



VIRTUAL ASTRONOMICAL OBSERVATORY

Remote Storage

Matthew J. Graham, Caltech



The VAO is operated by the VAO, LLC.



Use cases

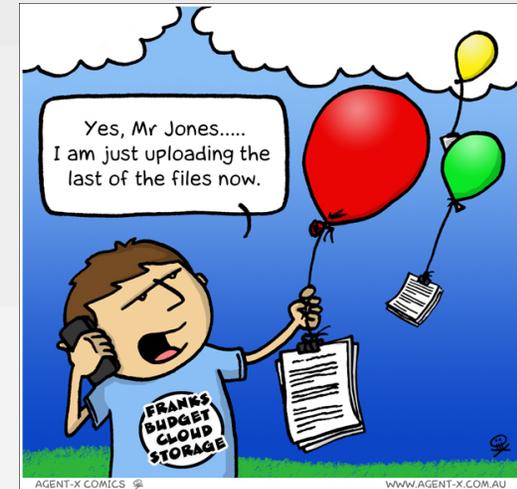
- Traditional astronomy (1990 - 2000):
 - I have data somewhere else and I want to retrieve it to a local store (travelling to Baltimore or Garching to reduce HST data)
- Modern astronomy (2000 - 2010):
 - I have data that is stored locally and I want to upload it somewhere (to a crossmatch service such as OpenSkyQuery)
- New astronomy (2010 -):
 - I have data somewhere and need to work with it...and really never want to bring it anywhere near my local store





Aspects of cloud storage - I

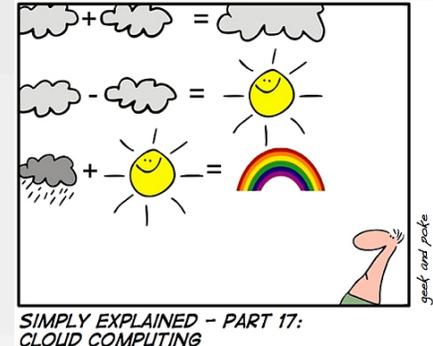
- Global
 - Filestore
 - Database
- Features
 - Nested containers
 - Appending
 - Locking
 - Permissions
 - Transcoding
- Usability
 - Specific client: usable from Windows, Linux or Mac
 - Browser-based control panel
 - Programmatic via API
- Access
 - Available as a network drive (with/without third-party software (WebDAV))
 - Available via physical media
 - Content delivery network (closest copy)
 - Peer-to-peer (BitTorrent)





Aspects of cloud storage - II

- Security
 - Authentication
 - Encryption: service provider cannot look at data
 - Access URLs with time-bounded validity
- Cost
 - Different payment models: per GB, per server, per month
- Others
 - Data duplication for redundancy (data deduplication)
 - Predictive caching
 - Data compression and bandwidth throttling to reduce upload and download times
 - Multiple connections
 - Access logs





VOSpace

- IVOA interface to distributed storage:
 - VOSpace 1.0 – basic flat store (FTP site)
 - VOSpace 1.1 – added containers and links
 - VOSpace 2.0 – RESTful version
- Operations:
 - create, move, copy, delete
 - get, set, find
 - pushTo, pullFrom, pullTo, pushFrom
- Resources
 - properties, views, protocols
 - searches, nodes, transfers



Cloudspace

- Variety of cloud storage solutions defeats concept of cloud – which cloud is my data in?
- Two possible solutions:
 - Proxy service (appears as a VOSpace) for control flow
 - Centralised
 - Potential security issues
 - Deploy VOSpace instances in popular clouds, e.g. Amazon S3 and SimpleDB
 - Cost

