

THE US NATIONAL VIRTUAL OBSERVATORY

# Some thoughts on clouds and workflows

Matthew J. Graham (Caltech, NVO)



# Motivation

- 09/28/09: 180GB subset of SDSS DR6 released as a AWS Public Data Set in the form of a MDF SQL Server file  
(<http://developer.amazonwebservices.com/connect/entry.jspa?externalID=2797&categoryID=287>)
- Flurry of activity on mailing list concerning what can we actually do with this and is this the way of the future, etc.



# Haven't we been here before?

- Trieste – joint session with AstroRG
  - create page on Wiki with Taverna tutorials
  - provide Astrogrid plug-ins to IVOA
  - initiate discussions on a common approach to grid/cloud abstraction



# Why workflows?

- Workflows provide:
  - a systematic and automated means of conducting analyses across diverse datasets and applications
  - a way of capturing this process that that results can be reproduced and the method can be reviewed, validated, repeated and adapted
  - a visual scripting interface so that computational scientists can create these pipelines with low-level programming concern
  - an integration and access platform for the growing pool of independent resource providers so that computational scientists need not specialize in each one
    - Carole Goble and David de Roure in *The Fourth Paradigm*



# Clouds

- Local clusters
- National-level facilities
  - Teragrid
- The “IVOA Cloud”
  - VOSpace, VOPIpe
- Third-party environments
  - EC2/S3, etc.

