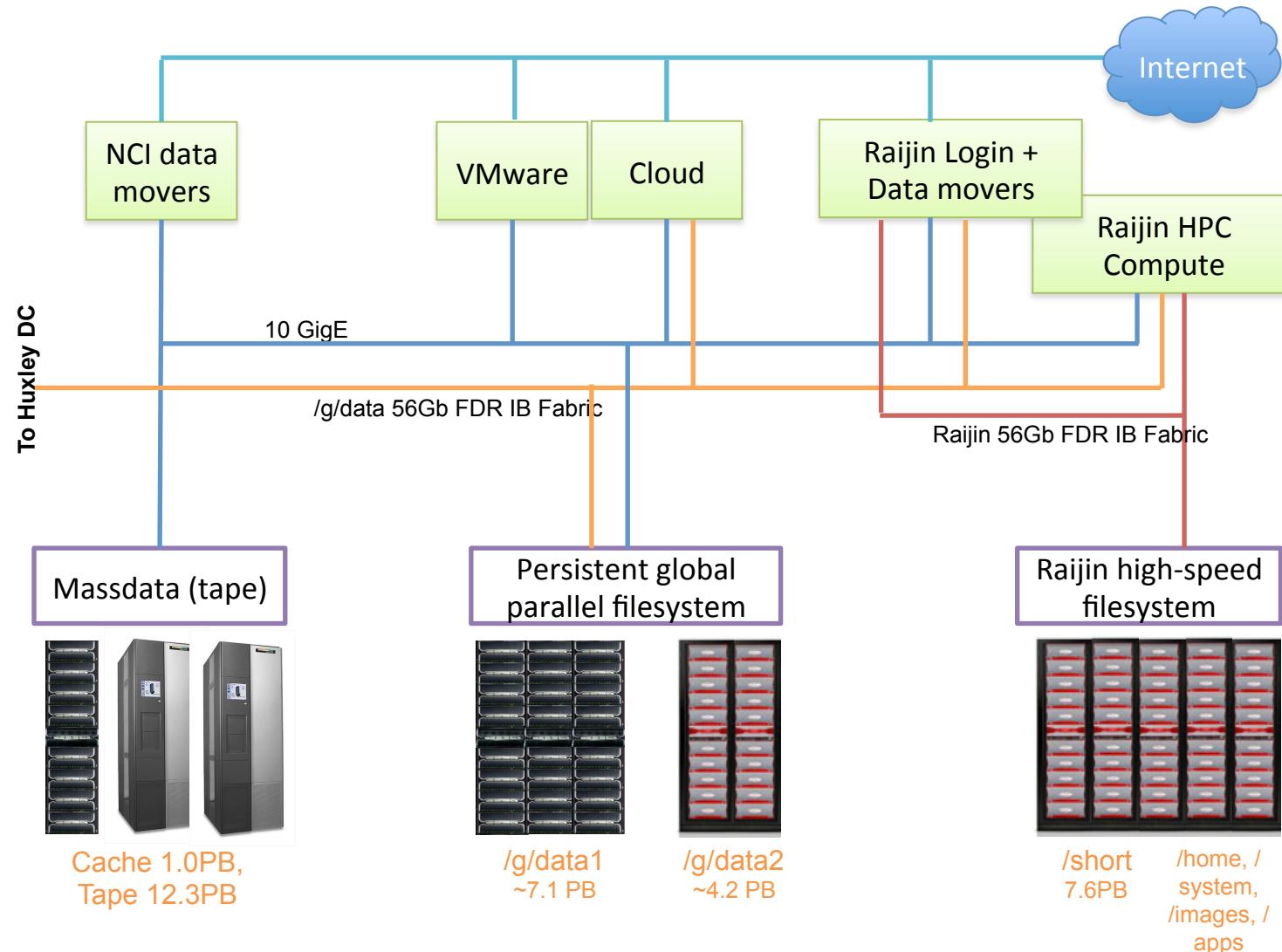




NATIONAL COMPUTATIONAL INFRASTRUCTURE

# NCI ESGF update

Ben Evans  
9 December 2014



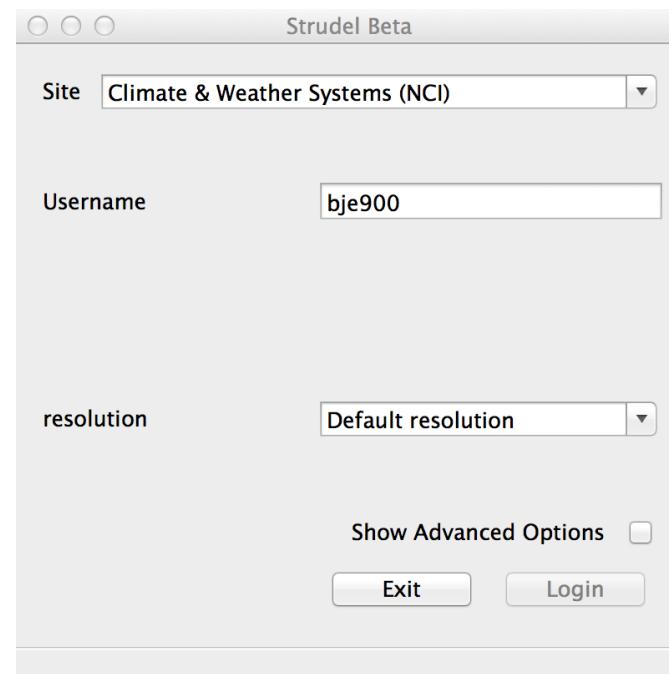


## VDI – Visual Display Interface

Secure connection to per user dedicated nodes



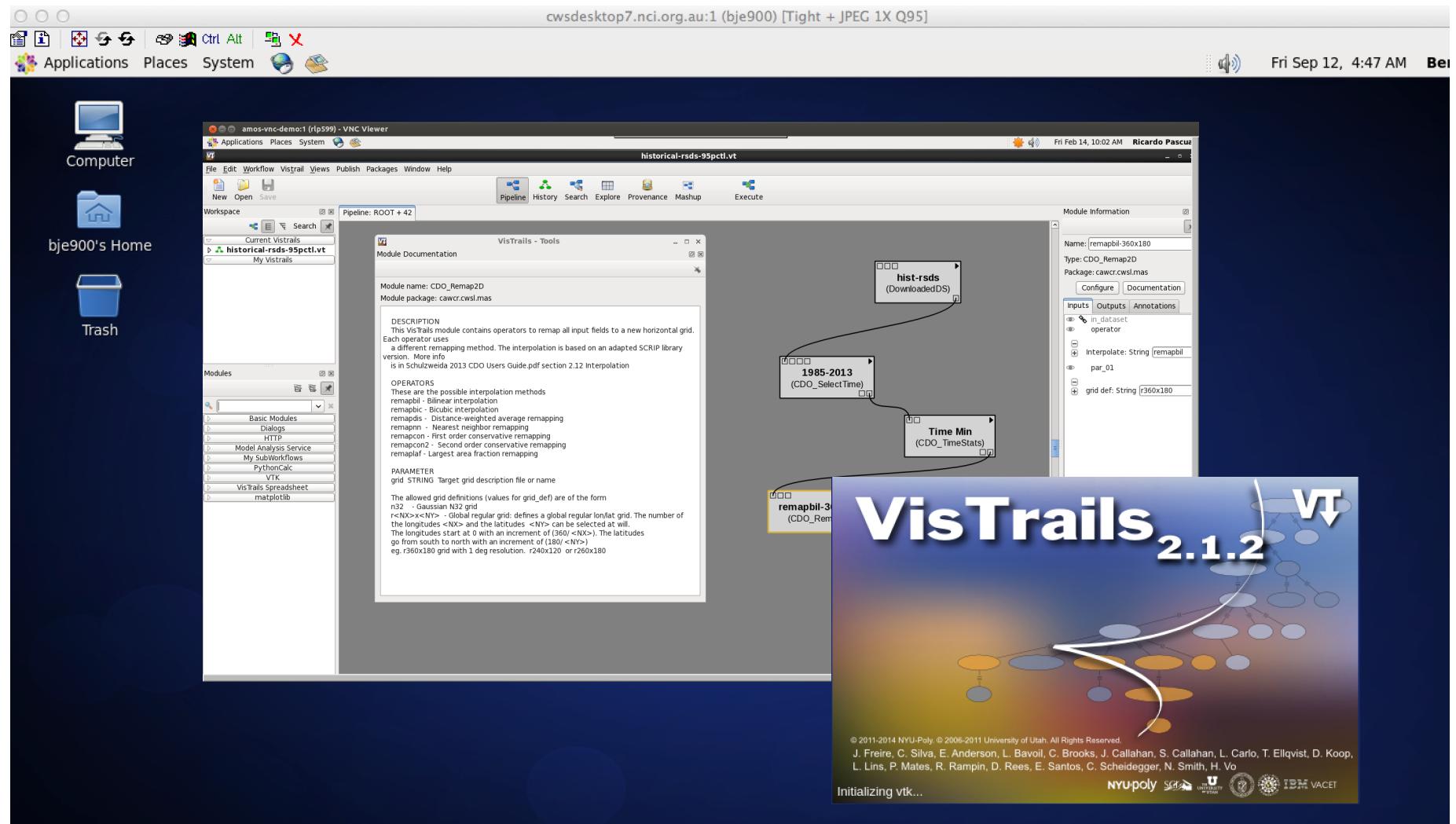
**Strudel**



© National Computational  
Infrastructure 2014

“NCI ESG update”  
Ben Evans, 9 December, 2014

[nci.org.au](http://nci.org.au)



© National Computational  
Infrastructure 2014

"NCI ESG update"  
Ben Evans, 9 December, 2014

nci.org.au

## Installation

- ESGF installation in the OpenStack cloud. Mix of puppet, esg-install, snapshot, config
  - Production:
    - IDP+portal
    - 4x international cached supernodes (1=PCMDI, 1=replicate NCI publisher)
    - 1x private publisher index node
    - 1x data (TDS) + LAS server (why colocated?)
    - 2x gridftp + globus online (internal ESGF private node replication)
  - Test Federation:
    - Similar setup
  - Development:
    - Severed from Federation
- perfSONAR (ICNWG)
- UV-CDAT installed (integrated in the VDI – 10s of nodes, and growth).

## Separate (larger) installation for data services for NCI

- Catalogue based on geonetwork and other indexing
- Open access, but moving more to restricted with local LDAP, Shib, and OAUTH2
- Data services: DAP (TDS, Hyrax), OGC (WMS, WCS, WFS, WPS), plus micro-services, other re-usable bits, and Virtual Laboratories



## Operational Status

- (Re)publishing Australian data
- Data will have LAS-enabled
- December to properly rejoin the federation, including NCI IDP and sharding

## Underway

- Ensuring local cache of data is up-to-date (gridftp's with Livermore)
- Items to be debugged/understood/resolved at this meeting
  - LAS to find our local data (index seems to be aware, can't search metadata)
  - Group permissions management and authorisation

## User Support and Development (ESGF is in scope, but not the largest)

- Increased investment and focus on user support and operation – 3 EFT (2 new)
- Virtual Laboratory staff (Development) – 4 EFT (in place)
- Research Data Collections – 3 EFT (1 new)
- High Performance Data – 3 EFT (2 new)
- And coordinating outcomes of other funding/projects

## ESGF Software Issues

- ESGF Installation and architecture:
  - install script is not scriptable.
  - needs both **cluster and cloud** awareness
  - configuration of local security vs federation needs to be separated
  - points of inter-dependency on federation needs to be clarified
- LAS status across the federation is poor
- Management of gridftp as standard publish point is not clear (needed at all?)

## Data issues to be understood

- What data can be under ESGF (and usefully managed)?
- Republishing of replicated data
  - ensures we are up-to-date on version and clear to users of the ESGF
  - Can better use local resources (eg LAS)
- DOI minting
  - Local DOI minting will occur.
  - Good processes needed for DOI records updates with replicas?
  - Agreed rules for model data
  - Consideration of long tail for user analyses



## Future:

- Are we really prepared for CMIP6, CMIP users, Satellite Earth Observation, other
  - Add more well-known satellite data collections: Sentinel, Himawari, MODIS
  - Add more well-known re-analysis products
  - More data formats and Datacubes for climate, satellite obs, elevation, bathy
- ESGF architecture
  - Need a reference framework plus trusted software model
  - Portals are useful, but focus needs to rich services and away from datacarts
  - Move to agreed standard services and possibly microservices
    - Eg ESRI services in the cloud
  - Emerging services for compute and data: WPS and server-side processing.
  - More consideration of handling of data licenses
- Identity management and authorisation issues
  - model that leverages national nodes and generic SSO (ie OAUTH2)
  - Programmatic access with authentication/authorisation integrated.
- Citation
  - Add citation of data services and software
  - DOI/Unique identifiers: both citing the data collections and output publication
- Vocabulary and Provenance Services

