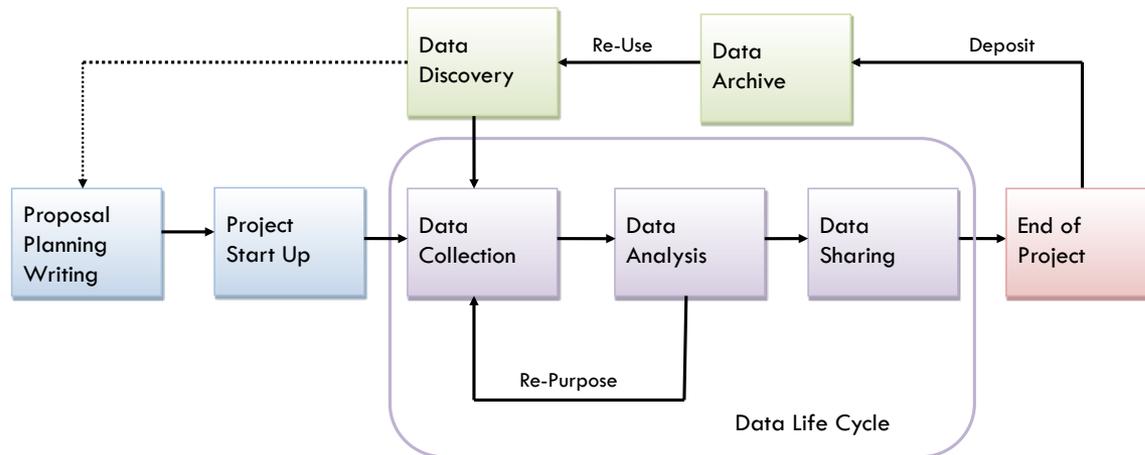


Workflow Systems for Engineering and Physical Sciences



Andrew Sallans

Head of Strategic Data Services and
Data Management Consulting Group

University of Virginia Library

als9q@virginia.edu

Goals for the workshop

- Learn about workflow and process
- Identify workflow tools
- See how workflow tools can support your data management
- Gain peer and expert feedback

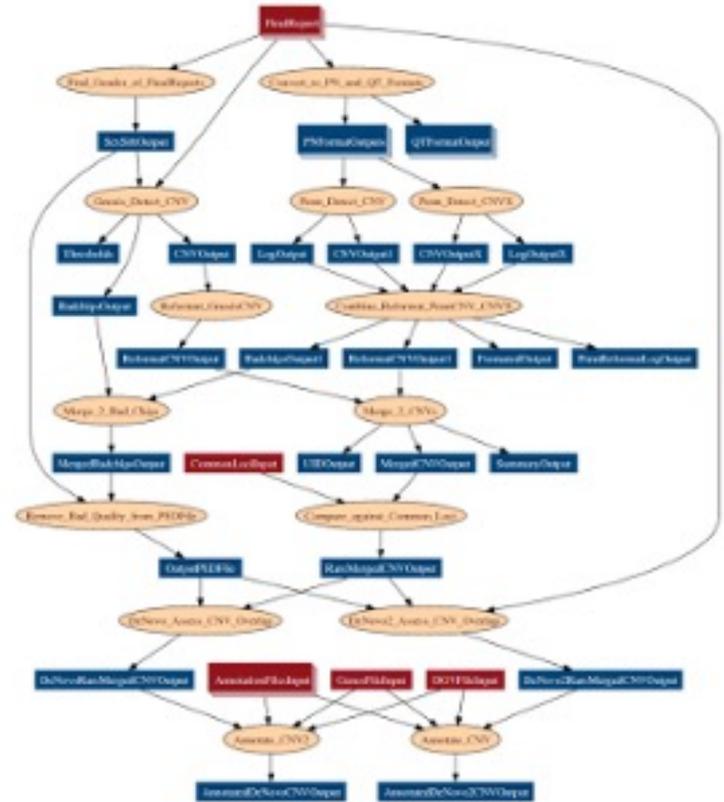
What is workflow or process?

According to the Pegasus Workflow Management System Project...

“allows users to easily express multi-step computational tasks, for example retrieve data from an instrument or a database, reformat the data, and run an analysis.”

Features from Pegasus (as an example)

- Portability / Reuse
- Performance
- Scalability
- Provenance
- Data Management
- Reliability
- Error Recovery

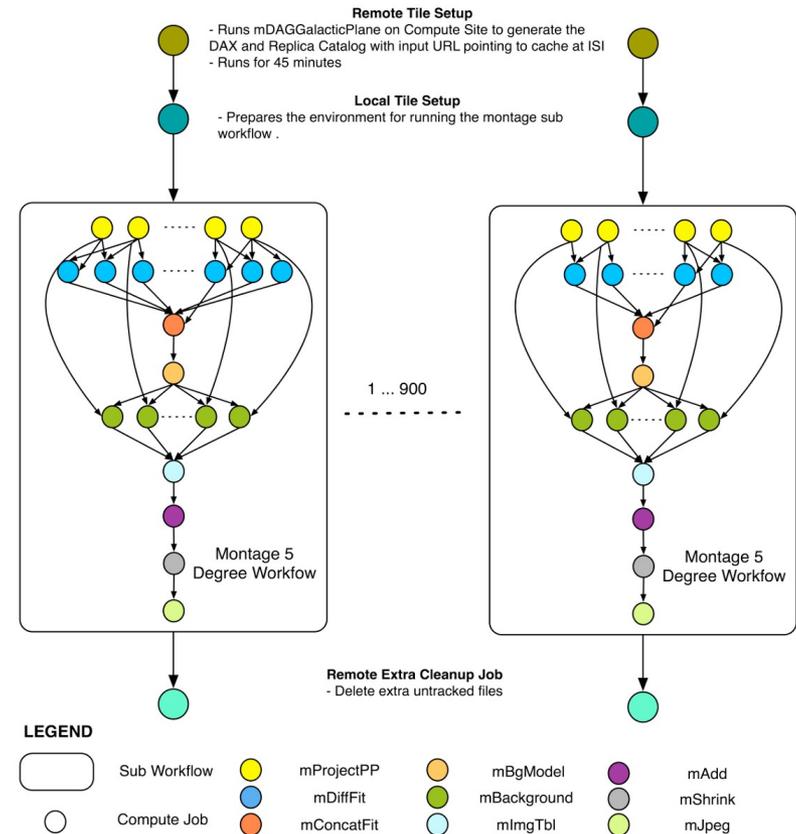


See for descriptions and image source:
<http://pegasus.isi.edu/>

Galactic Plane Workflow

- 18 million+ images from Spitzer Space Telescope
- Problems with inconsistent parameters
 - Differing shapes/sizes
 - Not of the same position on the sky
 - Differing projections
 - Sky/instrument noise
- Solution
 - Normalize pixel scale across 17 surveys
 - Assemble the images as a mosaic
 - Automate repeated tasks and track

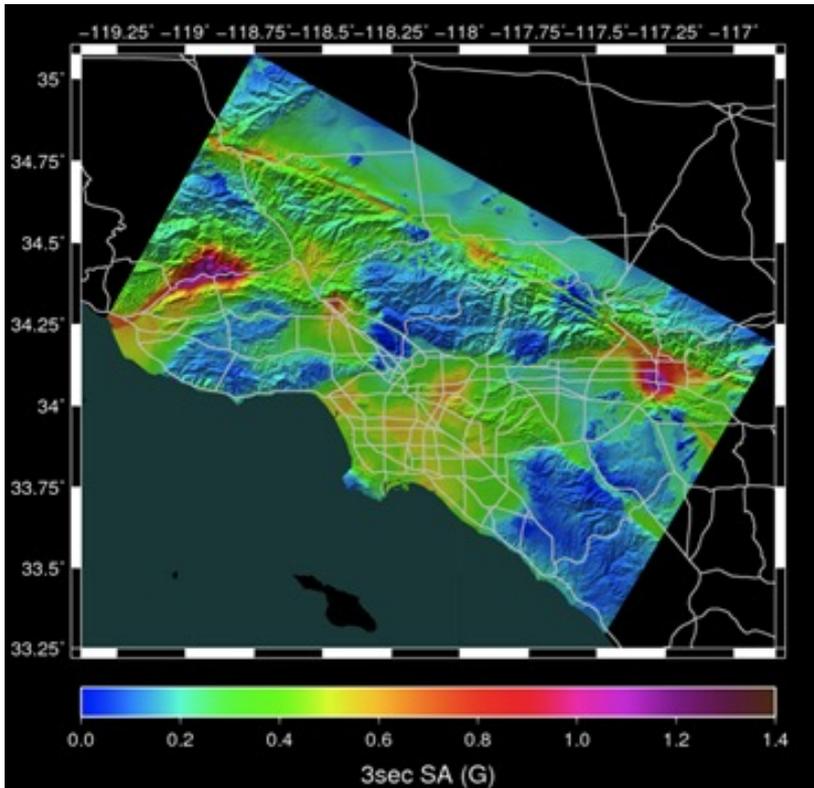
Montage Galactic Plane Workflow



Source: <http://pegasus.isi.edu/applications/galactic-plane>

CyberShake

- CyberShake
 - Cluster processing of MPI style jobs generating Strain Green Tensors
 - Post-processing of jobs for 10x10km regions
 - Hundreds of thousands of post-processed jobs (seismograms and peak spectral acceleration values creating hazard curves)
 - Several hundred hazard curves to create a hazard map
- Runs on TeraGrid and Open Science Grid
- More maps have been created through this process
- In 2006, two workflows ran on TeraGrid, took 23 days to complete, included 87,000 jobs.



What is your experience?

- Software solutions?
- Challenges you face with workflows?
- Barriers to using these solutions?

Where do workflows fail?

- Talk about common challenges
- Ask group about challenges

What workflow tools are available?

- Show some broad examples
 - Accelrys
 - MyExperiment
 - GoLIMS
- Show some domain specific examples
 - Social Sciences: Microsoft Visio,
<http://kieranhealy.org/files/misc/workflow-apps.pdf>
 - Life Sciences: VisTrails (SAHM), Kepler, Taverna, Galaxy, ESRI ArcGIS ModelBuilder, OpenMI, Science Pipes
 - Engineering: Platform LSF
 - Physical Sciences: ELNs, LIMS, SDMS, Pegasus
 - Humanities: Virtual Research Environment, Omeka
- Ask group if they have other examples

Demo 1: First Tool

- Have several examples ready to pull up demos, offer a few choices to meet audience interest

Mailing List Subscription

- Please check the box on our sign-in sheet to receive occasional emails to keep up with our services, training, and news.
- Please encourage others to subscribe:
<http://eepurl.com/CJwYT>

More Research Data Services in the Library

Offering expert data assistance at every stage of the research process.



PLANNING

Need a data management plan?

We can assist you with developing a data management plan that meets increasingly stringent criteria from funding agencies, including:

- Implementation of procedures, tools, and workflows for managing data sets
- Designing a strong study that yields reliable statistics

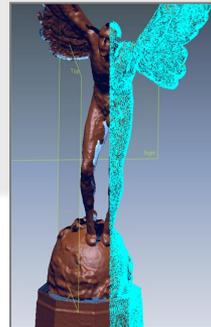
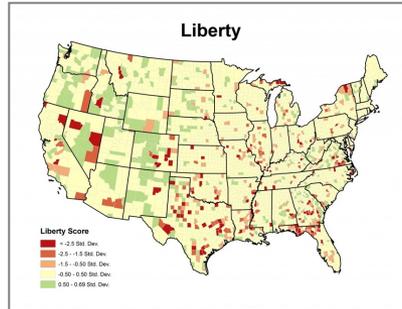


FINDING & COLLECTING

Need help finding data or collecting your own?

We have thousands of sources with the data you seek and experts who will help you:

- Locate, evaluate and format data
- Design metadata and data documentation protocols for new data collection
- Capture data with the appropriate technology tools for your needs



SHARING

Ready to share or archive your data?

We can consult with you on strategies to help others discover or access your research by:

- Adhering to data sharing policies and norms
- Selecting a data-sharing repository
- Making your data easier to discover and link



ANALYZING

Want help uncovering unique and compelling insights?

Get expert assistance from statistical, spatial, or media specialists to analyze your data and convey your research message:

- Learn how to use cutting-edge tools and methods
- Experiment with high-resolution visualization technologies
- Develop graphical representations that bring impact to your analysis

QUESTIONS?

Data Management Consulting Group
University of Virginia Library

<http://dmconsult.library.virginia.edu>

dmconsult@virginia.edu