iRODS Strategy and Future

Reagan W. Moore

University of North Carolina at Chapel Hill

rwmoore@renci.org

Topics

- Open source collaboration
- Sustainable version
- Future enhancements
- Integration with storage and networks

Open Source

- **Software available from**
  - http://irods.diceresearch.org
- **Contributions from users**
  - New features needed for a specific application
  - Bug fixes
  - Collaborative development
  - Co-funded development

Example Contributions

- **IN2P3, Jean-Yves Nief**
  - Monitoring system
  - Simple Mass Storage System interface
- **Distributed Bio, Chris Smith**
  - Integration with file system
  - PAM integration
- **SLAC**
  - Port to mySQL

Sustainable Version

- **RENCI E-iRODS development**
  - Collaborated on port to C++ and boost libraries
  - Software testing environment
  - Generation of binary release
  - Development of new features
    - Pluggable micro-services
    - Pluggable storage drivers
  - Developing model for consortium-based co-development of the software

DICE Group

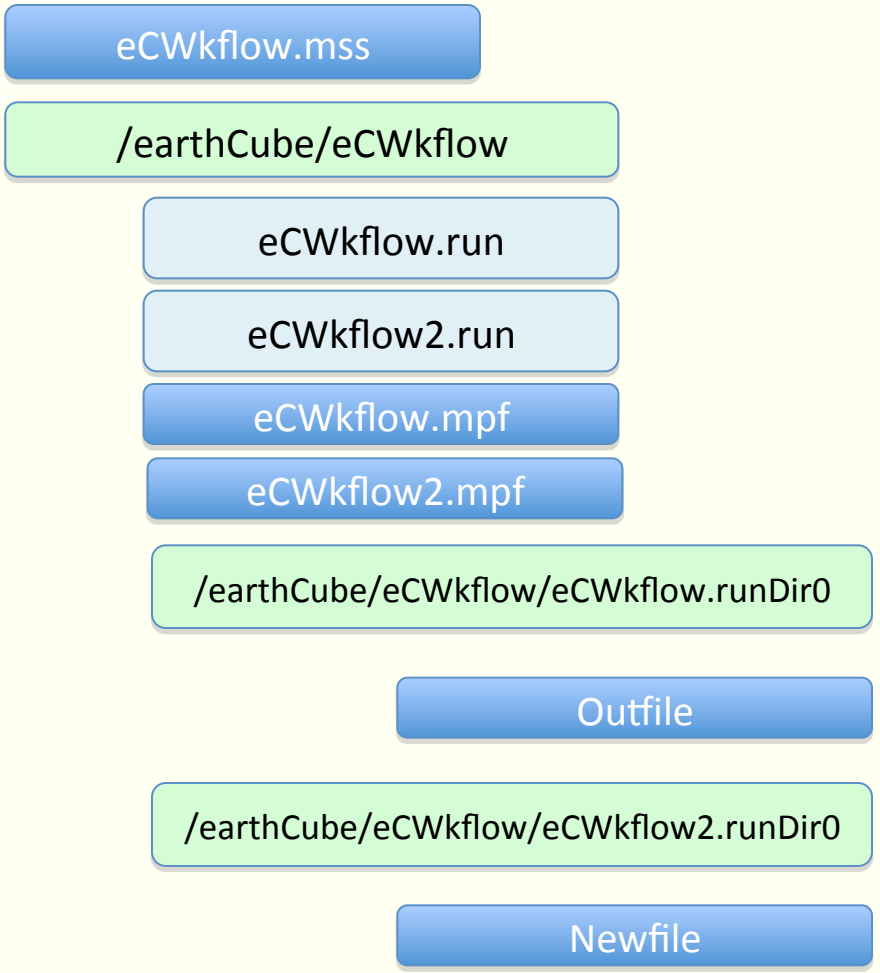
- **Supported by NSF funding:**
  - NSF OCI-1032732 grant, "SDCI Data Improvement: Improvement and Sustainability of iRODS Data Grid Software for Multi-Disciplinary Community Driven Application," (2010-2013)
  - NSF Cooperative Agreement OCI-094084, “DataNet Federation Consortium”, (2011-2013).
- **Actively seeking additional funding sources**
  - NSF EarthCube Initiative
  - NSF Big Data solicitation
  - NSF DIBBS solicitation

D·I·C·E renci UCSD DFC DataNet FEDERATION CONSORTIUM THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL THE NATIONAL ARCHIVES ARCHIVES.GOV NSF

Future enhancements

- **iRODS version 3.2 scheduled for release at the end of September**
  - Support for registration and sharing of workflows
  - Support for Pluggable Authentication Modules
  - Support for pluggable micro-services
  - Support for pluggable storage resource drivers
  - Support for storage drivers for THREDDS, OpenDAP (pyDAP andERDDAP), NetCDF

Workflow Management & Registration



Workflow file

Directory holding all input and output files associated with workflow file (mounted collection that is linked to the workflow file)

Automatically generated run file for Executing each input file

Input parameter file, lists parameters and input and output file names

Directory holding all output files generated for invocation of eCWkflow.run, the version number is incremented

Output file created for eCWKflow.mpf

Future Technology

- **Integration with DDN storage controllers**
  - Creates intelligent storage through the ability to automate application of policies and procedures within the storage system
  - A disk spin that is not used is a revolution lost forever
- **Apply policies to enable feature-based indexing of data**
  - Apply the procedures that detect the presence of relevant features
  - Index data based on the features that are present
  - Rebuild index when new procedures become available to detect new features

Future Technology

- **Integration with the Future Internet Architecture**
  - Embed policies within the network through use of Software Defined Network Overlays
  - Policies can include network quality guarantees (bandwidth, latency, jitter)
  - Policies can include accessing data by name instead of IP address
  - Define a virtual circuit that implements policies across a network subset
- **Map the virtual circuit to a virtual collection**
  - Enable guarantees on data delivery within a virtual collection
  - Enable access by file name, or file property instead of by IP address
  - Moves data grid access policies into the network

Your Future Technology

- **What requirements do you have for the next generation data grid?**
  - Policy enforcement
  - Automation of administrative functions
  - Validation of assessment criteria
- **Are there standard policy sets that can be used as the basis for building your data management application?**