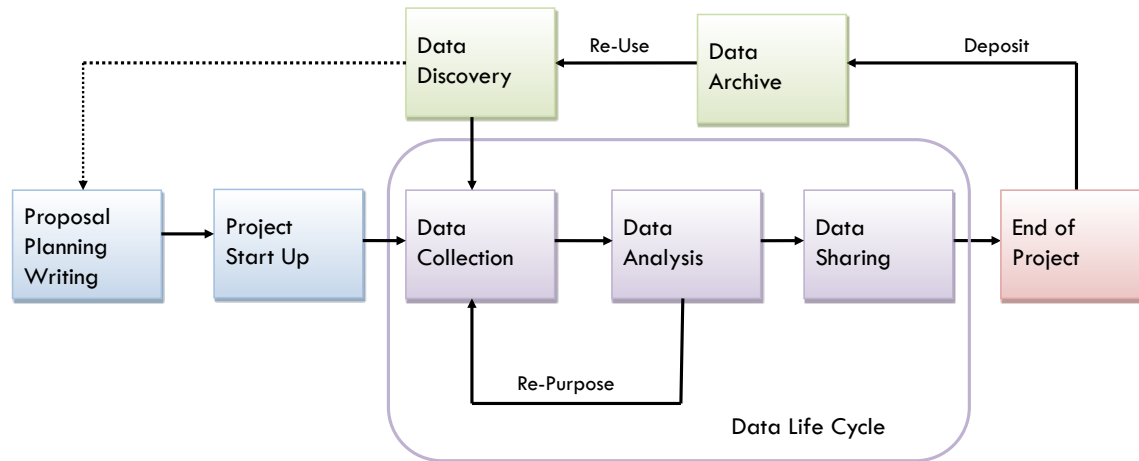# Workflow Systems for Life Sciences and Social Sciences

Bill Corey
Data Consultant
Data Management Consulting Group
University of Virginia Library
wtc2h@virginia.edu

Andrea Horne Denton
Health Sciences Data Consultant
Claude Moore Health Sciences Library
ash6b@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

Workflow versus Process

Workflow: A sequence of connected steps where each step follows the next, or where steps occur concurrently. It is a systematic pattern of activity.

Process: A workflow that has well-defined inputs, outputs and purposes.

http://en.wikipedia.org/wiki/Workflow

Glossary

- **Workflow Management System (WMS):** Computer system that manages & defines a series of tasks.

- **Scientific Workflow System (SWS):** Computer system designed to compose & execute a series of computational or data manipulation steps in a scientific application. Also called a Scientific Resource Management System (SRM) or a Scientific Workflow Management System (SWMS).

- **Electronic Lab Notebook (ELN):** Computer program designed to replace paper laboratory notebooks.

- **Semantic Workflow System (SWS):** Computer system that designs computational experiments, validates workflows, and generates metadata.

- **Laboratory Information Management System (LIMS)**: Computer program that serves as a laboratory and information management system. Often includes ELNs, SWSs, data management, and data analysis tools. May also be called a Laboratory Information System (LIS) or Laboratory Management System (LMS).

What is a scientific workflow?

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."

Scientific workflows allow scientists to model, design, execute, debug, re-configure and re-run analysis and visualization processes.

Typical features of workflows

- Portability / Reuse

- Flexible design

- Performance

- Scalability

- Provenance

- Data Management

- Reliability

- Reproducibility

image source: http://pegasus.isi.edu/
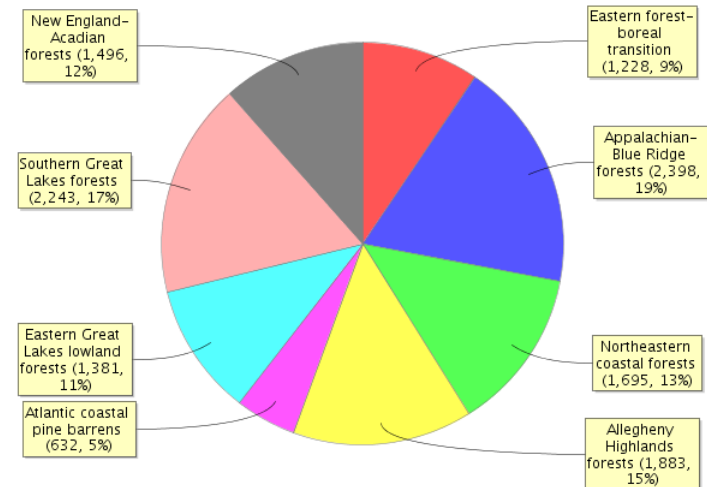
Science Pipes and Kepler

Based on the Kepler Scientific Workflow Management System (SWMS)

Science Pipes allows anyone to access, analyze, and visualize primary biodiversity data using the data analysis tools of Kepler. Analyses and visualizations can be shared, modified, reused and enhanced.

The Kepler Project

- Create, Execute and Share models and analyses
- Scientific and Engineering disciplines
- Data sources can be local or online
- Integrate different software components, such as "R" and "C" code
- Facilitate remote and distributed execution of models
- Graphical User Interface (GUI)


Species Richness for New York Ecoregions

New England–Acadian forests (1,496, 12%)

Eastern forest–boreal transition (1,228, 9%)

Appalachian–Blue Ridge forests (2,398, 19%)

Southern Great Lakes forests (2,243, 17%)

Northeastern coastal forests (1,695, 13%)

Eastern Great Lakes lowland forests (1,381, 11%)

Atlantic coastal pine barrens (632, 5%)

Allegheny Highlands forests (1,883, 15%)

Sources: https://kepler-project.org/ & http://sciencepipes.org/beta/home

Image source: http://sciencepipes.org/beta/pipes/detail?workflowID=b1e1c4c6-8931-11df-8902-cb7a96ae555e

UNIVERSITY of VIRGINIA LIBRARY

irisnote SRM Platform

Enterprise Electronic Lab Notebook (ELN)

and Content Management System (CMS)

irisnote

- Simple, powerful replacement for paper notebooks
- Highly scalable architecture
- Da...
- Sa...
- Co... me...
- Manage & protect institutional knowledge -- fully compliant with 21 CFR Part 11, GLP & ISO regulatory requirements, encryption
- Cross-platform -- Mac, Windows, iPad

Irisnote announced Oct. 23, 2013 that the basic cloud service will be removed from service on Oct. 25, 2013.

Sources: http://irisnote.com/how-it-works/

Image source: http://irisnote.com/how-it-works/
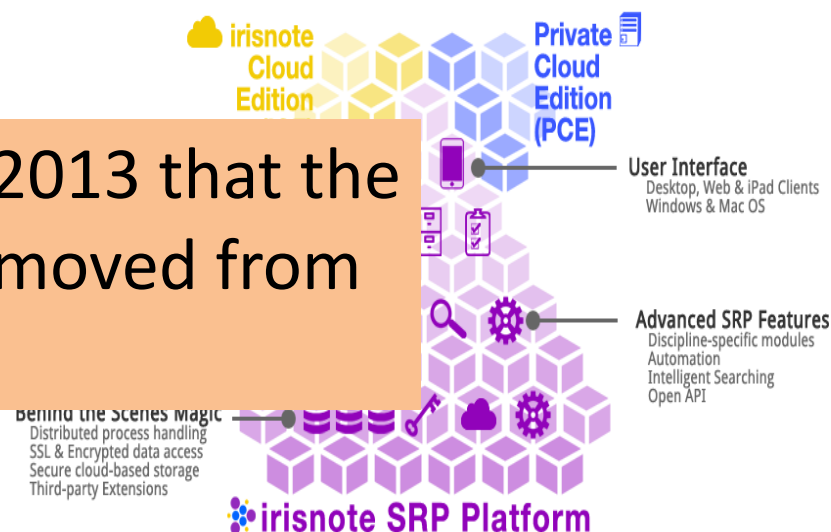
WINGS

Workflow Instance Generation & Specialization

- **Semantic Workflow System** that assists scientists with the design of computational experiments
- Multi-discipline: sociology, biology, education, environmental sciences
- Uses Pegasus or OODT as the execution engine
- Can submit workflows in scripted format for local, secure execution
- Generates metadata attributes for all new data products
- Modular design
- Designed to interface with external data & component catalogs

Source: http://wings-workflows.org/about
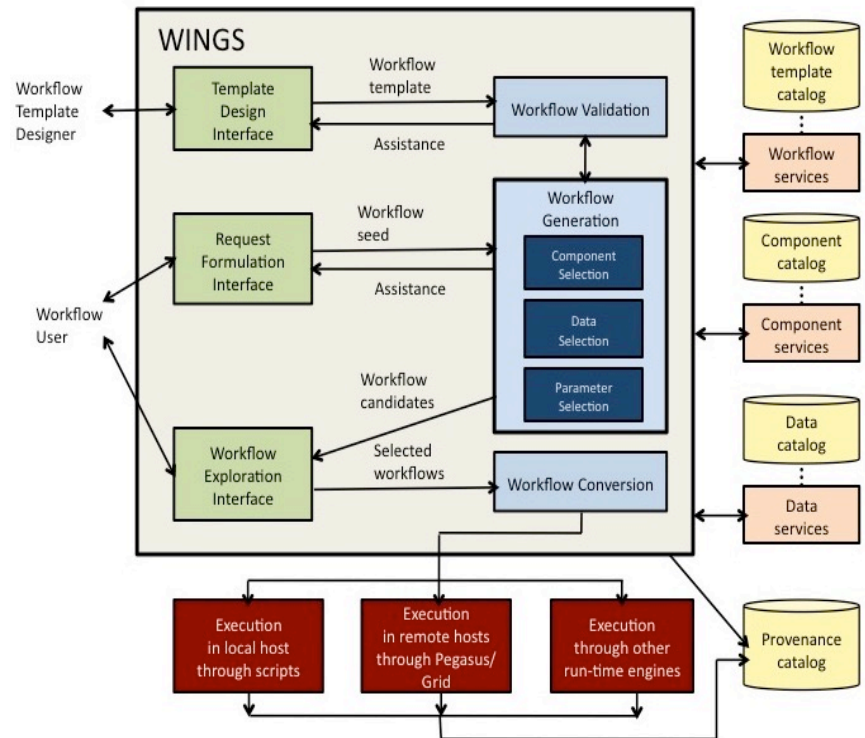
Open Science Framework

The Open Science Framework (OSF) is part network of research materials, part version control system, and part collaboration software. The purpose of the software is to support the scientist's workflow and help increase the alignment between scientific values and scientific practices.

- Document and archive studies
- Share and find materials
- Detail individual contributions
- Increase transparency
- Register materials
- Manage scientific workflow

Collaborator Network for Public Projects

Projects with more than 10 connections are used to represent the connections between each of the collaborators on that project for clarity. Node radius represents total collaboration overall, while number of lines connecting to a node represent the individuals they collaborate with.

Source: https://openscienceframework.org/explore/

UNIVERSITY of VIRGINIA LIBRARY
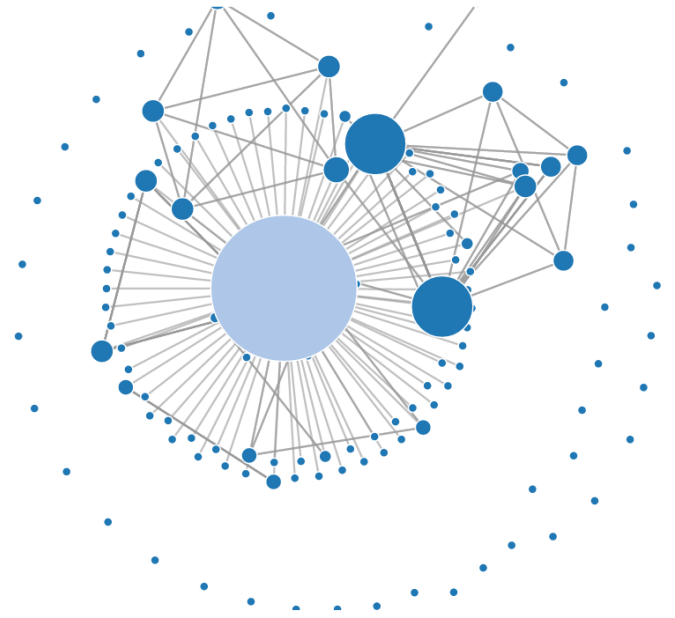
What is your experience?

- Does anyone have a tool or workflow system that they are using now?

- What challenges have you encountered using a tool or workflow?

- What are the barriers to using these solutions?

- What features are you looking for in your 'ideal' workflow system tools?

What workflow tools are available?

- **Scientific Workflow Management Systems -- General**:
  - Pegasus http://pegasus.isi.edu/
  - Taverna http://www.taverna.org.uk/
  - Kepler https://kepler-project.org/
  - accerlys http://accelrys.com/
  - Open Science Framework https://opensscienceframework.org/
  - ESRI ArcGIS http://www.esri.com/software/arcgis

- **Scientific Workflow Systems -- Domain-specific**:
  - Social Sciences: Microsoft Visio, WINGS, Argos Ontology, text editors (Emacs, notepad)
  - Life Sciences: VisTrails, Kepler, Galaxy, Science Pipes, Scilligence
  - Engineering: Triana, GridANT, GridFlow, Askalon, Karajan, Platform LSF
  - Physical Sciences: ELNs, LIMS, SDMS, Pegasus
  - Humanities: Virtual Research Environments, Omeka

What workflow tools are available?

- **Workflow Modules and Tools:**
  - Karma Provenance Collection http://d2i.indiana.edu/provenance_karma
  - myexperiment: scientific workflows and research objects http://www.myexperiment.org/
  - Microsoft Visio: flowchart software
  - Microsoft SharePoint: document management and scientific workflow integration

- **ELNs, LIMs**
  - Golims http://www.golims.com/
  - Labarchives http://www.labarchives.com/
  - Labguru http://www.labguru.com/
  - Notebookmaker http://www.notebookmaker.com/
  - LabCollector http://labcollector.com/
  - Thermo Scientific Watson http://www.thermoscientific.com/

Demo 1: Science Pipes/Kepler

- Video: Introduction to SciencePipes
  http://info.sciencepipes.org/help/2012/10/video-introduction-to-sciencepipes.html

- Video: How to create a pipe
  http://info.sciencepipes.org/help/2012/12/video-how-to-create-a-simple-pipe.html

Demo 2: WINGS

- WINGS: http://wings-workflows.org/about
- WINGS example: http://seagull.isi.edu/marbles/
- WINGS Sandbox: http://www.wings-workflows.org/sandbox/

- OODT: http://oodt.apache.org/
- Pegasus: http://pegasus.isi.edu/

Demo 3: Open Science Framework

- OSF: Scientists

- https://www.youtube.com/watch?feature=player_embedded&v=c6lCJFSnMcg

- OSF: Developers

- https://www.youtube.com/watch?feature=player_embedded&v=WRadGRdkAIQ

More ideas!

- William J. Turkel Doing research with digital sources http://williamjturkel.net/how-to/

- Bamboo DiRT   Digital Research Tools http://dirt.projectbamboo.org/

- Transforming Scholarly Communication http://msrworkshop.tumblr.com/tagged/resources

- Mindmeister Mind mapping & brainstorming http://www.mindmeister.com/

- Scrivener http://www.literatureandlatte.com/scrivener.php

- Zotero http://www.zotero.org/

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

RESEARCH DATA SERVICES

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

1: PLANNING

We can assist you with developing a data management plan and designing your planned data analysis, including:

- Implementing plans, using tools, and creating workflows for managing research data
- Advising on study design, power analysis, and choice of statistical methods
- Helping to meet increasingly stringent criteria from funding agencies

2: FINDING & COLLECTING

We have access to thousands of sources of data and experts who will help you:

- Locate, evaluate and format data
- Create metadata and data documentation protocols for new data collection
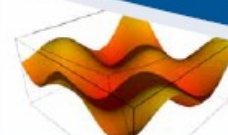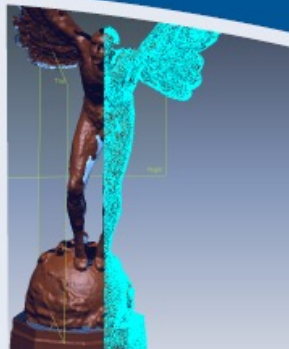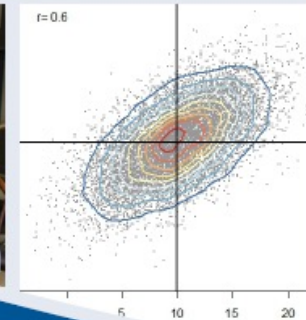- Capture data using best practices and appropriate technology

3: ANALYZING

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

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

4: SHARING & ARCHIVING

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 reuse

UNIVERSITY OF VIRGINIA LIBRARY

Workshops • 1:1 Consultations • Class Presentations
Contact us at researchdataservices@virginia.edu

QUESTIONS?

Bill Corey
Data Management Consulting Group
University of Virginia Library
wtc2h@virginia.edu

Andrea Horne Denton
Health Sciences Data Consultant
Claude Moore Health Sciences Library
ash6b@virginia.edu

Data Management Consulting Group
University of Virginia Library
http://dmconsult.library.virginia.edu
dmconsult@virginia.edu