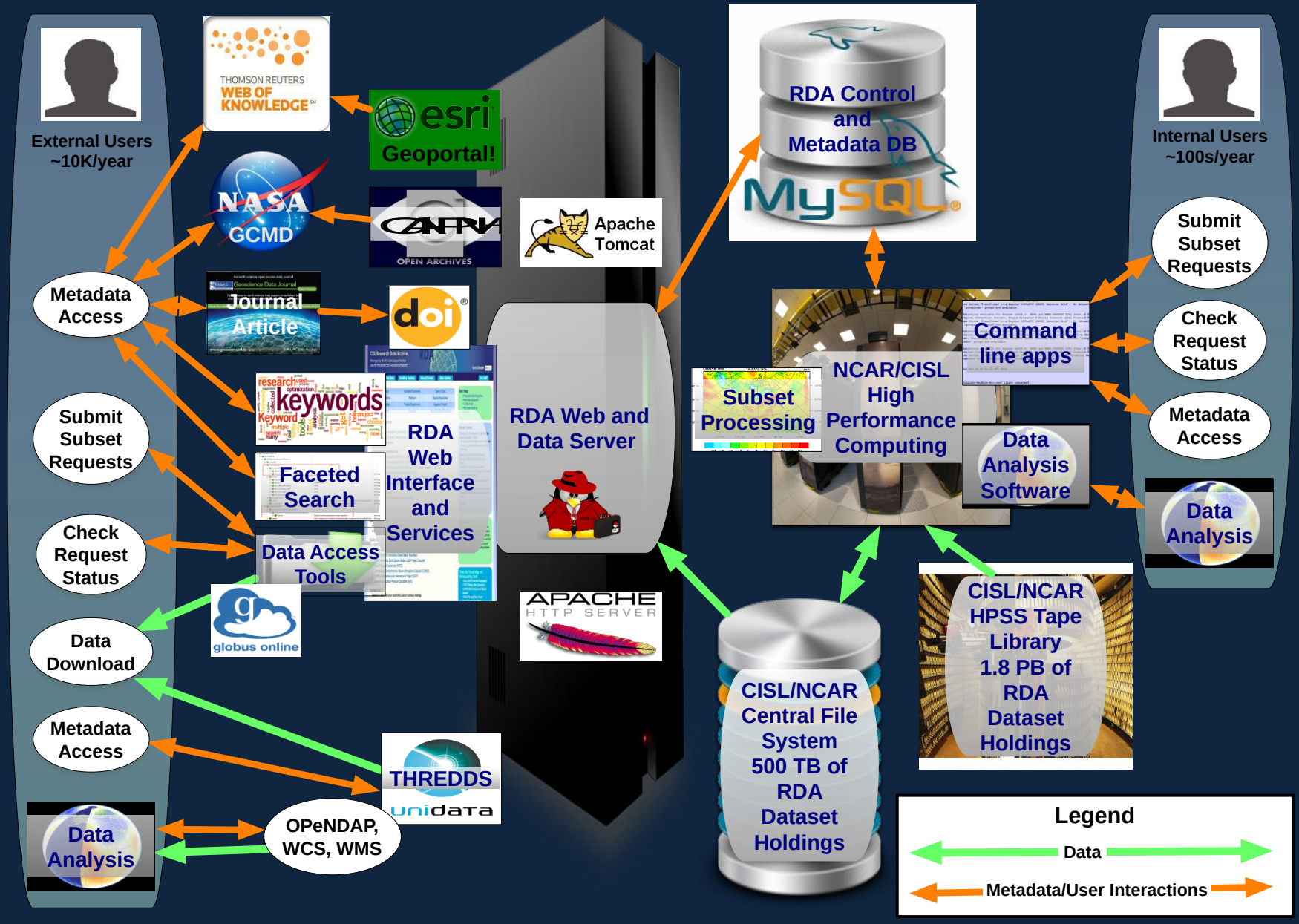
Data/Metadata Storage

Data Discovery

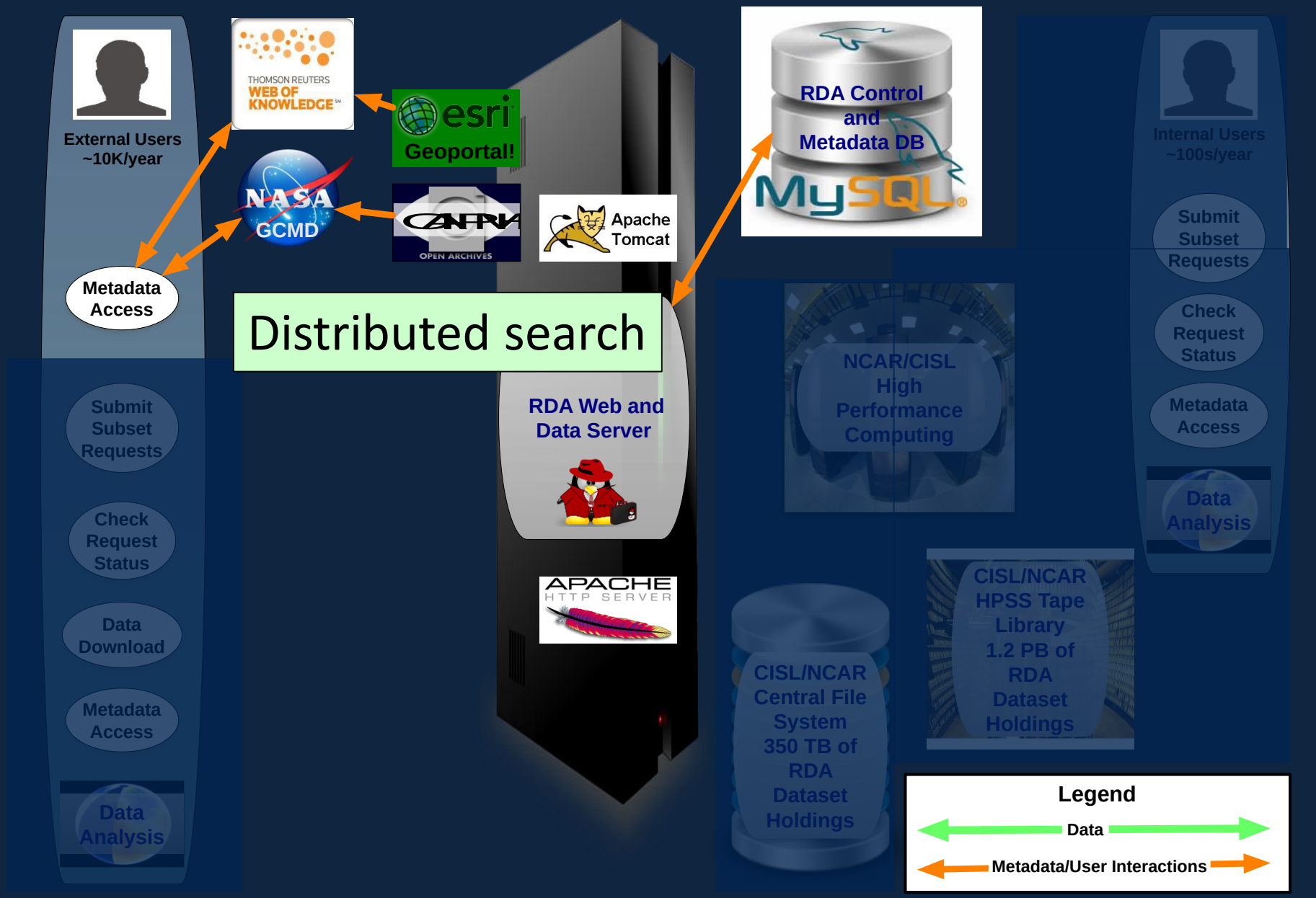
Data Delivery

http://

Current RDA Services



RDA Data Discovery Pathways



RDA Data Discovery Pathways

RDA Data Discovery Pathways

-Server Generated Data Citation

How to Cite This Dataset:

RIS

BibTeX

Compo, G. P., and Coauthors, 2009: NOAA CIRES Twentieth Century Global Reanalysis Version 2. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO. [Available online at http://dx.doi.org/10.5065/D6QR4V37.] Accessed† dd mmm yyyy.

†Please fill in the "Accessed" date with the day, month, and year (e.g. - 5 Aug 2011) you last accessed the data from the RDA.

Bibliographic citation shown in American Meteorological Society (AMS) style

Get a customized data citation

Dataset:

NOAA CIRES Twentieth Century Global Reanalysis Version 2 (ds131.1)

Your Access History for July 2013:

Choose a day to get a citation for this dataset. You will also see more details about your downloads on that day, which will help you verify that this is the citation you want.

July 2013						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

For Data Accessed on 2013-07-03:

Dataset Citation: RIS

Compo, G. P., et al. 2009. *NOAA CIRES Twentieth Century Global Reanalysis Version 2*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. http://dx.doi.org/10.5065/D6QR4V37. Accessed 3 Jul 2013.

Bibliographic citation shown in Federation of Earth Science Information Partners (ESIP) style

Data Access Detail:

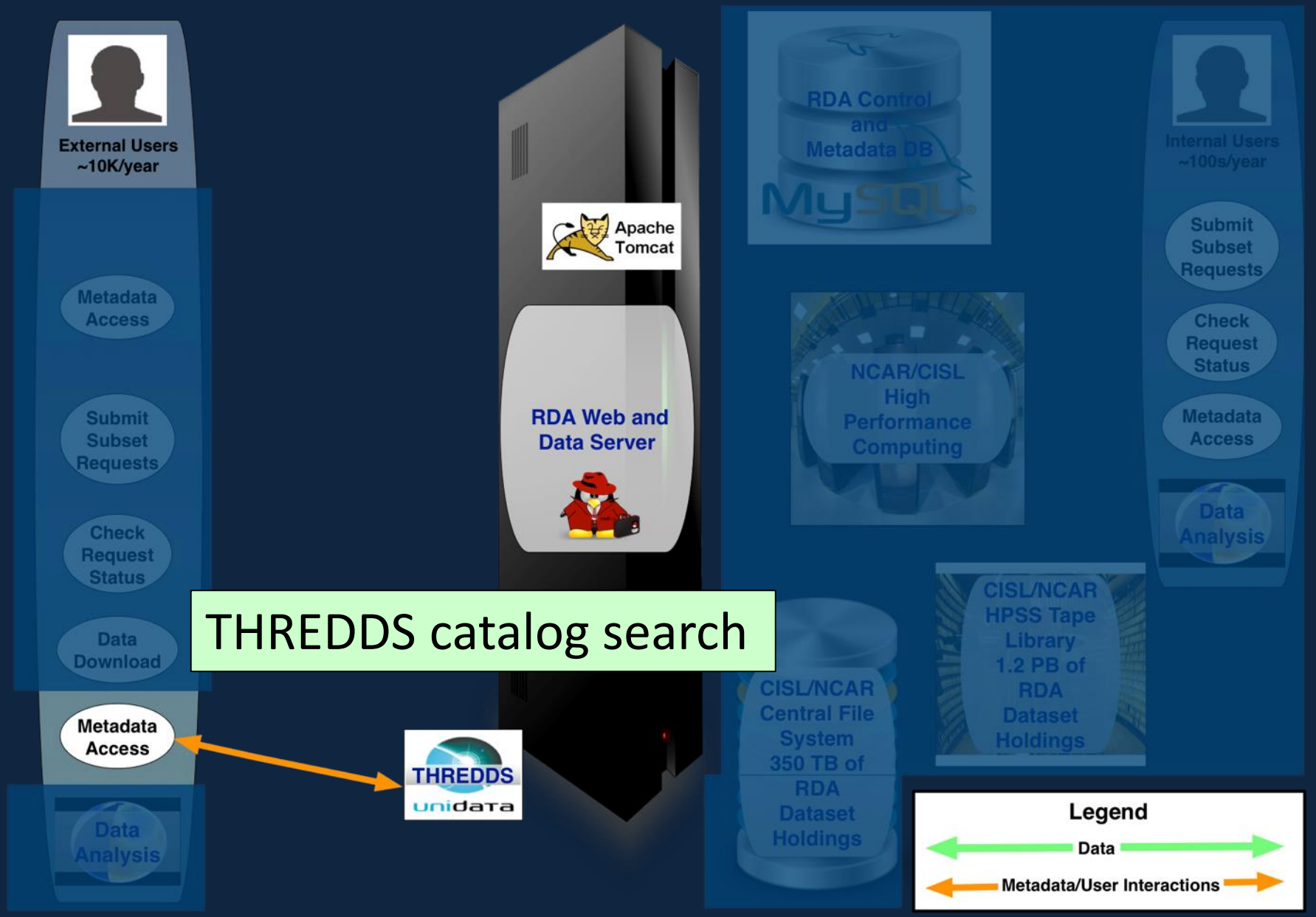
1 subset request:

- 18 files, 3.44 MB

```
Date Limits          :  1975-06-15 00:00 to 1975-06-30 12:00
Parameter            :  TMP
Level Type           :  HGTI
```

RDA Data Discovery Pathways

RDA Data Discovery Pathways



THREDDS catalog search

RDA Data Discovery Pathways

-Advise User of Best Dataset Collection(s) to Support Needs

NCAR UCAR | ClimateDataGuide

inform • compare • discover

CLIMATE DATA | ANALYSIS TOOLS | MODEL EVALUATION | EXPERT CONTRIBUTORS | ABOUT | Site-wide Search >>

Data Discovery Guided by Experts > >

Search and access 174 data sets covering the Atmosphere, Ocean, Land and more. Explore climate indices, reanalyses and satellite data and understand their application to climate model metrics. This is the only data portal that combines data discovery, metadata, figures and world-class expertise on the strengths, limitations and applications of climate data. Discover it now.

See data pages with guidance from these experts:

Ricciardulli, Lucrezia | Meiyappan, Prasanth | Vicente-Serrano, Sergio M. | Huang, Boyin | Kay, Jennifer | von Schuckmann, Karina

Dai, Aiguo | Keyantash, John | Meier, Walter | Kent, Elizabeth | Hassler, Birgit | Tian, Baijun

Data Set Overviews > >

Compare the attributes, strengths and limitations of multiple data sets.

Atmospheric Reanalysis: Overview & Comparison Tables | Global Temperature Data Sets: Overview & Comparison Table | Precipitation Data Sets: Overview & Comparison table | Sea Ice Concentration data: Overview, Comparison table and graphs

JOIN US

Multiply the impact of your work by announcing new data sets and sharing your knowledge of the strengths, limitations and applications of particular data sets.

Ways to make an impact

Become a registered user of this site

Contribute a data set & assessment

Post a comment to any data set page

SHARE CLIMATE DATA GUIDE

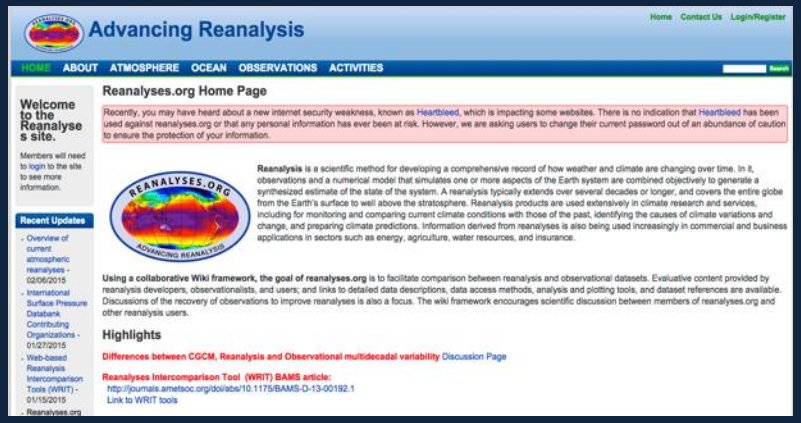
FOLLOW US FOR UPDATES

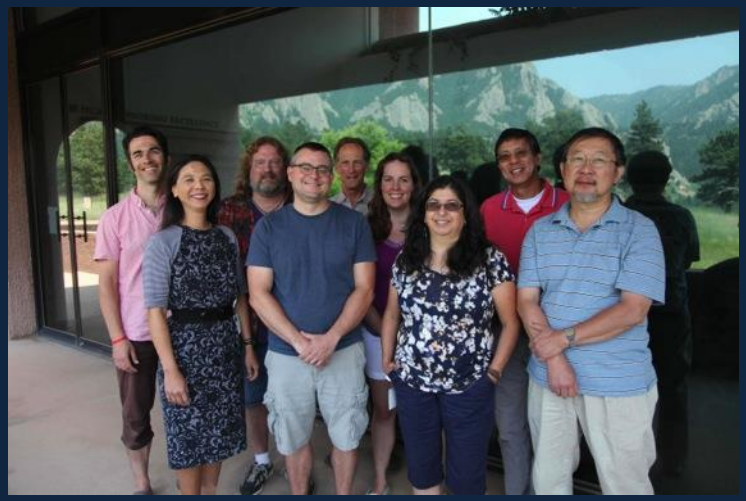
Follow us on Twitter for content updates & relevant data news.

Tweets | Follow

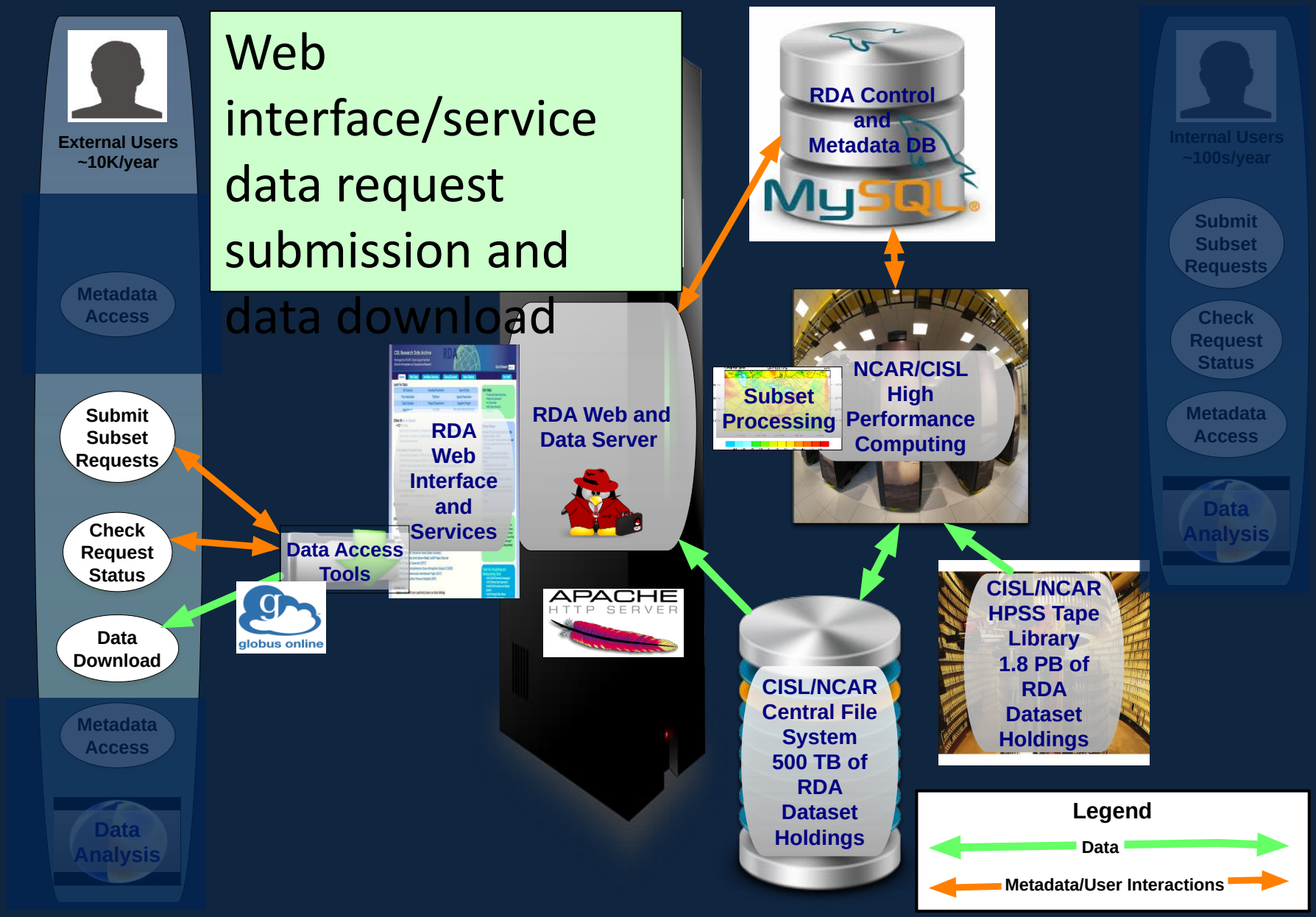
AGU's Eos @AGU_Eos 12 Dec

How well do climate models simulate present-day climate? A new diagnostic tool may give insight: ow.ly/FNdlG

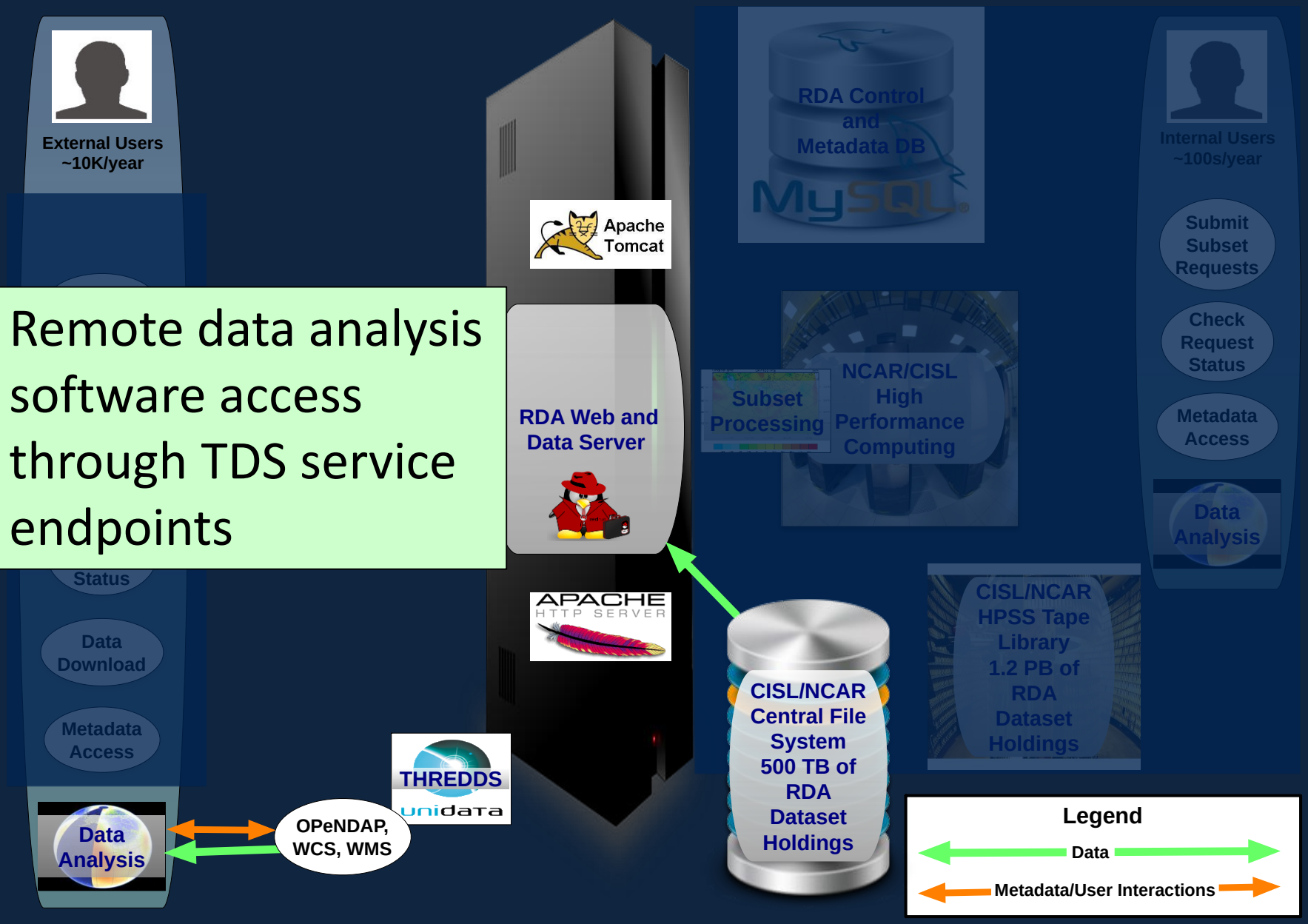




RDA Data Access Pathways



RDA Data Access Pathways

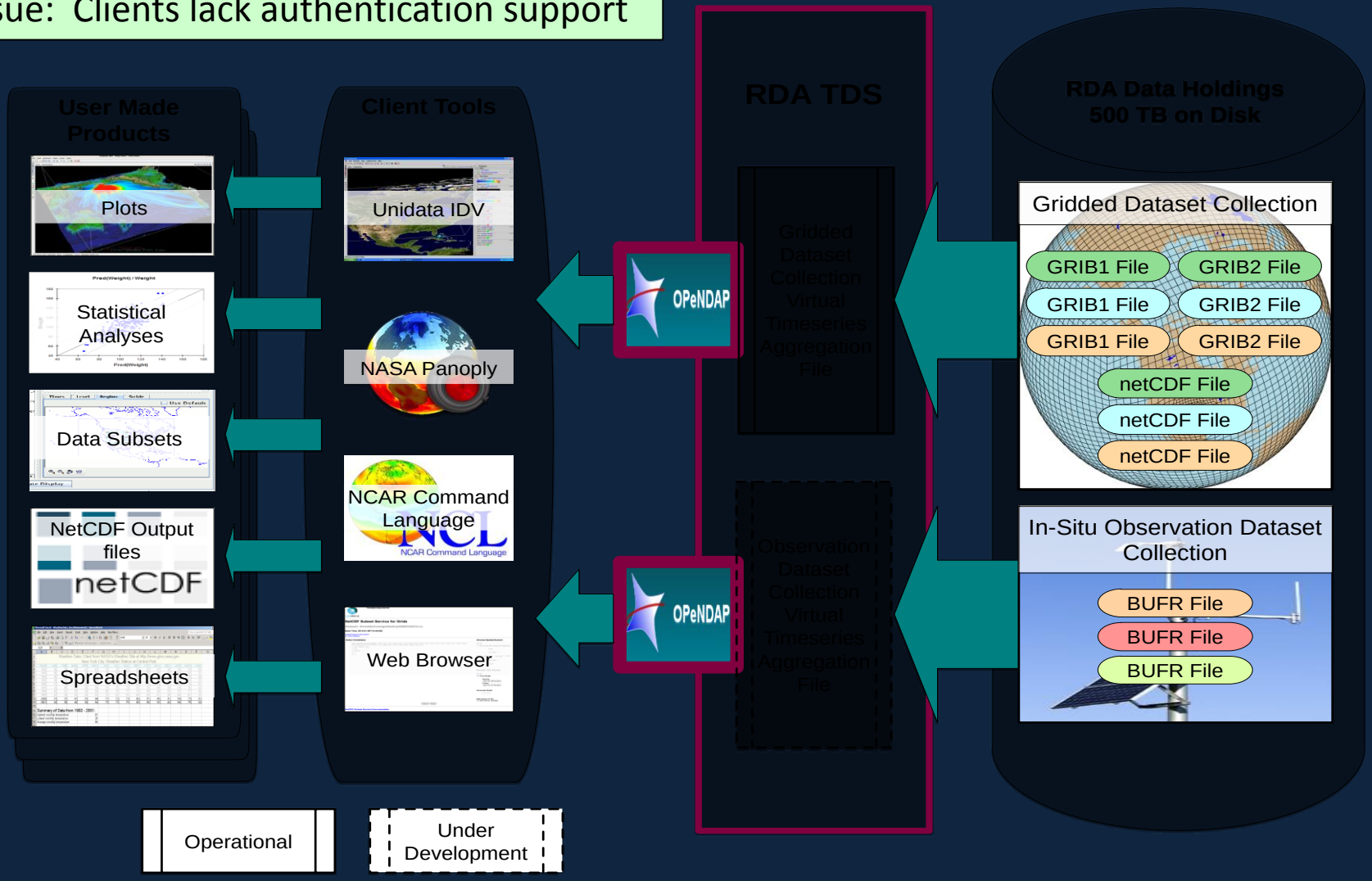


Remote data analysis software access through TDS service endpoints

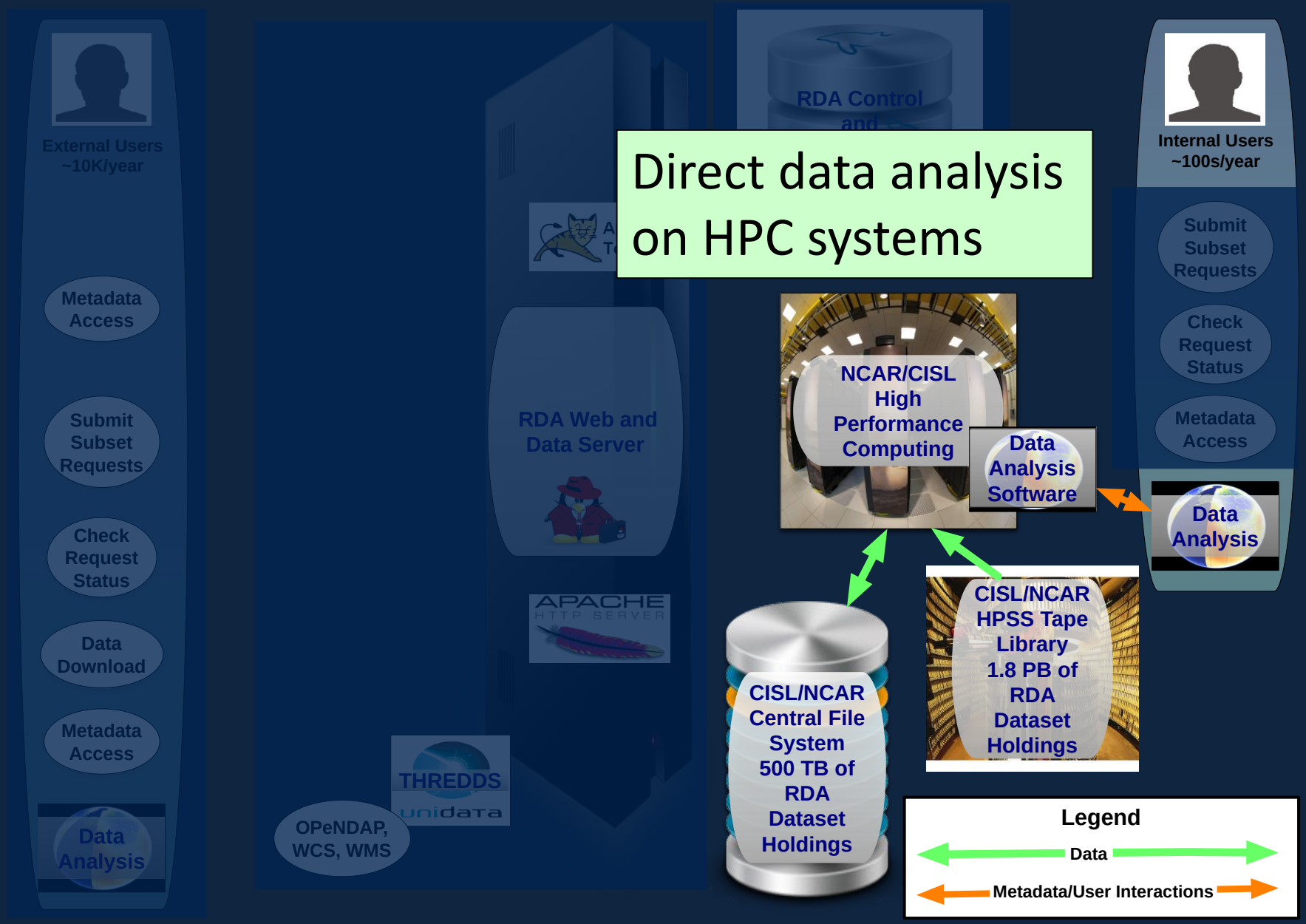
RDA Data Access Pathways

-Unidata TDS Dataset Collection Virtual Aggregations

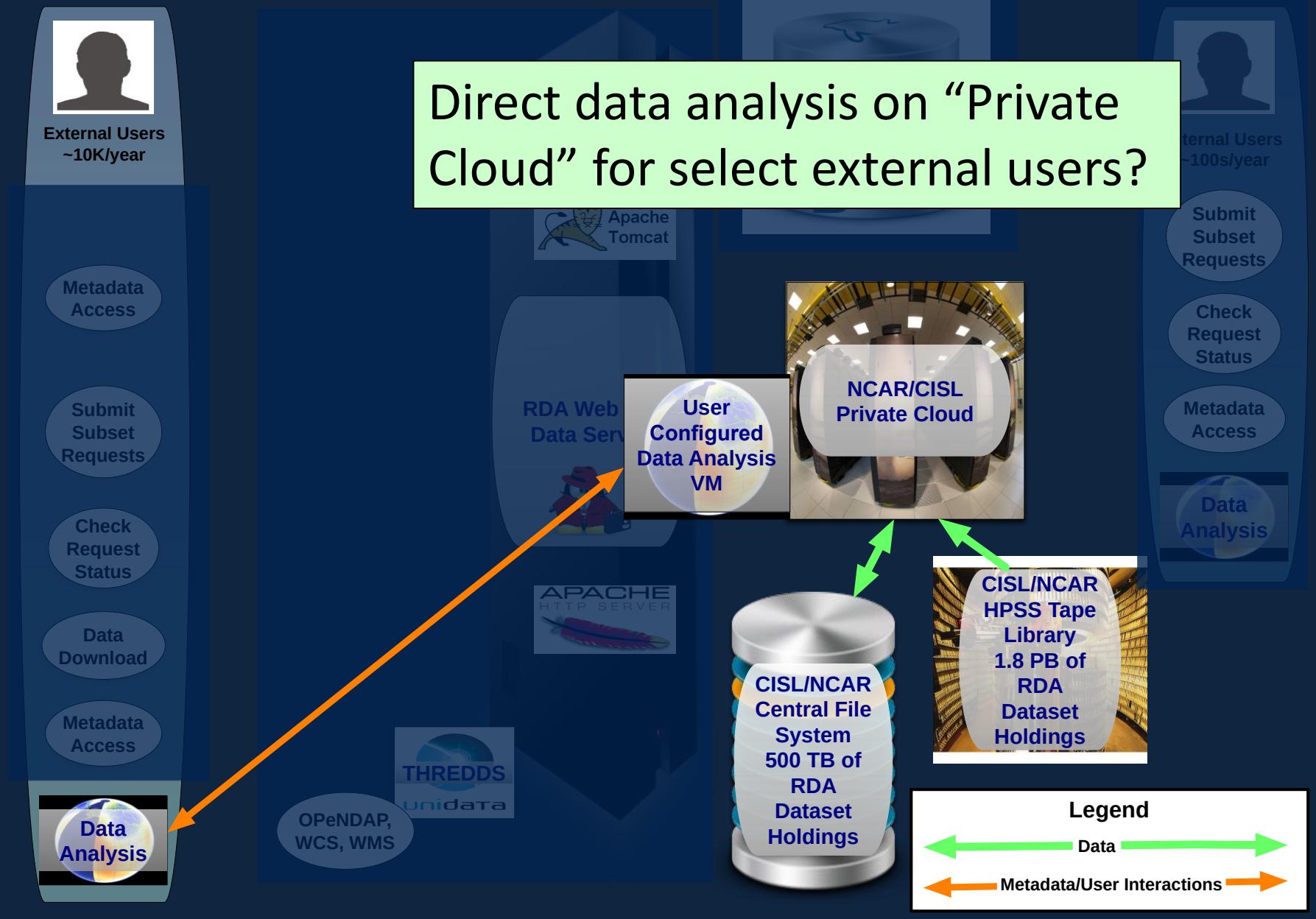
Issue: Clients lack authentication support



RDA Data Access Pathways



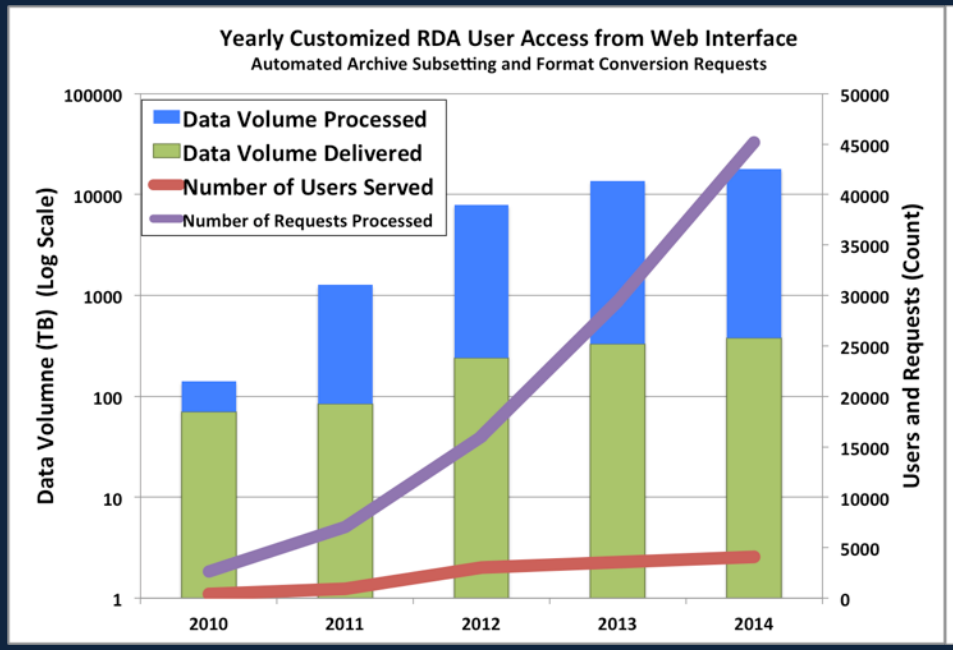
Future RDA Data Access Pathways

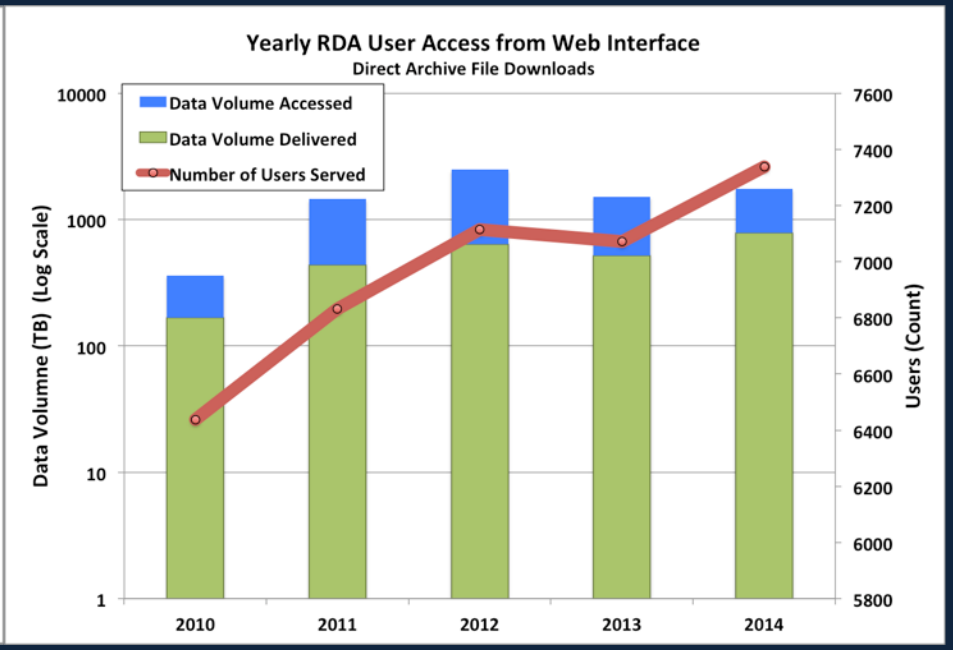


Format management and usage

- Archive files stored in native data formats
  - GRIB-1, GRIB-2, netCDF, HDF, BUFR, various binary and ASCII formats
  - Archive may be restructured, depending on end-user research needs
    - Time-series for single variable, or synoptic-time arrangement
    - Climate research versus weather research preference!
  - Prefer on-demand format conversion over maintaining datasets in multiple formats
    - Offer format transformation as an output option
  - Leverage protocols (e.g. OPeNDAP) that make data-format transparent to the end user

Current Services Overview –Usage metrics





In 2014

- 17+ PB (virtual) processing
- 4000 users received data subsetting and format conversion services
- 45K requests were processed and 380 TB of data delivered
- 7300 users downloaded 750TB of data from archive files
- Metrics are now being collected to determine user data discovery pathways and TDS use

Policies (User Registration, Access/Use, Licensing, Limits, T&Cs)

- User Profile – provided during registration (RDA only)
  - Email address
  - First and Last Name
  - Organization
  - Country
  - Agree to UCAR Corporate Terms of Use
  - Special Authorizations (access to restricted datasets)
- Implement requirements per third party policy agreements
  - Enable RDA to hold and serve unique research products
- Analyze metrics as a basis for scientific impact assessment and economic justification
- Free and open access to all archive holdings
  - We reserve the right to request payment – Why?

Policies (User Registration, Access/Use, Licensing, Limits, T&Cs)

- Future – Might instantiate access control to services in tiered levels – dependent on resource strain.
  - Highest priority for NCAR and NSF supported education and research
  - Broadly support all US users
  - Use unused capacity for world-wide service

Future Directions

- Improve data access
  - General GLOBUS connect access for all RDA holdings
  - Expand OPeNDAP and OGC based access
  - ESGF – integration of RDA archives with climate model data
  - Data brokering – e.g. distributed satellite data matchup service with marine surface observations (ICOADS)
- Improve data discovery
  - More data journal articles and RDA DOIs
  - Additional metadata standards available through harvestable endpoints – ISO 19115
  - Use web standards (e.g. defined by Schema.org) to support discoverability by commercial search engines
- Support “single-sign-on” identity management
- Tutorial videos to enhance consulting with user forum (,YouTube, edx.org)
- More hosted data processing
  - Improved multi-node/MPI data preparation
    - Sub-setting, format conversion, customized file structures
  - Private “tiered” cloud structure for non-HPC users
    - Data analysis (analytics, regridding, user code execution environments)
- Commercial partnerships –public cloud
  - Focus on new communities – general science interested public, more visualization,

Questions?

http://rda.ucar.edu

Doug Schuster
schuster@ucar.edu