International

Virtual

Observatory

Alliance

Scientific Workflows in the VO

Version 1.00

IVOA Note 2013 April 1st

This version:
http://www.ivoa.net/documents/Notes/ScientificWorkflows/20130401/

Latest version:
http://www.ivoa.net/Documents/Notes/ScientificWorkflows/

Previous version(s):

Author(s):
André Schaaff, Jose Enrique Ruiz et al.

Editor(s):
André Schaaff, Jose Enrique Ruiz

Abstract

We will soon be facing a new generation of facilities and archives dealing with huge amounts of data (ALMA, LSST, Pan-Starrs, LOFAR, SKA pathfinders...) where scientific workflows will play an important role in the working methodology of astronomers. A detailed analysis about the state of the art of workflows in the frame of the VO involves languages, design tools, execution engines, use cases, etc. A major topic is also the preservation of the workflows and the capability to replay a workflow several years after its design and implementation. Several talks concerning these issues have been presented during the past IVOA Interoperability meetings. In order to undertake this task within our community we

have decided as a first step to write this Note. We have collected experiences (including use cases, tools, etc.), references and remarks from the community.

Status of This Document

This is a Note. The first release of this document was 2013 April 1st.

This is an IVOA Note expressing suggestions from and opinions of the authors. It is intended to share best practices, possible approaches, or other perspectives on interoperability with the Virtual Observatory. It should not be referenced or otherwise interpreted as a standard specification.

A list of current IVOA Recommendations and other technical documents can be found at http://www.ivoa.net/Documents/.

Acknowledgements

All the participants registered on the workflow@ivoa.net mailing list.
Inputs from attendants to ADASS 2011 BoF about Scientific Workflows in Astronomy

Contents

1 Introduction

One of the current challenges in Astronomy is the efficient exploitation of the huge volume of data currently available. This is needed in order to ensure the prompt return of the big investments made in terms of facilities to obtain those data, something that clearly the traditional methods of analysis are not currently achieving. This is one of the most important reasons why scientific workflows are becoming a need in Astronomy.

Publishing the orchestration of data and processes as the methodology used in an astronomical digital experiment will need Virtual Observatory standards for the characterization of workflows, in order to be indexed, shared, and retrieved.

In this Note we intend to provide a very quick revision of the state of the art in the domain of scientific workflows, from general technical topics like languages and formalisms, composition tools and engines, through more astronomy specific related initiatives and concerns in the frame of the VO, as well as in different VO Working Groups.

2 State of the Art

2.1 Definition

"Workflow" is used to refer in general to modelling and IT management of all tasks and actors in the composition of a business process. The final goal is to automate the best working method as a concatenation of operations, often using distributed resources. In this note we will use the term workflow for both process and software.

There are two main types of workflows: business workflows and scientific workflows. We provide below a quick definition of business workflows, even if the goal of this note is to focus on scientific workflows. Workflows have been present in IT and in the scientific community for many years. This document is not intended to provide an exhaustive list of all workflow languages, engines, integrated tools, etc.

2.1.1 Business workflows

The business workflows (BWFs) appeared in the 70s and the definition that we retain is that given by the WFMC:

The automation of a business process, in whole or parts, where documents, information or tasks are passed from one participant to another to be processed, according to a set of procedural rules.

More practically, they can automate work processes within companies, which were previously done by hand. The BWFs are concatenated software-oriented tasks to perform complex workflows with a major control-flow policy.

2.1.2 Scientific workflows

The scientific workflows (SWFs) are a variant of BWFs. They are relatively similar but have different features that are not present in BWFs. We retain Bertram Ludäscher's [8] definition:

These are networks of analytical steps that may involve, e.g., database access and querying steps, data analysis and mining steps, and many other steps including computationally intensive jobs on high performance cluster computers.

This type of workflow is designed for scientists and, therefore, is able to meet their specific needs. Therefore, while BWFs are control-flow oriented, the SWFs are in contrast, data-flow oriented. They give the opportunity for users to operate easily in a large number of complex and heterogeneous data, with computationally intensive and distributed processing (e.g. grids, clouds...)

Workflows are useful to capture scientific methodology and to provide provenance information for their results. They provide also a formalization of the scientific analysis (routines to be executed, dataflow, execution details…) and they are very useful digital entities for managing computation at a large-scale. A large number of projects have defined their workflow language and the associated tools (engine, design and composition tools...)

2.1.3 Towards a new type of workflow

In recent years, given the popularity of workflows and to meet new expectations, a new type of workflow has emerged, which we can call "adaptive workflows". In the literature, we can see it under different names, i.e. "WDOs" (Workflow-Driven Ontologies)[1] or "flexible workflow". The main characteristic of this type of workflow is to offer the ability to change, more or less automatically, the structure of a workflow during its execution. It takes into account the execution environment of the workflows.

2.2 Languages and formalisms

A workflow **language** provides a way to describe a workflow and to make its execution possible through a workflow engine. It is like a programming language. A workflow could be defined at least with a simple script language (e.g. *Sculf* is an XML-based language associated to *Taverna 1.x* execution engine) Other

[1] http://trust.utep.edu/wdo/

examples: AGWL, BPEL4WS, BPML, DGL, DPML, GFDL, GJobDL, GSFL, ,GWorkflowDL, MoML, OWL-S, PIF, PSL, SWFL, SwiftScript, WPDL, WSCI, WSCL, WSFL, XLANG, xWFL, YAWL…

The workflow **formalism** is present at the modelling level. UML activity diagram is a well-known example, as well as Directed Acyclic Graphs (DAG). Other examples: Petri net, BPMN, IPO, GPSG, Workflow Patterns, Pi Calculus, Finite-State Machine, Gamma calculus…

The need for standards in this context is justified by the fact that all the workflow tools are based on a language of their own as well as a model of relations between objects and a set of commands for the transfer of information between participants. Workflows typically have their own user interfaces/APIs, description languages, provenance strategies, and enactment engines, which are not standard and do not interoperate. Workflow integration or reuse therefore is currently impractical, thereby inhibiting the growth and proliferation of workflows in scientific practice.

2.3 Workflow composition and enactment

2.3.1 Design tools

The process definition tools are tools to model the workflow to be performed. Thus, in most cases, these tools have graphical features for easy drag and drop tasks and actors in the composition of their processes. Existing communications between the entities are then defined by linking them just as easily. While some users like using graphical features to compose and describe workflows, others like scripts. For this group, writing a script is easier than using a GUI.

Examples: CAT, GWUI, ilog's BPMN Modeller, JOpera, Taverna Workbench, Triana, Platform Process Manager, XBaya GUI for Workflow Composition …

2.3.2 Workflow engines

The workflow engine is a software service that provides and controls all or only a part of the runtime of a workflow instance.

Examples: 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, WebAndFlo, WFEE, wftk, YAWL Engine…

2.3.3 Workflow Enactment System

At the heart of the workflow is the Workflow Enactment System. It is a service to create, manage, run instances of procedures and manage their interactions with the outside. It is composed of one or more workflow engines that allow maintaining an internal control on centrally or distributed data.

2.4 User tools

Our goal here is not to provide an exhaustive list of all existing workflows tools for final users. We focus on scientific workflows and we include a few as examples, but there are many others.

Taverna[2] is a strongly typed bioinformatics workflow management system developed by the European Bioinformatics Institute and the University of Manchester. It aims to provide a language and software tools to facilitate the use of workflows and distributed computing in the scientific community. The Taverna suite includes the Taverna Engine that powers both the Taverna Workbench and Server which allows remote execution of workflows.

Kepler[3] is a generic science oriented workflow system (ecology, bioinformatics, geology...), which would tend to be universal. It is based on Ptolemy II system developed by researchers at the University of California at Berkeley and collaborators. A set of actors is defined and their performances are under the supervision of one or more directors who determine the semantics to apply to the links between the actors.

Triana[4] is a workflow system originally built to provide a tool for rapid analysis of data from gravitational waves. At the beginning, the procedures were modelled and executed locally or remotely using RMI. Recently, Triana has been extended to incorporate components that are distributed, grid computing-oriented or Web Services oriented.

MyExperiment[5] is a social networking site for workflow exchange and sharing, with 3000 members and 1000 workflows representing 10 workflows management systems. As in the case of Taverna, this Virtual Research Environment is mainly used by bioinformatics, enabling users to upload and find publicly shared workflows, promoting building of communities, forming of relationships and collaboration.

[2] http://www.taverna.org.uk
[3] https://kepler-project.org
[4] http://www.trianacode.org
[5] http://www.myexperiment.org

3 Related initiatives

3.1 ESO Reflex

ESO Reflex[6] [5] [6] is a graphical workflow system for running ESO reduction recipes and related tools in a flexible manner. It was initially developed within the SAMPO[7] project, which performed a feasibility study for the integration of ESO pipeline processing tools with the Taverna 1 workflow engine. It allowed the user to define and execute a sequence of recipes using an easy and flexible GUI. Instead of running the recipes one at a time, a sequence of recipes can be run as a workflow where the output of a recipe is used as an input to another recipe. It was focused on ESO pipelines for astronomical data reduction. The power of the workflow as an entity encompassing the tasks typically assigned to scripts, combined with the additional semantics which actual encode the data reduction recipes, have made ESO continue the incarnation of ESO Reflex, this time based on the Kepler[8] workflow engine.

3.2 AstroGrid

AstroGrid[9], the UK's Virtual Observatory System, developed the AstroGrid Workflow System [21], a multi-user batch system for the execution of potentially long-running astronomical workflows. A file describing which remote applications — data collections and processing packages — are going to be used, is needed in order to enact and execute the workflow. These applications may be distributed throughout the VO, some of them may be implemented in CEA servers. The CEA (Common Execution Architecture) [7] [9] defines the Web service interfaces, message protocols, and formats that an executable application must support in order to be fully compliant with VO standards. The results and intermediate products of the workflow are stored in MySpace.

AstroGrid also developed a version of the Taverna 1 workbench with VO plug-ins [1] [20], which added a number of significant capabilities. The AstroGrid implementation of Taverna relies on the Astro Runtime, a client side library of functions to access the Virtual Observatory.

3.3 VO France & CDS

[6] http://www.eso.org/sci/software/sampo/reflex/
[7] http://www.eso.org/sci/software/sampo/
[8] https://kepler-project.org
[9] http://www.astrogrid.org

A Workflow working group[10] [19] has started to work in 2005 in the frame of OV France. The aim was to provide use cases and to implement them as workflows with a (VO enabled or other) workflow tool. The CDS has developed AÏDA (Astronomical Image processing Distribution Architecture) during the MDA (Masses de Données en Astronomie) French Ministry funded project and the European VOTECH project.

AÏDA has 2 sides, one at the server level to execute a workflow and one at the user level, as it provides a graphical composition tool based on JGraph. This tool is able to validate the data (FITS images) before the execution at each step of the workflow through the IVOA Characterisation implementation.

3.4 Helio-VO

The HELIO-VO[11] project is a domain-specific virtual observatory for solar physics that has been built, not only with data access and sharing in mind, but with the actual description of the knowledge in the field (via ontologies), and their processes (via workflows). One of its main achievements is having enabled Taverna to run on Grid or Cloud based resources, thus greatly expanding its potential in Astronomy. Processing and storage services allow the users to explore the data and create the products. These capabilities are orchestrated with the data and metadata services using the Taverna workflow engine.

3.5 CyberSKA

CyberSKA [12] is a project aimed at exploring and implementing the cyber-infrastructure that will be required to address the evolving data intensive science needs of future radio telescopes such as the Square Kilometre Array. They are developing a web based workflow builder that supports image segmentation, image mosaicking, spatial reprojection, and plane extraction from data cubes. These actions and processes contained in the workflow are provided as web services, which automatically determine the most efficient course of action regarding where data is to be retrieved from and processed.

3.6 Wf4Ever

The EU FP7 funded project "Wf4Ever: Advanced Workflow Preservation Technologies for Enhanced Science"[13] [11] [13] [14] started in December 2010 with the main intend to contribute to the development of standards and models for the preservation of scientific workflows. Wf4Ever considers complex digital objects (Research Objects) that include workflow models, the provenance of their

[10] http://www.france-ov.org/twiki/bin/view/GROUPEStravail/Workflow
[11] http://www.helio-vo.eu
[12] http://www.cyberska.org
[13] http://www.wf4ever-project.org

executions, and interconnections between workflows and related resources. This project will investigate and develop technological infrastructure for the preservation and efficient retrieval and reuse of scientific workflows in a range of disciplines, including Astronomy.

On-going efforts to improve seamless building of scientific workflows in the Astronomy domain have produced the first ready-to-install AstroTaverna [14] plugins [2] [3]. AstroTaverna developments have triggered the interest of a similar initiative carried out in the VAMDC Consortium[15].

Among the most important features are:

- Access to Virtual Observatory Registry
- Access to ConeSearch, SIA and SSA VO Services
- Efficient visualization of VOTables data and description
- VOTable data manipulation, extraction and filtering
- SAMP connectivity to other astronomical software

Upcoming versions will cover access to TAP and CDS SOAP Services, execution of Aladin scripts and tasks via local services, as well as other local services allowing files format conversion and standard astronomical functions. Moreover, because astronomers are found to be heavy users of Python scripting language, some exploration studies are considering to provide Taverna users the possibility to add their own Jython beanshells when building Taverna workflows.

The work undertaken in the frame of the Wf4Ever project could foster the development of astronomical workflows, favoring the use of Virtual Observatory standards for interoperability among astronomical data and process.

3.7 Pegasus

Pegasus[16] is a highly fault tolerant workflow management system that runs workflow applications in many different environments including desktops, campus clusters, grids, and now clouds. In a workflow application, the output from one component becomes the input to another component, as in a pipeline application. Pegasus enables scientists to construct workflows in abstract terms without worrying about the details of the underlying execution environment. This is what makes it so powerful as a science tool, and why it has found applicability in many fields, including astronomy, bioinformatics, earthquake science, climate modeling and others.

[14] http://wf4ever.github.com/astrotaverna
[15] http://www.vamdc.eu
[16] http://pegasus.isi.edu

3.8 Montage

Montage[17] is a toolkit for assembling Flexible Image Transport System (FITS) images into custom mosaics. This toolbox of components has been well studied in computer science workflow systems, and is used in a number of production astronomy systems.

3.9 ER-Flow

The ER-flow[18] EU FP7 funded project aims to build a European Research Community through interoperable workflows and data sharing. The project targets major research communities that use workflows to run their experiments on a regular basis (Astrophysics, Computation Chemistry, Heliophysics and Life Sciences), and collaborates with the National Grid Infrastructures through EGI.eu.

The targeted research communities select workflows which can be used as pilots to demonstrate how to develop, use and share workflows. The project will port these pilot workflows to a simulation platform and publish them in a workflow repository. ER-flow will also collect and analyse requirements of the supported research communities towards interoperability of scientific data in the workflow domain.

ER-flow may be seen as the continuation of already finished "SHIWA: Sharing Interoperable Workflows for large-scale scientific simulations on Available DCIs"[19] EU FP7 funded project.

4 Workflows preservation

The preservation of workflows as complex digital experiments is an important issue where methodology, processes and data need a common preservation strategy in order to achieve reproducible procedures and repeatable results through large periods of time.

Workflows and their components, as digital entities, need specific applications to be interpreted and re-executed. These, in turn, need specific libraries installed on a specific operating environment**,** which runs on very specific hardware configurations for which drivers are provided. All of these factors combine to ensure that workflows are severely vulnerable to obsolescence: if any of the layers in the dependency tree is lost, the entire object ceases to be accessible and usable. In this context, Virtual Machines have been considered as a method for capturing a workflow in an executable, mostly portable form. But there could

[17] http://montage.ipac.caltech.edu
[18] http://www.erflow.eu
[19] http://www.shiwa-workflow.eu

also be vulnerabilities regarding the interpretation of workflows and data, documenting their provenance and limitations, and ensuring that they are trustworthy.

As a first approach to preservation of workflows we can consider the basic steps for software preservation: preserve, retrieve, reconstruct and replay. [12] For retrieval, in addition to knowledge of general software architecture, there is a need for explicit information on the software's functionality. With reconstruction there is a need for understanding the dependencies and components, details on program language and the libraries required to ensure the correct output. Replay will also need sufficient documentation and might be used as a benchmark to assess the success of the preservation method.

We should consider the preservation of all digital entities involved in a workflow, taking into account the provenance of the final results, which is especially complex in a cloud of services. Given a predicted rise in the number of openly available web services and workflows, it would seem necessary, to curate processes as effectively as we curate the data they consume and the publications they generate. We should be able to find a workflow or process based on what it does, what it consumes as inputs and produces as outputs, and find copies or similar services usable as alternates [10].

Other issues to be considered are permissions and licenses concerning infrastructure requirements or proprietary data, versioning of workflows and of its components, classification and indexation in semantic repositories for them to be retrievable, referenced and acknowledged.

5 Workflows in the VO

Unlike traditional pipelines, which tend to produce scientifically exploitable results, most of the scientific workflows in the Virtual Observatory should be aimed at producing scientific insight. They should be easily accessible to a wide range of non-highly specialised technical users, allowing an effortless design, composition and execution. The complete digital characterization of workflows should describe the scientific methodology used in an experiment in its entirety.

The classical vision of a workflow as the orchestration of tools and tasks running either locally, on a cluster, or on a grid may be greatly improved if considering the VO as an infrastructure of web services and data.

VO services could be used as components for internet-based workflows. Since their execution is independent of the investigator's platform, they ensure the reproducibility of the results and their dissemination given their modularity, and their universal availability.

5.1 Distributed data analysis workflows

In this case a user or a client defines and executes a distributed workflow, which invokes services on multiple remote sites via the VO infrastructure. The workflow would be entirely in VO-space, driving simpler services at the individual sites.

The AstroTaverna developments provide a graphical tool for the composition and design of workflows based on VO services and data from different archives and facilities.

5.2 Data processing pipelines

Traditional data processing pipelines, e.g., instrumental or survey data processing pipelines, which produce higher, level data products. At present there are many variants of these and they have little or no direct connection to VO, aside from possibly producing VO-compliant data or being optionally driven from VO.

It is not clear how much VO mechanisms are needed at this level (VO compliant data and metadata, modelling provenance, etc.)

5.3 Driving data processing pipelines from VO

In this case we have a traditional data processing pipeline and the remote user or client software invokes a job to do some pipeline reprocessing, e.g., to custom reprocess an instrumental dataset to produce a new image, cube, etc. The "workflow" in this case runs at a single site, and VO is used to drive the job remotely (SSO, UWS) and manage the results (VOSpace, VO data services).

We could think on integrating the traditional data processing pipelines we already have with VO, to allow VO users to do on-the-fly reprocessing to generate data products which can be analyzed with VO (custom reprocessing of observatory data for example)

Some attempts to integrate general processing applications have been made with CEA and UWS.

6 Workflows and IVOA standards

Several contributions [4] [16] [17] [18] have been presented in the past in VO related forums, as well as in astronomy data analysis conferences [15] [19]. We provide below a short enumeration of IVOA standards and ongoing works susceptible to contribute to the development of standards for workflows.

- **Data Modelling**: Characterisation, Provenance
- **Semantics**: UCD, Ontologies, Vocabularies
- **Grid and Web Services**: UWS, VOSpace, SSO
- **Theory**: Self-descriptive Web Services
- **Data Curation and Preservation**: Permanent identifiers
- **Applications**: SAMP, Workflow Management Apps

7 Knowledge Discovery in Databases

Towards Virtual KDD Workflow Web-based Warehouses

In the KDD context a workflow is a precise and well-codified description of the multi-step process, which is needed to execute and supervise multiple tasks, acting like a sophisticated scripting resource. Each task represents the execution of a computational process, such as running a program, submitting a query to a database, submitting a job to a distributed computing infrastructure, like for instance cloud or grid platform, or simply invoking a web service as a remote resource.

In the data mining practice workflows are a powerful way to systematically, iteratively and accurately run complex data mining procedures: managing dataset and meta-data creation, feature subset extraction, normalization, machine learning and validation of data, safe and efficient archiving of output data, data comparisons across repeated runs and finally regular and incremental update of data warehouses.

Furthermore, in the VO context, standardized workflows could be helpful to gather and aggregate data from distributed datasets and data-generating algorithms, to engage multi-epoch and multi-band comparative astrophysics. Moreover, beyond data assembly, workflows may codify data mining and knowledge discovery pipelines across predictive algorithms. And last but not least, SWFs could transform the implicit multi-step processing sequence of a KDD application into an explicit and reusable along time specification over a standardized software farm and shared infrastructure.

A typical KDD SWF is based on three main components: an **execution environment**, a **visual design toolset**, and a **Software Development Kit (SDK)**.

The environment physically executes the workflow on behalf of applications and handles common computing concerns, including (i) invocation of the service

applications and handling the heterogeneity of data types and interfaces; (ii) monitoring and recovery procedures from failures; (iii) optimization of memory, storage, and execution nodes, including concurrency and resource sharing; (iv) general data handling, for instance mapping, referencing, normalization, streaming, and staging; (v) logging of process status and data production tracking; and (vi) monitoring of access policies for security.

A crucial aspect of SWFs for data mining is that they must be able to handle long-running processes in volatile environments and thus must be able to achieve asynchronous interaction with users, robust and capable of fault tolerance and recovery. They also need to evolve continually to harness the growing capabilities of underlying computational and storage resources, delivering greater capacity for analysis.

In such dynamical context we believe that the use of virtual machines, i.e. software drivers able to virtualize the underlying computing infrastructure to the high-level software workflow applications, could provide an efficient and easy way to exploit hybrid distributed processing platforms, by also minimizing the technical knowledge about their configuration and use by astronomers.

For what KDD is concerned, the design toolset in the VO SWFs should provide visual scripting applications for authoring and sharing KDD workflows and setup the components that are to be incorporated as executable steps. The aim is indeed to minimize the complexity of the underlying applications and enable users to design and fully understand workflows without commissioning specialists or hiring software engineers. This for sure could empower astronomers to build their own pipelines in an autonomous way.

As a matter of fact web 2.0 technologies (for instance web applications) offer the best solution to provide such SWF scripting design tools, mainly because they do not require local computing resource by permitting in principle to build, configure and execute SWFs by a personal Smartphone or tablet.

Finally, standardized SDKs enable developers to extend the capabilities of the system and enable workflows to be embedded into applications, Web portals, or databases. This has the potential to incorporate sophisticated knowledge seamlessly into the tools that astronomers use routinely.

As concluding remarks, we recall that SWFs should offer techniques to support the new paradigm of data-centric science. In a data-centric environment, it should be as much as possible minimized the massive data flow on the network. It is indeed much more convenient and fast to move applications towards the data centers, especially if they are organized as KDD application warehouses. This of course requires a well-defined standardization process, in order to organize applications and SWFs in a fully interoperable way.

By having the possibility to share on demand applications between standardized and interoperable KDD warehouses, it may engage a virtuous mechanism in which users may operate by remote, through a simple web browser, sharing resource on the network (not data), building flexible workflows and launching them in the virtual computing cloud, by interacting with these resource in an asynchronous way (for example by exploiting the web containers based on AJAX technology). By this way SWFs can be always maintained updated, replayed and repeated. Results and secondary data can be computed as needed using the latest sources, providing virtual data (or on-demand) warehouses by effectively providing distributed query processing. The workflows themselves, as main actors in the data-centric science, can be generated and transformed dynamically to meet the requirements at hand.

8 Proposal

The quantitative leap in volume and complexity of the next generation of archives will need analysis and data mining tasks to live closer to the data, in computing and distributed storage environments, but they should also be modular enough to allow customization from scientists and be easily accessible to foster their dissemination among the community.

Astronomy is a collaborative science, but it has also become highly specialized, as many other disciplines. Sharing, preservation, discovery and a much simplified access to resources in the composition of scientific workflows will enable astronomers to greatly benefit from each other's highly specialized know-how, they constitute a way to push Astronomy to share and publish not only results and data, but also processes and methodologies.

This disruptive transformation in the way digital experiments are designed, performed, shared and preserved in Astronomy cannot be done outside the Virtual Observatory, where workflows, processes and services should benefit of the same privileges acquired by data.

References

[1] K. Benson and N.A. Walton, Euro-VO DCA Theory and Grid Workshops 2008, http://sait.oat.ts.astro.it/MSAIt800209/PDF/574.pdf
[2] Garrido, IVOA Sao Paolo Fall Interop 2012, Applications session http://wiki.ivoa.net/internal/IVOA/InterOpOct2012Applications/astrotaverna_interop_2012.pdf
[3] Garrido, IVOA Sao Paolo Fall Interop 2012, GWS session http://wiki.ivoa.net/internal/IVOA/InterOpOct2012GWS/pdl_interop_2012.pdf
[4] M.J. Graham, IVOA Garching Spring Interop 2009, GWS session http://www.ivoa.net/internal/IVOA/InterOpNov2009GWS/Garching-GWSWorkflow.pdf
[5] R. Hook et al., Euro-VO DCA Theory and Grid Workshops 2008, http://sait.oat.ts.astro.it/MSAIt800209/PDF/578.pdf
[6] P. Järveläinen et al., ADASS London 2007, http://adsabs.harvard.edu/abs/2008ASPC..394..273J
[7] P. Harrisson, *A Proposal for a Common Execution Architecture*, IVOA WG Internal Draft 2005-05-12 http://www.ivoa.net/Documents/Notes/CEA/CEADesignIVOANote-20050513.html
[8] Ludäscher, B., Berkley, C., Jones, M., & Lee, E. A. *Scientific Workflow Management and the Kepler System*. Concurrency and Computation: Practice & Experience - Workflow in Grid Systems, 18, 1039 – 1065. (2006) http://users.sdsc.edu/~ludaesch/Paper/kepler-swf.pdf
[9] G. Rixon, *Introduction to CEA and UWS*, IVOA Note October 2007 http://www.ivoa.net/Documents/Notes/IntroductionCEA_UWS/IntroductionCEA_UWS-20071005.pdf
[10] J.E. Ruiz, IVOA Sao Paolo Fall Interop 2012, GWS session http://wiki.ivoa.net/internal/IVOA/InterOpOct2012GWS/CharWS.pdf
[11] J.E. Ruiz, IVOA Sao Paolo Fall Interop 2012, DCP session http://wiki.ivoa.net/internal/IVOA/InterOpOct2012DCP/ROs.pdf
[12] J.E. Ruiz, IVOA Pune Fall Interop 2011, DCP session http://www.ivoa.net/internal/IVOA/InterOpOct2011DCP/Wf4EverPune.pdf
[13] J.E. Ruiz, IVOA Naples Spring Interop 2010, DCP session http://www.ivoa.net/internal/IVOA/InteropMay2011DCP/Wf4EverNaples.pdf
[14] J.E. Ruiz, IVOA Nara Falll Interop 2010, DCP session http://www.ivoa.net/internal/IVOA/InterOpDec2010DCP/Wf4Ever.pdf
[15] A. Schaaff et al., ADASS Paris 2011- Scientific Workflows in Astronomy BoF http://www.eso.org/sci/php/meetings/adass2011/html/display.php?topic=BoF_Schaaff_1314702958.html
http://amiga.iaa.es/FCKeditor/UserFiles/File/WfsADASS.pdf
[16] A. Schaaff and F. Bonnarel, Baltimore IVOA Fall Interop 2008, Application session 2008, http://www.ivoa.net/internal/IVOA/InterOpOct2008Applications/GWS-DM-REG-301008.pdf
[17] A. Schaaff et al., Euro-VO DCA Theory and Grid Workshops 2008, http://sait.oat.ts.astro.it/MSAIt800209/PDF/559.pdf
[18] A. Schaaff, Trieste IVOA Spring Interop 2008, GWS sessionhttp://www.ivoa.net/internal/IVOA/InterOpMay2008GridAndWebServices/GWS-Charac-Workflow-20May08.pdf
[19] A. Schaaff et al., ADASS London 2007, http://adsabs.harvard.edu/abs/2008ASPC..394...77S
[20] N. A. Walton et al, ADASS London 2007 http://adsabs.harvard.edu/abs/2008ASPC..394..309W
[21] N. Winstanley, *Design and Implementation of the AstroGrid Workflow system,* IVOA Note February 2006 http://www.ivoa.net/Documents/Notes/AstrogridWorkflow/AstrogridWorkflow-20060227.pdf