ADASS XXI Paris, 6-10 November 2011

Scientific Workflows in Astronomy BoF

A. Schaaff (CDS), L. Verdes-Montenegro (IAA-CSIC), J.E. Ruiz (IAA-CSIC), J. Santander Vela (ESO)

Contributions from Pascal Ballester (ESO), Sami Maisala (University of Helsinky),
Francesco Pierfederici (STSCI) and Christian Surace (LAM)

Session planning

- Introduction and Workflows@IVOA, André Schaaff
- Wf4Ever, Juande Santander Vela
- IceCore, Sami Maisala
- UVES workflow, Pascal Ballester
- Different Wf levels in a project, Christian Surace
- OWL, Francesco Pierfederici
- Discussion points

ADASS XXI Paris, 6-10 November 2011

Introduction and Workflows@IVOA

A. Schaaff (CDS)

Workflow

- Workflow is a general denomination used in several domains, from business to science
  - A jungle of definitions, languages, engines, workbench, etc.
- From a simple script to elaborated design / language / execution / control tools

A few acronyms / terms in the Wf world…

- Languages : AGWL, BPEL4WS, BPML, DGL, DPML, GJobDL, GSFL, GFDL,GWorkflowDL, MoML, SWFL, WSCL, WSCI, WSFL, XLANG, YAWL, WPDL, PIF, PSL, OWL-S, xWFL…
- Formalism : AGWL, BPEL4WS, BPML, DGL, DPML, GJobDL, GSFL, GFDL,GWorkflowDL, MoML, SWFL, WSCL, WSCI, WSFL, XLANG, YAWL, WPDL, PIF, PSL, OWL-S, xWFL…
- Design tools : ilog's BPMN Modeller, CAT, GWUI, XBaya GUI for Workflow Composition, Taverna, Triana, JOpera, Platform Process Manager. . .
- Engines :BioPipe, BizTalk, BPWS4J, DAGMan, GridAnt, Grid Job Handler, GRMS, GWFE, GWES, IT Innovation Enactment Engine, JIGSA, JOpera, Kepler, Karajan, OSWorkflow, Pegasus (uses DAGMan), Platform Process Manager, ScyFLOW, SDSC Matrix, SHOP2, Taverna, Triana, wftk, YAWL Engine, WebAndFlo, WFEE. . .

Examples of Wf tools used in astronomy

- Taverna
  - Bioinformatics closely oriented at the beginning
- Triana
  - A workflow system originally built to provide a tool for rapid analysis of data from gravitational waves
- Kepler
  - Generic science oriented workflow system

Related initiatives in astronomy

- ESO Reflex, UVES workflow see Pascal's presentation
- AstroGrid
- Helio-VO
- VO France Workflows WG
- CyberSKA
- Montage
- Wf4Ever (see Juande's presentation)
- And others

Workflow@IVOA

- Talks about workflows in the past in Grid and Web Services WG and in Data Curation and Preservation IG
- Workflow initiative started in July 2011
  - Call for discussion and collaboration (over 30 interested people)
  - Creation of a mailing list, workflow@ivoa.net, of interested people (26 registered)
  - Writing of a Note draft published during the last Pune Interop
    - Talks and discussion
    - A final version will take into account remarks, ideas, questions, …, from both IVOA meeting and ADASS BoF
  - Planned during next Interop meeting
    - Discussion about what we need in the VO : new standards ?, best practice guide ? work on methodology ? preservation ?, etc.

Remarks from last IVOA meeting (1)

- Workflows are very “popular” in other communities
  - Example : 
    - Social website for workflow experiment sharing
    - Mainly bioinformatics oriented (a few Wfs in astronomy)
    - 2141 workflows (8 November 2011)
- In astronomy, several projects are using workflows
  - Collaboration / sharing about methodology, tools, experiments ??

Remarks from last IVOA meeting (2) Workflows and VO Standards

- Data Modeling
  - Characterisation and Provenance (Wf final or intermediate data checking)
- VOTable
  - (tabular data exchange between the different steps of the Wf execuction)
- Semantics
  - Ontologies, Vocabularies, Annotations (adaptative Wfs)
- Data Access Layer
  - Table Access Protocol, Simple Image Access Protocol,…, (self-descriptive data sources)
- Grid and Web Services
  - Universal Worker Service (asynchronous Wf job execution), VOSpace (intermediate or final Wf data storage)
- Applications
  - Simple Application Messaging Protocol (interaction with VO tools (visualisation, etc.))
- Knowledge Discovery in Databases
  - cf. Raffaele D'Abrusco talk this morning
- Data Curation and Preservation
  - Persistant identifiers, preservation of VO resources (replaying of Wfs)
- Registry
  - Several standards (dynamic selection of data resources and services during a Wf execution)

Discussion points

- How to boost the creation / sharing of Wfs knowledge in our community ?
  - People could be reluctant to fill with care the information required as part of the workflow preservation (data used, services, references, etc.) or to make the effort to ensure that it is reusable by others
- Wf4Ever is a European project with a limited time life
  - How to preserve Wfs on the long term ? What kind of things (user datasets, methodology, tools for backward compatibility, Licenses??, etc.)
- What can we learnt / what have we learned from other communities ?
- Publishing Wfs : review methods (peer review, volunteer, etc.)
- What parameters would you use to rate a Wf ? Or which ones would you like a rating system to provide you, so that you can select based on their quality / rating ?
- …