MontageMeeting033009

Meeting to discuss provenance issues in Montage and astronomy applications in general.

Vahi 14:30, 30 March 2009 (PDT)

Attendees

- Bruce Berriman (IPAC)
- Ewa Deelman
- Karan Vahi
- Paul Groth
- Raphael Bolze

Background

- NSF Astronomy money
- As part of Montage, developed a mosaic engine.
- Montage is technically not funded. John Good and Bruce do it in their own free time.

Provenance in Astronomy

Historical Perspective

- Astronomy has a standard format for data. One of the pioneers on it.
- However fallen behind the curve. Lot of Astronomy applications will suffer from this.

Examples from Astronomy

- Data from Spitzer observatory. Nothing on horizon to replace it. Most of the data products delivered are not the images out of telescope. But combine images from various points of sky, times of observation. People who delivered them have no idea where the data came from. Any astronomer who wants to understand these images, cannot really reproduce them as no provenance is there, as provenance is gone. How to record provenance ? Someone will find in the images that cannot be edited. Automated pipelines for creating the images exist.

- Montage Example
    - Users download montage code,data and make their own mosaics. Some astronomers took data from 2mass, SDSS and created an image mosaic to study gravitational lensing. The image was published but the source of provenance was lost. Someone might use this image to combine with another image and publish it.
    - Project will use montage as an automated product. Some people part of Spitzer project put montage in the pipeline but dont keep information about what went into the input. One of teams is GLUN?

Metadata in Montage

Currently, a fits image is converted to jpeg. John stretches images till it looks nice. no standard metadata in astronomy for recording metadata for jpegs. There is proposal called AVM (Astrophysical Virtual Metadata) extension of XMP (the standard that digital camera's use). JPEG supports writing of metadata as text in the image. But people dont write it. Most products are in fits format. Astronomers create JPEG out of them.

What to log for Provenance

How to figure out what will be useful. Otherwise problem of too much provenance. 2mass data is fully archived. That is fully documented. The problem is if people create new products from the mission / 2 mass data and then information is lost. Sometimes people exclude input images that have air glow problem. The background image is black art, where they try to flatten out the background.

Scenario to trace defects in images

- you need to go back to original set of images
- which images were excluded, how was re projection done (determined from version of montage) ,
- how was background rectification done.
- version of codes. Since montage is a toolkit one can mix various components from different versions of montage.
    - What happens if a new compute system comes in the future and u need to compare old runs. Or if you are seeing an image generated from old version of montage code and now that code is no longer supported.

In general since the montage codes are deterministic we can only record provenance about the output data and not worry about intermediate data.

Things to do

- Paul will look at what environment/ecology folks are doing.
  - Send a link to Open Provenance Model.
- Bruce will work on usecases.
- Paul will also look at different approaches to provenance
- Karan sets up a separate montage wiki.

Time Frames

- June 2009 - Astronomy Meeting where Bruce presents some ideas
- June 2010 - prototype with Montage and 2 mass data. Pipeline focussed approach.
- June 2011 - Using these approaches to provenance in operation so that astronomers can use them.

Future Collaboration

New survey, satellite will be launched later this year. Focus on infrared astronomy. A new project pipeline is being built.
ST pipeline . Space Telescope pipeline to do automated processing of data from the Space Telescope.