Kepler WRP: Workflow Management and Reporting

Ben Leinfelder

Ilkay Altintas(2), Derik Barseghian(1), Daniel Crawl(2), Matthew B. Jones(1), Aaron Schultz(1), Debi Staggs(1)
(1) National Center for Ecological Analysis and Synthesis; (2) San Diego Supercomputer Center

National Center for Ecological Analysis and Synthesis,
University of California Santa Barbara

Ptolemy Mini-conference
April 16th, 2009
ATTACK OF THE OUTPUT!

Hello, world!
Hello, world!
Hello, world!

```
> setwd("/Users/leinfelder/.kepler/05-LinearRegression_1239293297419/"
> pdf(file = 'R_linear_regression1.pdf', width = 6.66666666666667, height = 6.66666666666667,
> T_AIR <- c(15.0, 13.4, 13.4, 12.4, 11.7, 11.4, 11.5, 11.5, 12.2, 17.4, 20.1, 23.3, 23.1,
> BARO <- c(953.4, 953.8, 954.0, 954.3, 954.5, 954.7, 954.8, 954.8, 954.9, 953.7, 952.6, 951.7
> BARO
> [1] 953.4 953.8 954.0 954.3 954.5 954.7 954.8 954.8 954.9 953.7 952.6 951.7
> [13] 951.2 950.7 950.3 950.6 951.8 952.3 952.8 953.1 953.3 953.4 953.6
> [25] 953.7 954.0 954.3 954.7 954.9 955.1 955.2 955.3 955.2 954.0 952.8 952.1
> [37] 951.5 950.9 950.5 950.4 950.3 950.7 952.2 952.8 953.1 953.3 953.5 953.9
> [49] 954.2 954.4 954.4 954.6 954.8 954.9 955.2 955.4 955.4 954.1 952.8 952.0
> [61] 951.5 951.0 950.6 950.2 951.2 952.4 952.8 953.1 953.3 953.4 953.7
> [73] 954.3 954.5 954.7 954.9 955.1 955.3 955.4 955.5 955.4 954.2 952.9 952.1
> [85] 951.6 951.1 950.5 950.3 950.5 951.2 952.2 952.9 953.3 953.8 954.1 953.9
> [97] 953.9 954.0 954.0 954.3
> res <- lm(BARO ~ T_AIR)
> res

Call:
  lm(formula = BARO ~ T_AIR)

Coefficients:
(Intercept)       T_AIR
  958.3772    -0.3244

> plot(T_AIR, BARO)
> abline(res)
```

Output, tamed
SANParks TPCs

• Thresholds for Potential Concern – River Flow exceedance
while (...) {
    edit workflow model
    edit report layout
    execute workflow
    view results
    adaptive management
}
Workflow Run Manager

- “Provenance Browser”
- Search
- View
- Annotate
- Share
  - Publication Ready Archive
  - Repository

- REAP
  - Manage
  - Share
  - Publish
  - Archive
Workflow Run Manager

Calculate and Plot Growing Degree Days using temperature data from REAP01 station at Baskett Slough.

5 degree C developmental threshold for Aphids
Provenance

- Record
  - Executions
  - Workflows
  - Actors
  - Ports
  - Parameters
  - Tokens

- Provide
  - Workflow history
  - Data for Report generation
• Simple Layout for Workflow items
  – Ports
  – Parameters
  – Labels
  – Text
• Layout
  – BEFORE run

• Instance
  – AFTER run
    + Layout
    + Provenance
    + Token value[s]
Report Rendering

- Visual Rendering
  - Kepler GUI
  - PDF
  - HTML

- Apache FOP
  + XML
  + XSLT
  + Magic
Title: Report for workflow 'central'

Hello, world!
• Publication Ready Archive
  – Workflow
    • MOML (model)
  – Data
    • Input
    • Output
  – Report
    • ROML (layout)
    • RIOs (instance)
    • PDFs (instance)

✓ Sharable
✓ Replicable
✓ Archival
Sharing Widely

- Web interface [built upon Metacat]
  - Publish
    - workflows
    - archives
  - Manage
    - Schedule workflow executions
    - Set workflow parameters
    - View completed execution results
    - View rendered reports
Acknowledgements

• This material is based upon work supported by:
  – The National Science Foundation under Grant Number 0722079 (Kepler/CORE)
  – The National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant Number 0072909), the University of California, and the UC Santa Barbara campus.
  – The Andrew W. Mellon Foundation.

• Resources
  – http://kepler-project.org/
  – https://kepler-project.org/developers/interest-groups/provenance-interest-group
  – http://www.nceas.ucsb.edu/ecoinfo