

Activités du groupe de travail Workflow

André Schaaff, CDS

Franck Le Petit, Observatoire de Paris Meudon

Philippe Prugniel, Observatoire de Lyon

Eric Slezak, Observatoire de Nice

Christian Surace, Laboratoire d'Astrophysique de Marseille



Un workflow ?

■ Le plus trivial

- Un enchaînement de tâches
- Exemple : script exécutant successivement plusieurs applications

■ Plus compliqué

- Enchaînement de tâches avec synchronisation, branchements conditionnels, etc..

■ Dans le cadre de l' Observatoire Virtuel

- Le cas précédent mais avec prise en compte des standards du VO

Un domaine très prolifique...

■ Workflow &

■ & langages

- AGWL, BPEL4WS, BPML, DGL, DPML, GJobDL, GSFL, GFDL, GWorkflowDL, MoML, SWFL, WSCL, WSCI, WSFL, XLANG, YAWL, SCUFL/XScufl, WPDL, PIF, PSL, OWL-S, xWFL, &

■ & formalismes

- Petri net, UML activity diagram, BPMN, DAG, IPO, GPSG, Workflow Patterns, Pi Calculus, Finite-State Machine, Gamma-calculus, &

■ & mapping workflow abstrait vers concret

- CWG, ACWG, Grid Job Handler, GWES, &

■ & moteurs

- 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, &

■ & outils de composition/design

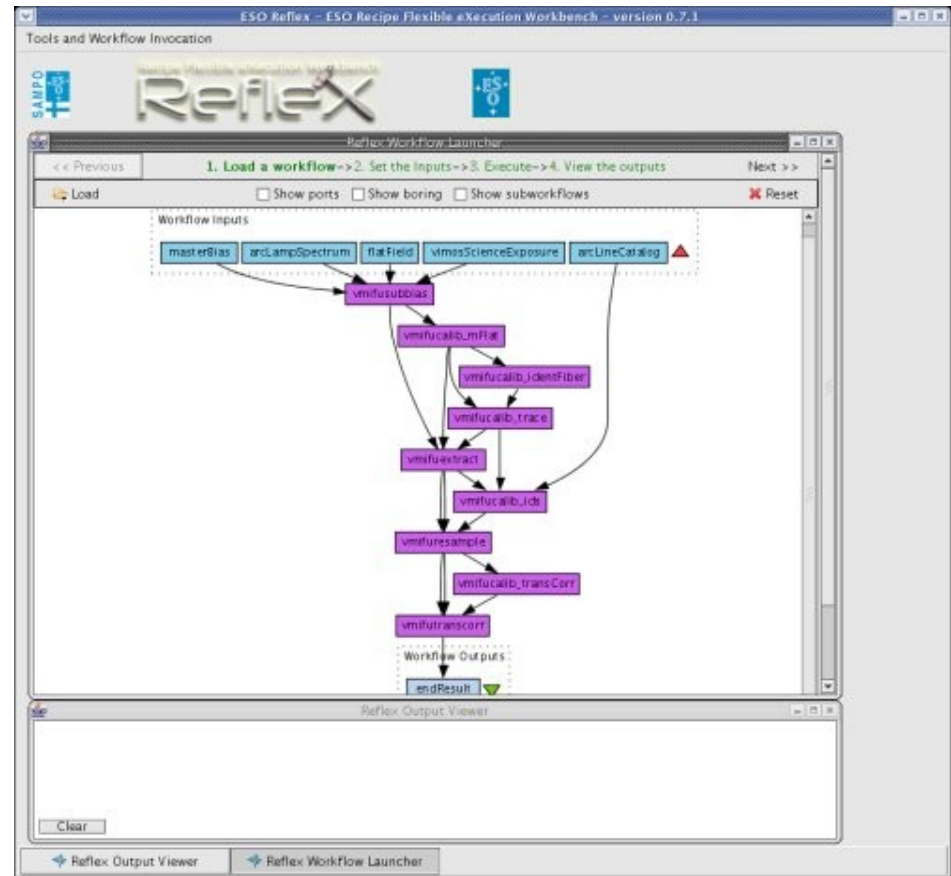
- ilog's BPMN Modeller, CAT, GWUI, XBaya GUI for Workflow Composition, Triana, JOpera, Platform Process Manager, ...

Objectifs du groupe de travail Workflow

- Dans un premier temps il était nécessaire d'affiner notre définition d'un workflow :
 - & a sequence of tasks executed within a controlled context by an architecture taking into account VO standards
- Objectifs
 - Définition de cas d'utilisation d'intérêt général dans divers domaines
 - Proposer des solutions pour définir et implémenter ces workflows
 - Définition de briques élémentaires
 - Collaborations avec d'autres OV
 - Collaboration avec la communauté STIC (expertise)
 - ...

Quelques outils en astronomie...

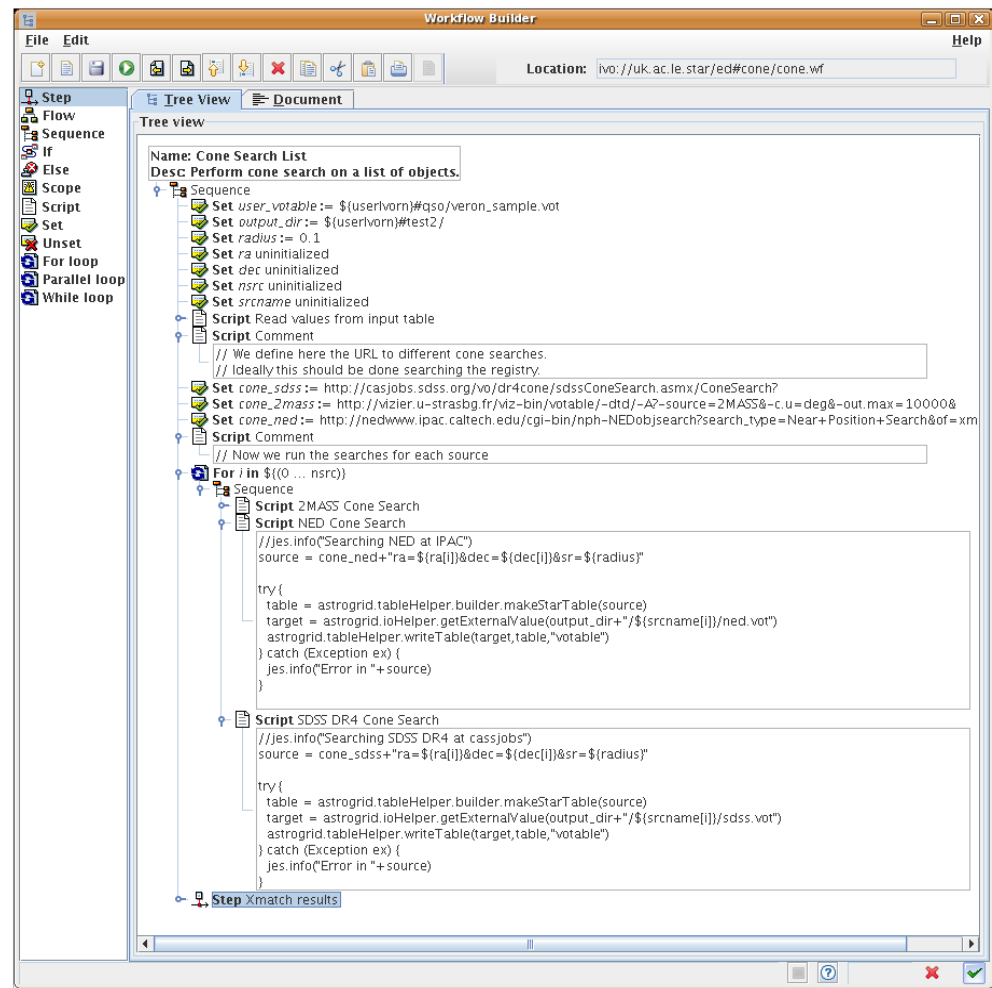
- « The ESO Recipe Flexible Execution Workbench » (initialement développé dans le contexte eScience en Bioinformatique).
- Ce projet se termine fin 2007, pas de précision sur une éventuelle intégration dans le VO.
- Discussions avec R. Hook en mars et lors d'ADASS.



<http://www.eso.org/sampo/reflex/>

Quelques outils en astronomie...

- AstroGrid : approche au travers d'un « workflow builder » dans les outils AG, basé sur un langage apparenté BPEL avec des inclusions de scripts (Groovy).
- Ils étudient actuellement Taverna.
- Discussions lors de la réunion VOTech d'Edimbourg.



Quelques outils en astronomie...

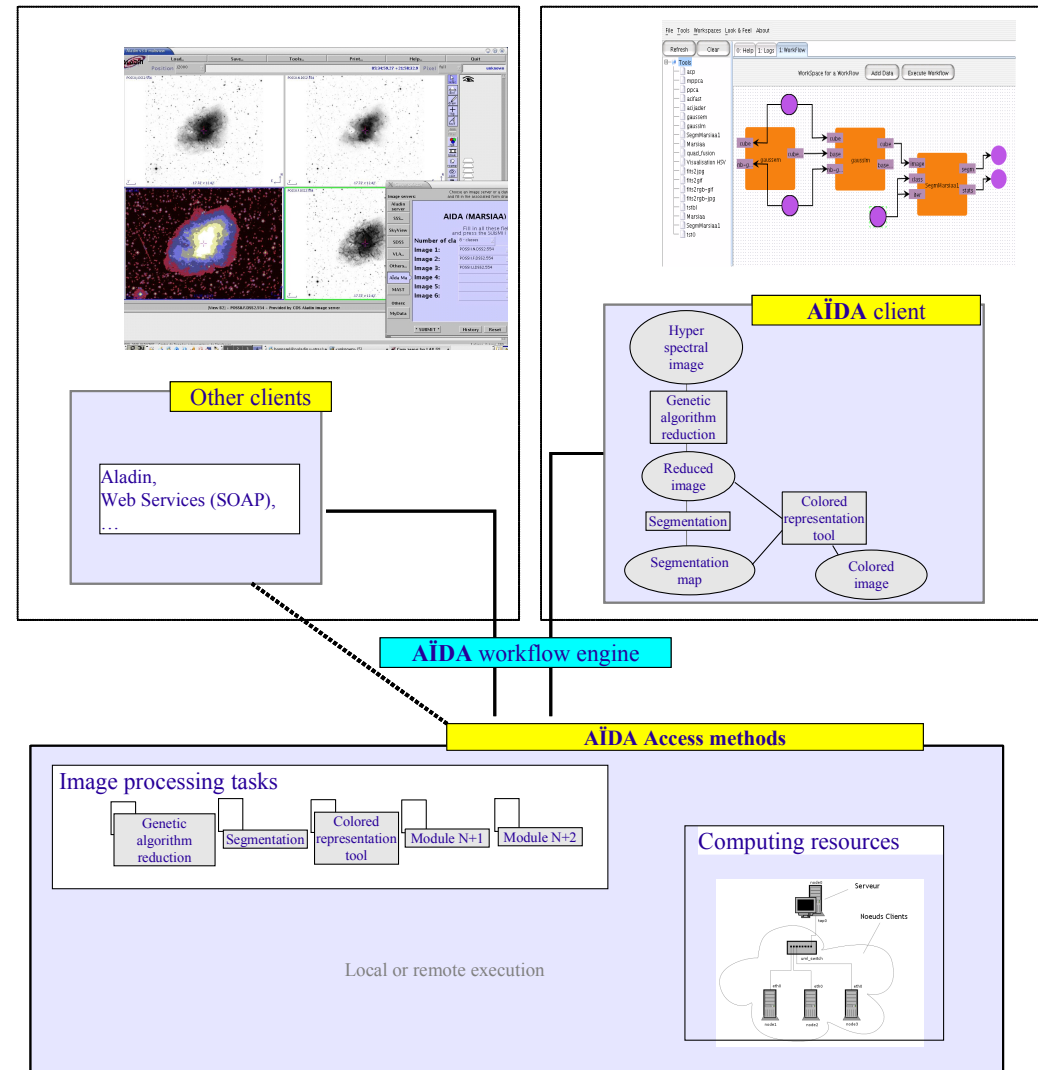
Astronomical Image processing Architecture

AIDA architecture

F. Bonnarel,
M. Louys,
A. Schaaff
CDS & LSIT

AIDA développements

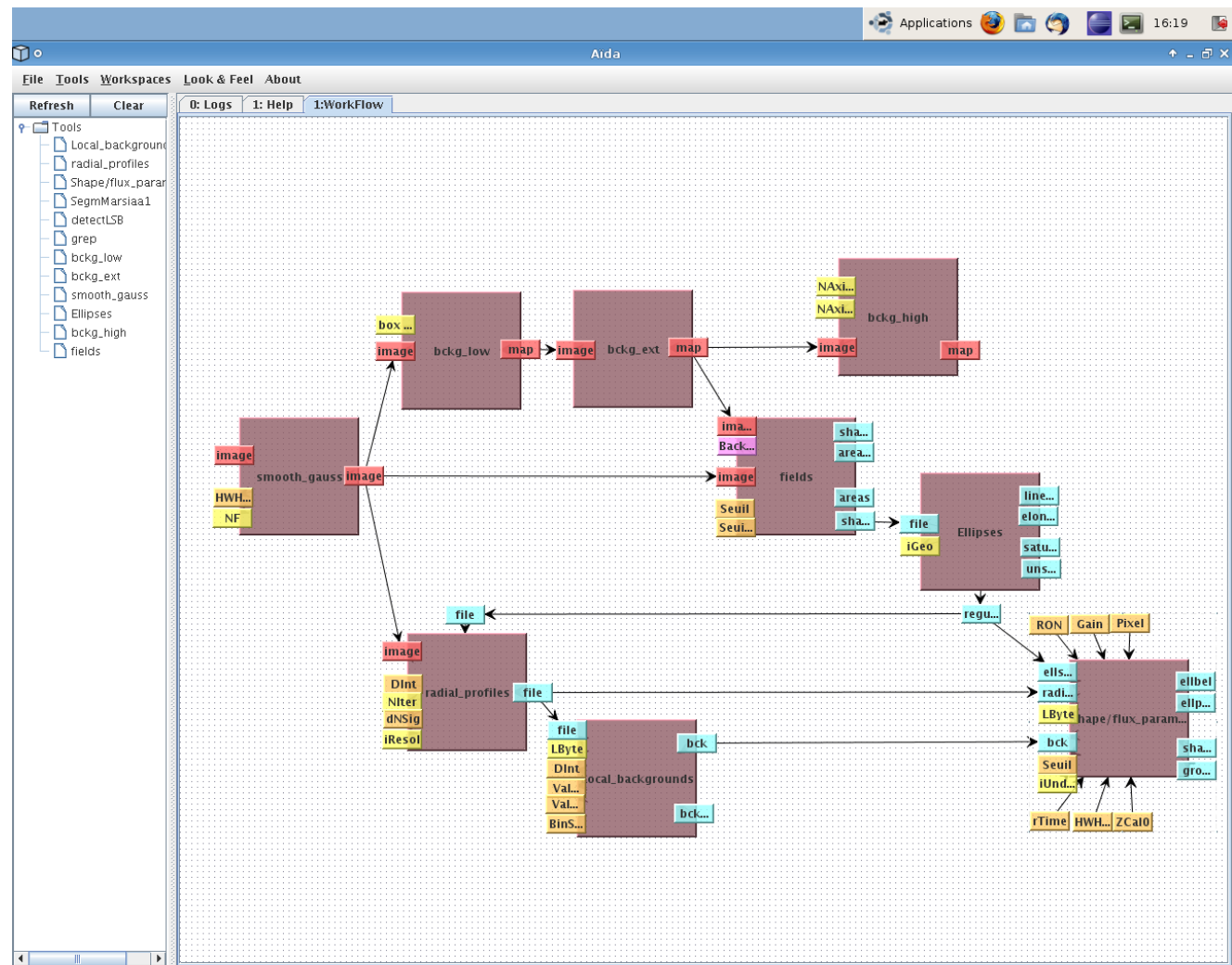
J. Beugnot
J.-J. Claudon,
G. Mantelet,
C. Pestel,
CDS



Travail réalisé dans le cadre de l'ACI Masses de Données, projet "Masses de Données en Astronomie" (2003-12/2006), OV France et VOTECH

AIDA : illustration de la partie cliente d'AIDA

Exemple de workflow créé par Eric Slezak



Cas d'utilisation et implémentation

- Le but du groupe de travail n'est pas de promouvoir un outil particulier
- Dans un premier temps nous avons décidé d'utiliser AïDA
 - Connaissance des sources, ajout de nouveaux types de données facilité, installation simple, ...
 - Standards VO pris en compte progressivement
 - Prototype de démonstration
 - &

Origine d'AIDA

- Il existe de nombreux outils de traitement d image dans divers langages (LSIIT : Matlab, C++, & , OCA : fortran, &)
 - Comment les rendre utilisables par d autres ?
 - Comment pérenniser ces outils ?
- AIDA est un environnement permettant
 - de standardiser l accès à ces outils (via CGI par exemple)
 - de les exécuter soit unitairement soit via la construction des workflows &
- Peu de contraintes
 - Un outil est potentiellement candidat s'il est exécutable sous Linux et non interactif

Ajout d'un outil dans AïDA

Un fichier descripteur décrit l'outil, en particulier les entrées/sorties

Fichier plat...

```

tool.label=mppca
tool.comment=Réduction par la technique de mélange de PPCA
tool.matlab_function=ProjectMPPCA
input.1.type=fits:seq
input.1.label=cube
input.1.comment=Cube d'image (une image fits par bande)
input.2.type=numeric:int
input.2.label=NBC
input.2.comment=Nombre de composantes dans le mélange
output.1.type=fits:seq
output.1.comment=Cube d'image après réduction (une image fits
par bande)
tool.pathname=/local/aida/lib/OtiamTools/mppca/wrap
  
```

...ou XML

```

<tool>
  <comment>Réduction par la technique de mélange de
  PPCA</comment>
  <label>mppca</label>
  <pathname>/local/aida/lib/OtiamTools/mppca/wrap</pathname>
  <matlab_function>ProjectMPPCA</matlab_function>
  <input id="1">
    <comment>Cube d'image (une image fits par
    bande)</comment>
    <label>cube</label>
    <type>fits:seq</type>
  </input>
  <input id="2">
    <comment>Nombre de composantes dans le
    mélange</comment>
    <label>nbc</label>
    <type>numeric:int</type>
  </input>
  <output id="1">
    <comment>Cube d'image après réduction (une image fits
    par bande)</comment>
    <type>fits:seq</type>
  </output>
</tool>
  
```

+ ajout dans la liste des outils => l'outil devient disponible dans AïDA

AIDA : travaux en cours

- Visualisation de la progression des traitements
- Remontée des erreurs (très dépendant des outils)
- Simplification de l'installation
- Standards VO (IVOA) (réalisation en cours dans le cadre de VOTECH)
 - Dans un premier temps une approche pragmatique avec intégration des standards suivant les besoins
 - Accès au « VOSpace » du CDS (basé sur iRODS/SRB)
 - Etude de l'application du standard IVOA « Characterisation » aux entrées/sorties des briques d'un workflow
 - UWS : Universal Worker Service pour les travaux asynchrones

Illustration

Quelques cas d'utilisation présentés dans le cadre du groupe de travail



Cas d'utilisation : Image processing

■ operation

- detection and evaluation of related objects in 1 band image

■ subjacent model

- diffuse disjoined tasks in emission on a bottom slowly variable without defects

■ method

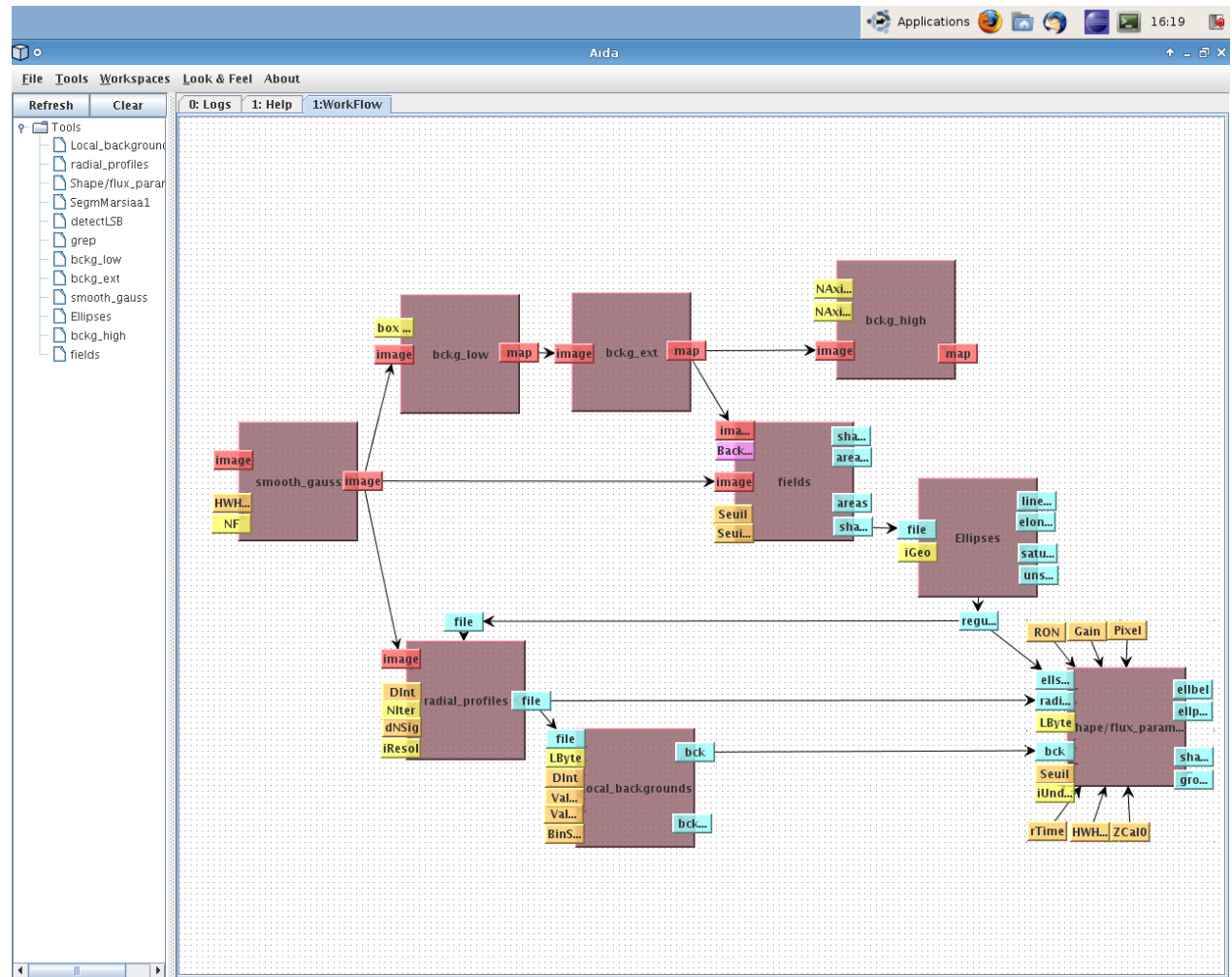
- cartography of the background
- thresholding by segmentation
- adjustment of an ellipse of form
- evaluation of the azimuth profile of brightness
- calculation of measurements of form and flow

E. Slezak
*Observatoire de
la Côte d'Azur*

...ce n'est qu'une partie d'un cas d'utilisation en cours de définition

Exemple d'implémentation du cas d'utilisation images

Use case
E. Slezak
Observatoire de la Côte d'Azur



Cas d'utilisation : Simulation

1 - PDR code

modélisation of diffuse clouds
gives theoretical absorption spectra

2 - Stellar spectra databases

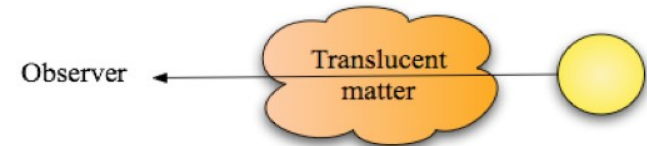
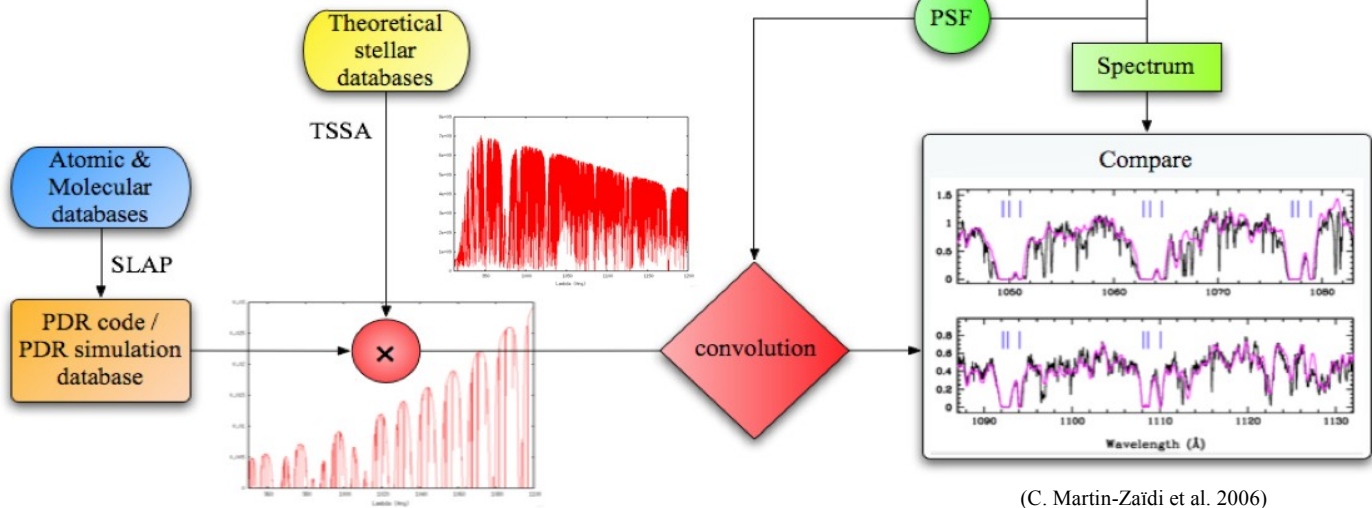
provides synthetic stellar spectra

3 - FUSE database

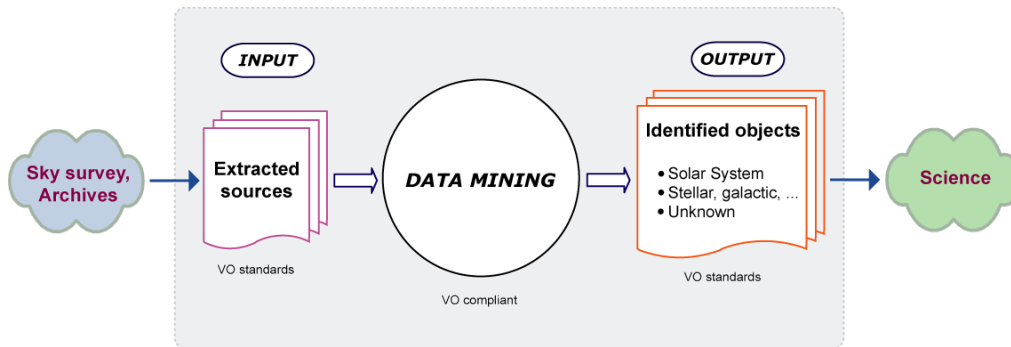
provides observations & PSF

4 - Compare directly models to observations

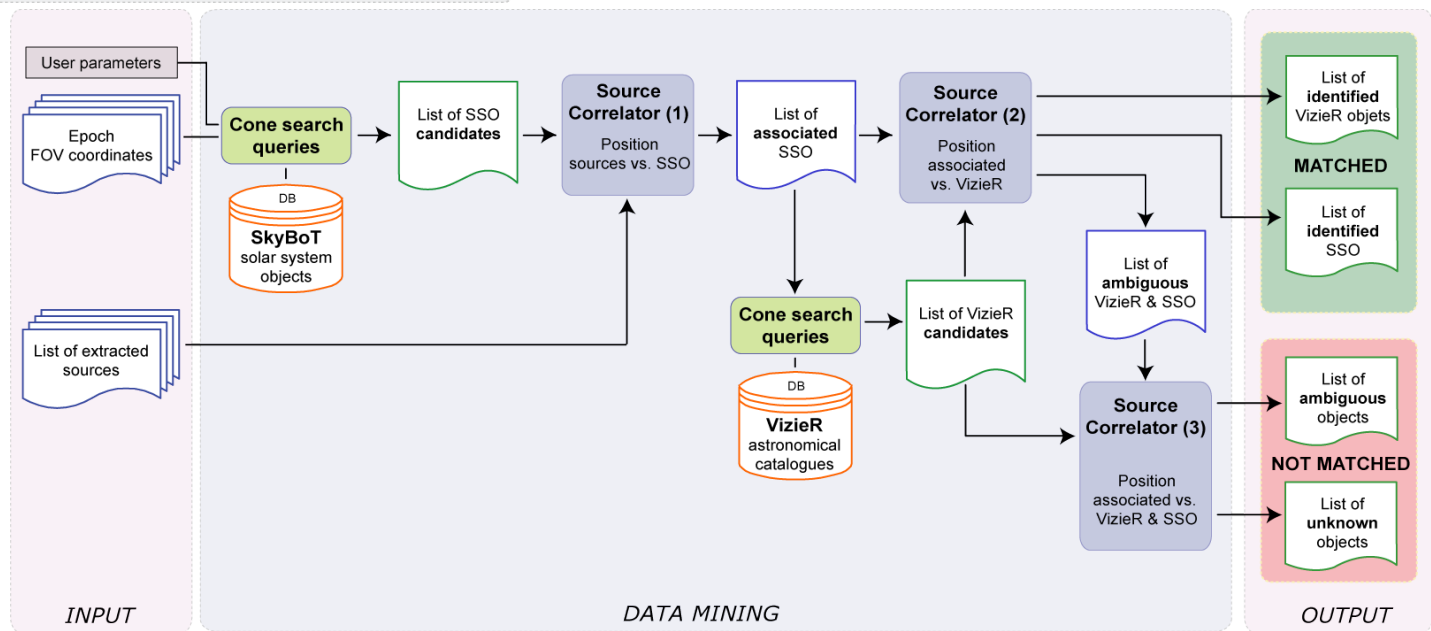
F. Le Petit et al.
*Observatoire de
Paris-Meudon*



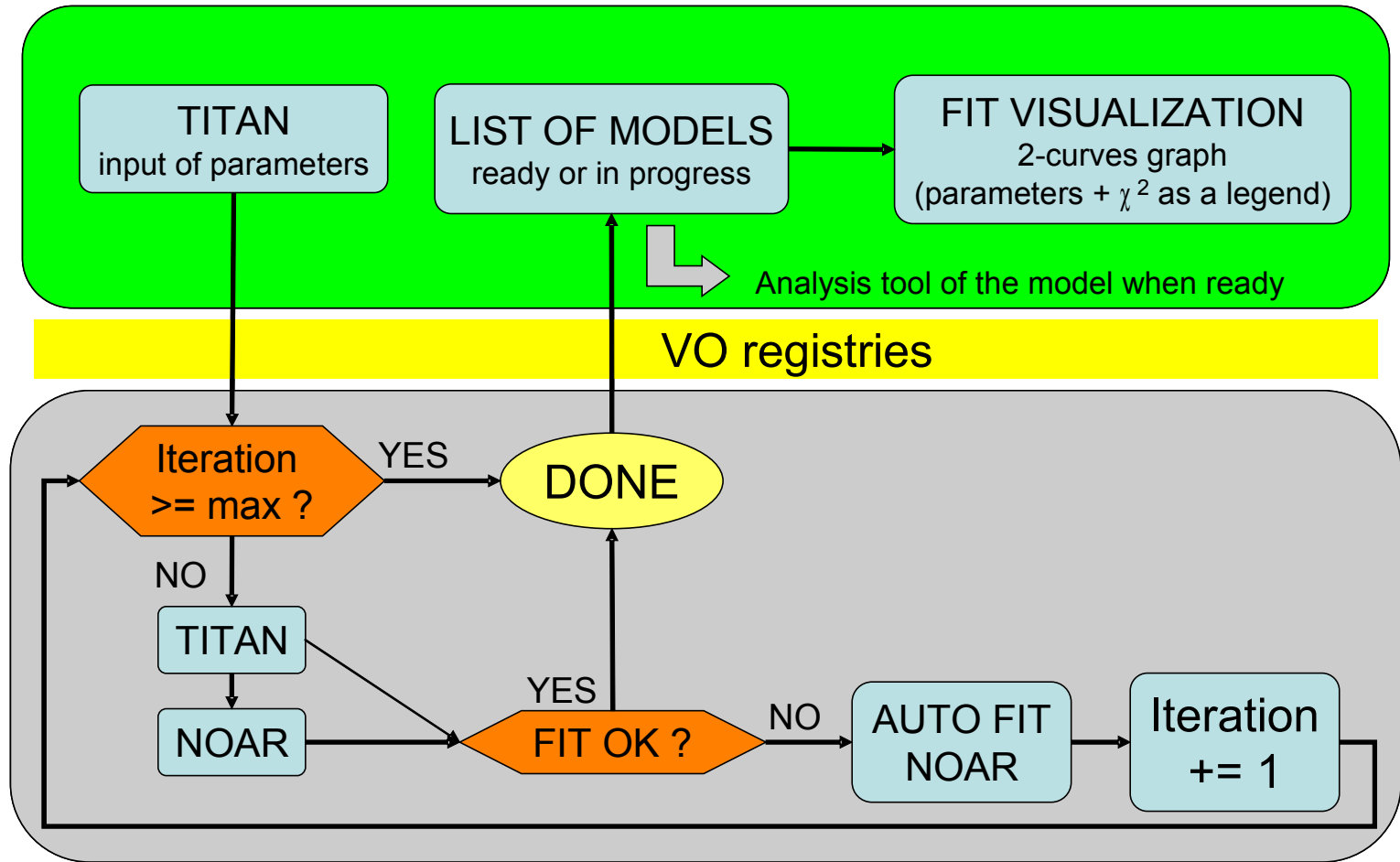
Cas d'utilisation : Data Mining



J. Berthier et al.,
IMCCE



Cas d'utilisation pour des codes TITAN et NOAR



L. Chevallier
Obs. Paris
Meudon

Workflow builder plugin dans Aladin

Basé sur les
bibliothèques
graphiques et le
moteur d'AIDA

C. Pestel
T. Boch
CDS

The screenshot displays the Aladin workflow builder v. 0.3 - Using JLOW library. The interface includes a menu bar (File, Run, View, Help), a task list on the left, and a central workspace for building workflows. The workflow diagram shows the following steps:

- Input:** A green circle labeled 'm51'.
- Task 1:** A green box labeled 'SIMBAD' with 'Load Simbad Data' below it.
- Task 2:** A green box labeled 'NED' with 'Load NED Data' below it.
- Task 3:** A yellow box labeled 'XMatch' with '1' and '2' in blue boxes on its left side.
- Task 4:** A second yellow box labeled 'XMatch' with '1' and '2' in blue boxes on its left side.
- Task 5:** A green box labeled 'Export Plane' with a right-pointing arrow.

Arrows indicate the flow: 'm51' feeds into both 'SIMBAD' and 'NED'. 'SIMBAD' and 'NED' feed into the first 'XMatch' box. The first 'XMatch' box feeds into the second 'XMatch' box. The second 'XMatch' box feeds into 'Export Plane'. Below the workflow diagram, the following command sequence is shown:

```
NED.m51.14.0'-get NED m51 14.0'
Simbad.m51-l-get Simbad m51
XMatch-1=xmatch 2 1 4
XMatch-2=xmatch 2 3 4
sync: export "XMatch results2" toto
```

The background shows the Aladin v4.0 interface with a star field and various toolbars.

Travaux en cours (1)

■ Workflow et Characterisation (IVOA)

■ Buts :

- Caractériser les entrées/sorties de chaque outil.
- Vérifier la cohérence entre les images candidates et celles que l'outil attend en entrée
 - ▶ En entrée, une image et un fichier « Characterisation », ce dernier est confronté aux contraintes (liste de règles) définies pour l'outil.
- Assurer le bon "fonctionnement" du workflow.
- => une sorte de compilation du workflow

■ Des retombées

- Optimisation, normalisation des images en entrées
- Applications à d autres types de données.
- ...

Travaux en cours (2)

- Développement de briques (extraction of subparts in n dimensions hypercube, separation of structures in n dimensions hypercube, resolution adaptation for imagery and spectrum, multi components fit, image addition, spectrum addition, ...)
- Interopérabilité entre les langages et outils de workflow
 - Travail démarré dans le cadre de VOTECH avec des collègues d AstroGrid
- Test d autres outils
- & ainsi que la continuation de la partie « Cas d utilisation
- Démarrage d'une activité Workflow dans le cadre du groupe de travail GWS lors du prochain interop à Trieste en mai 2008

Synthèse

■ Historique

- 07/2003 - 12/2006, **Masses de Données en Astronomie** projet MDA de l'ACI MD : développement de l'architecture AIDA
- depuis 2005
 - **Création du groupe de travail Workflow d'OV France**: définition de cas d'utilisation, étude d'outils (moteur de workflow, langages, &), collaboration avec des collègues STIC,&
 - VOTECH DS3 : développement JLOW / support AIDA, présentation des travaux du groupe de travail OVF, collaborations avec d'autres équipes VO, &

■ Réunions, publications et séminaires:

- Réunions du groupe de travail, 10/11/2005, 16/06/2006, 21/12/2006, 26/04/2007
- Image Processing and Scientific Workflows in the Virtual Observatory Context, E. Slezak et al., IAU 2006 (session orale)
- Implementing Astronomical Image analysis pipelines using VO Standards, M. Louys et al., IAU 2006 (session orale)
- Workflow in Astronomy, the VO France Workflow Working Group experience, A. Schaaff et al., ADASS 2007 (session orale)
- Séminaire « Workflow et Démo AIDA » dans le cadre de VO Paris, A.Schaaff et Cyril Pestel, 24/10/2007

■ Participants et contributeurs aux réunions OV France

- Christophe Barache – Obs. Paris-Meudon, Jérôme Berthier - IMCCE, Thomas Boch - CDS, Frédéric Boone – LERMA, François Bonnarel - CDS, Loïc Chevallier – Obs. Paris-Meudon, Jean-Julien Claudon - CDS, André Csillaghy – HES Switzerland, Bernard Debray - LAOB, Jean-Michel Desert - IAP, Pierre Didelon - CEA, Marie-Lise Dubernet – Obs. Paris-Meudon, Anabela C. Goncalves - LUTH, Jesus Iglesias - IMCCE, Gaëlle Labourot - UVSQ, Martin France - CRAL, Franck Le Petit – Obs. Paris, Pierre Le Sidaner – Obs. Paris-Meudon, Mireille Louys - CDS-LSIT, Guillaume Mella - LAOG, Areg Mickaelian – BAO Armenia, Johan Montagnat - ESSI Sophia Antipolis, Nicolas Moreau - LERMA, Steven Morin - UVSQ, Jonathan Normand – Obs. Paris-Meudon Cyril Pestel-CDS, Philippe Prugniel – Obs. Lyon, Fabrice Roy - LUTH, Lena Sargsyan - Yerevan State University, Alain Sarkissian - Service d'Aéronomie, Renaud Savalle - Obs. Paris-Meudon, André Schaaff – CDS, Eric Slezak, Obs. Nice, Christian Surace – LAM, Françoise Tran Minh - LERMA Obs. Paris-Meudon, Frédéric Vachier – IMCCE, Bruno Voisin - NUI Galway, ...