

Workflows en astronomie, groupe de travail Workflow OV France, AIDA, autres outils et standards de l'OV

André Schaaff
CDS



Points abordés

- Architecture de l'OV

- Les workflows
 - Définition d'un workflow, intérêt scientifique et état de l'art
 - Quelques travaux en astronomie, focus sur AIDA 2.0
 - Groupe de travail Workflow OV France, illustrations
 - Démo workflows et Characterization

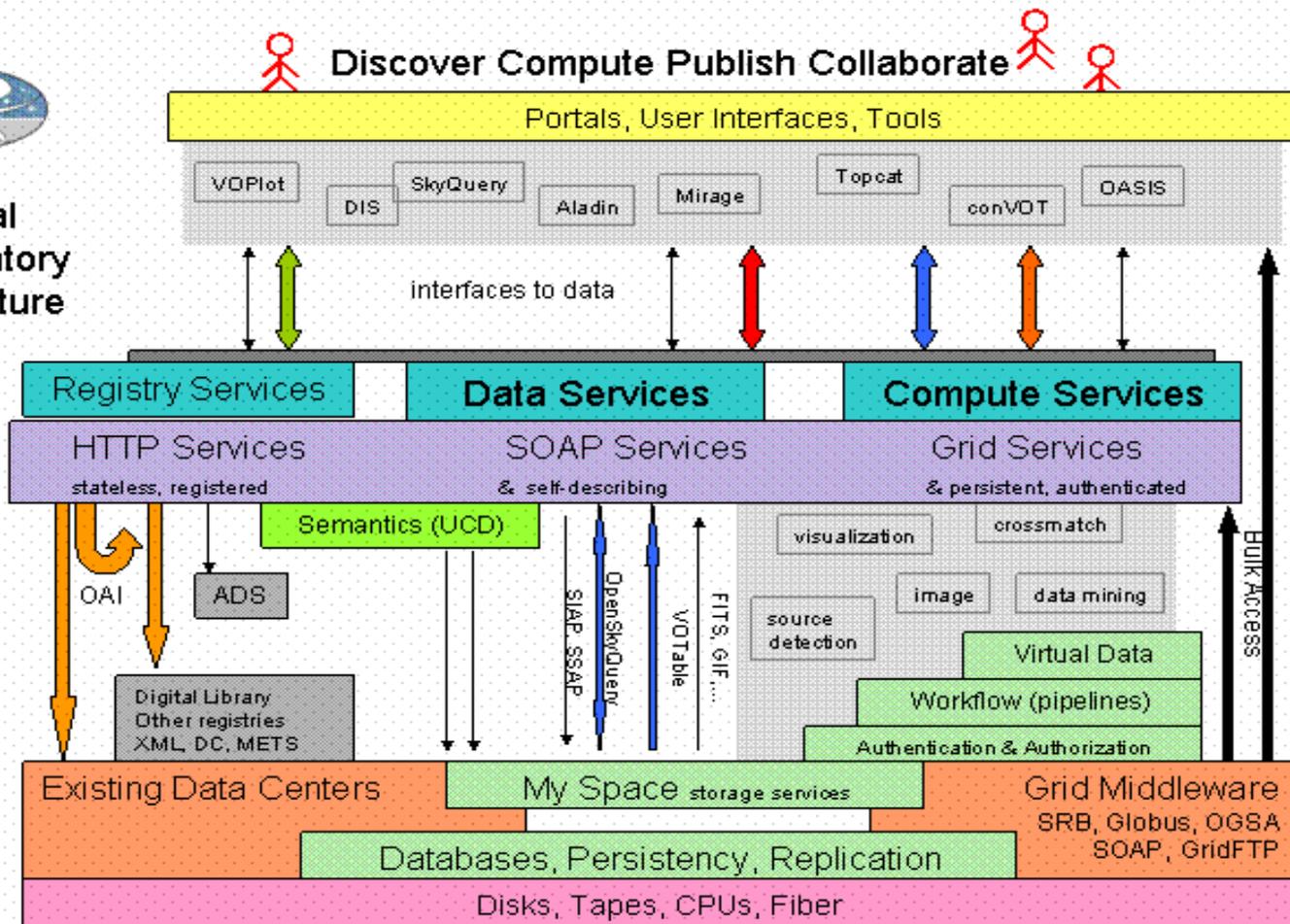
- Autres outils et applications de l'OV
 - Saada (démos création et garnissage d'une base de données)
 - Astrogrid desktop

- Questions / réponses

Architecture de l'OV



Virtual
Observatory
Architecture



Un workflow ?

■ Le plus trivial

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

■ Plus compliqué

- Un 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 de l'OV

Intérêt scientifique des workflows

- Capture de la méthodologie scientifique
- Formalisation de l'analyse scientifique
 - Routines à exécuter, data flow, détails de l'exécution, ...
- Aptitude au calcul à grande échelle
 - Découpage en briques, ...
- Favorise le travail collaboratif pour le design, l'assemblage, la validation, l'analyse, ...
- ...

Rapide état de l'art des workflows

■ Des langages

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

■ Des 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, ...

■ Des outils de composition et de design

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

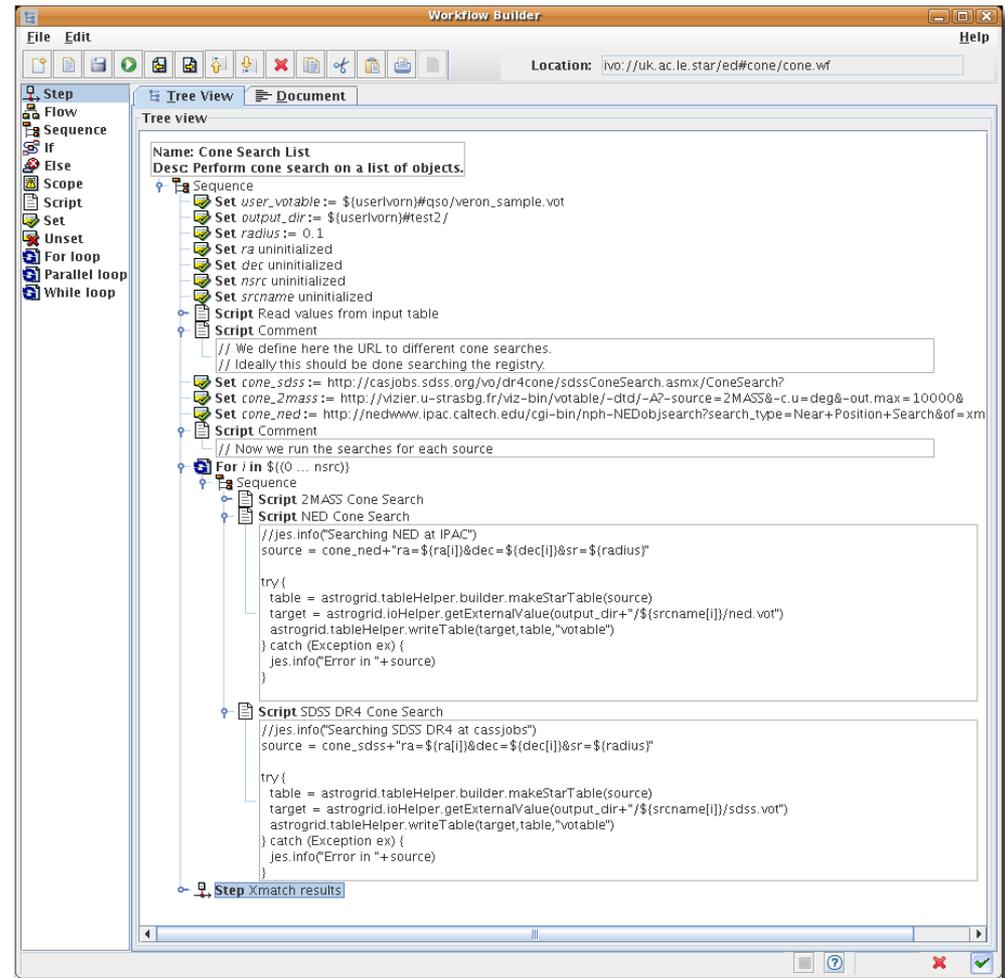
■ Des formalismes

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

Quelques travaux en astronomie...

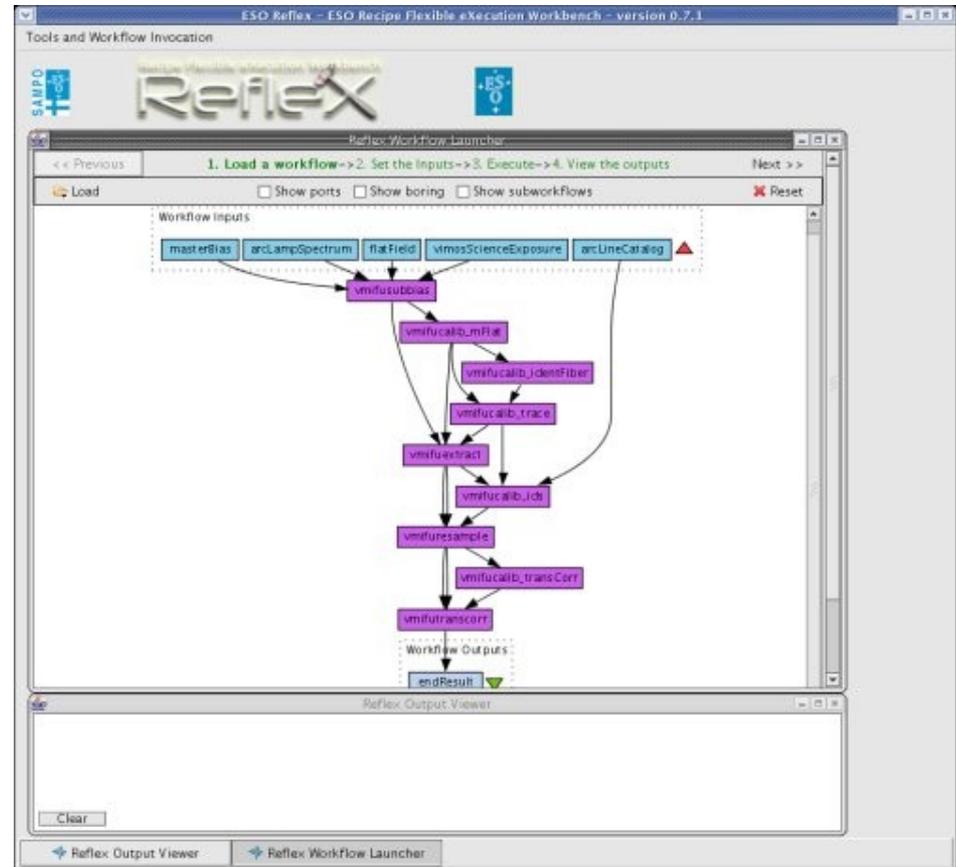
(avec une connotation OV)

- AstroGrid : approche au travers d'un « workflow builder » basé sur un langage apparenté BPEL avec des inclusions de scripts (Groovy)
- Ils travaillent actuellement avec Taverna (projet initialement développé dans le contexte eScience en Bioinformatique).



Quelques travaux en astronomie (2)

- «The ESO Recipe Flexible Execution Workbench» (basé sur Taverna).
- Ce projet s'est terminé fin 2007 mais des travaux sont en cours pour une intégration dans l'OV (interaction avec des outils de l'OV (e.g. Aladin))
- Une version packagée est disponible en ligne avec des exemples d'utilisation



<http://www.eso.org/sci/data-processing/software/sampo/reflex/>

Quelques travaux en astronomie (3)

AIDA 2.0 - Astronomical Image processing Architecture

Main contributors

O. Benjelloun, characterization integration

J. Beugnot*, packaging

F. Bonnarel, architecture & characterization

J.-J. Claudon*, core development

B. Gassmann, Camea & characterization

M. Louys, architecture & characterization

G. Mantelet*, characterization integration

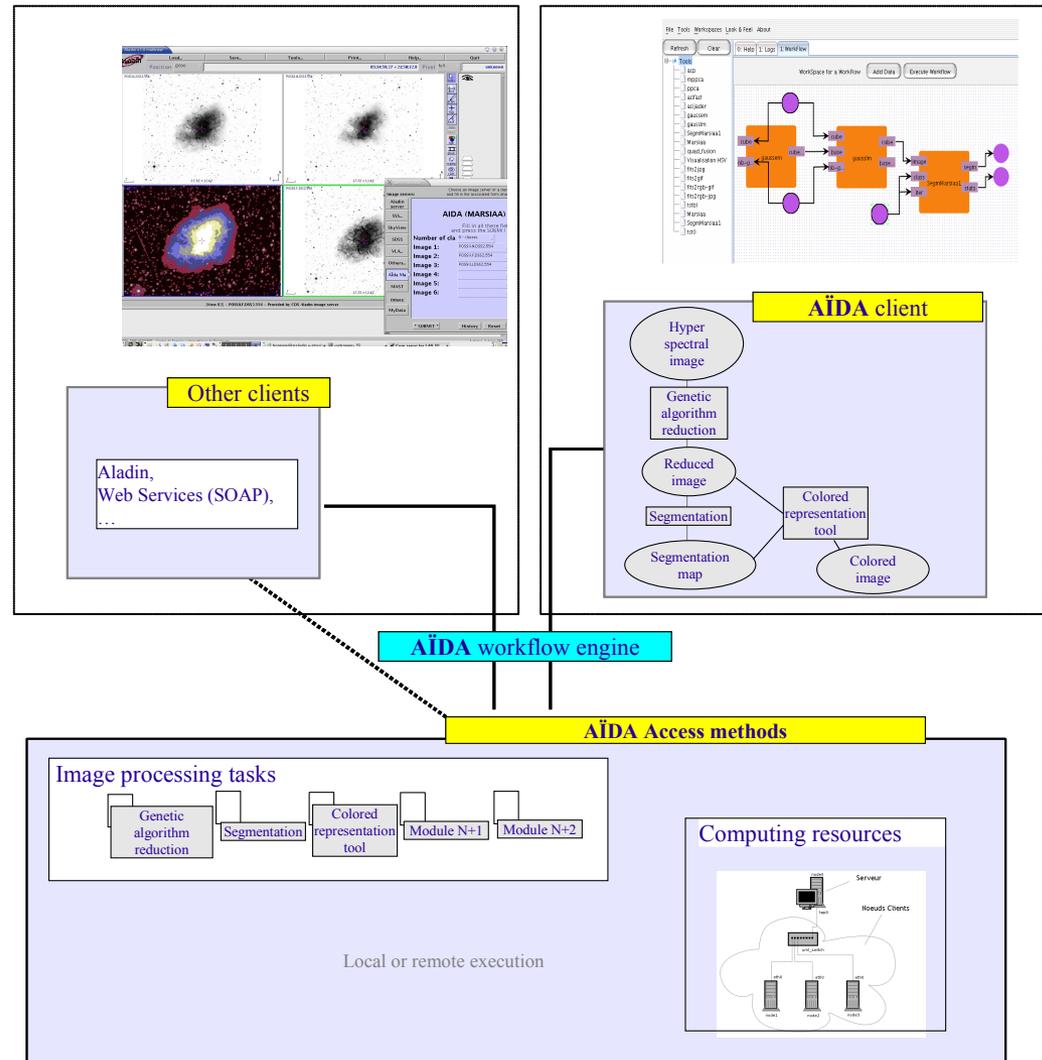
C. Pestel*, design capabilities, new developments

A. Schaaff, architecture & characterization

E. Slezak, use cases & characterization

(* have left)

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



Quelques travaux en astronomie (4)

■ Motivations à l'origine d'AÏDA 2.0

- **Nombreux outils de traitement d'image disponibles dans divers langages (LSIIT : Matlab, C++, ..., OCA : Fortran, ...)**
- **Comment les rendre utilisables par d'autres ? Comment pérenniser ces outils (souvent développés par des stagiaires, postdocs, ..., + départs en retraite, etc.. => perte de compétences)**
- **AÏDA est un environnement permettant de standardiser l'accès à ces outils (via CGI par exemple) et permettant de les exécuter unitairement ou de construire des workflows ...**
- **Peu de contraintes : outils exécutables sous Linux, MacOS, Windows et non interactifs**
- ...

Focus : AïDA 2.0 et l'ajout d'un outil

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>
  
```

Une procédure d'ajout via une page Web est en cours de validation

+ ajout dans la liste des outils => *l'outil devient disponible dans AïDA 2.0*

AIDA 2.0 : dernières avancées

- Visualisation de la progression des traitements
- Remontée des erreurs (très dépendant des programmes intégrés)
- Simplification de l'installation
- Prise en compte de standards IVOA
 - Accès au « **VOSpace** » du CDS (basé sur iRODS)
 - Prise en compte de « **Characterization** » aux entrées/sorties des briques d'un workflow
 - Mise en oeuvre d'UWS : **Universal Worker Service** pour les travaux asynchrones
 - **PLASTIC** pour l'interaction avec d'autres outils de l'OV (Aladin, TOPCAT, etc.)

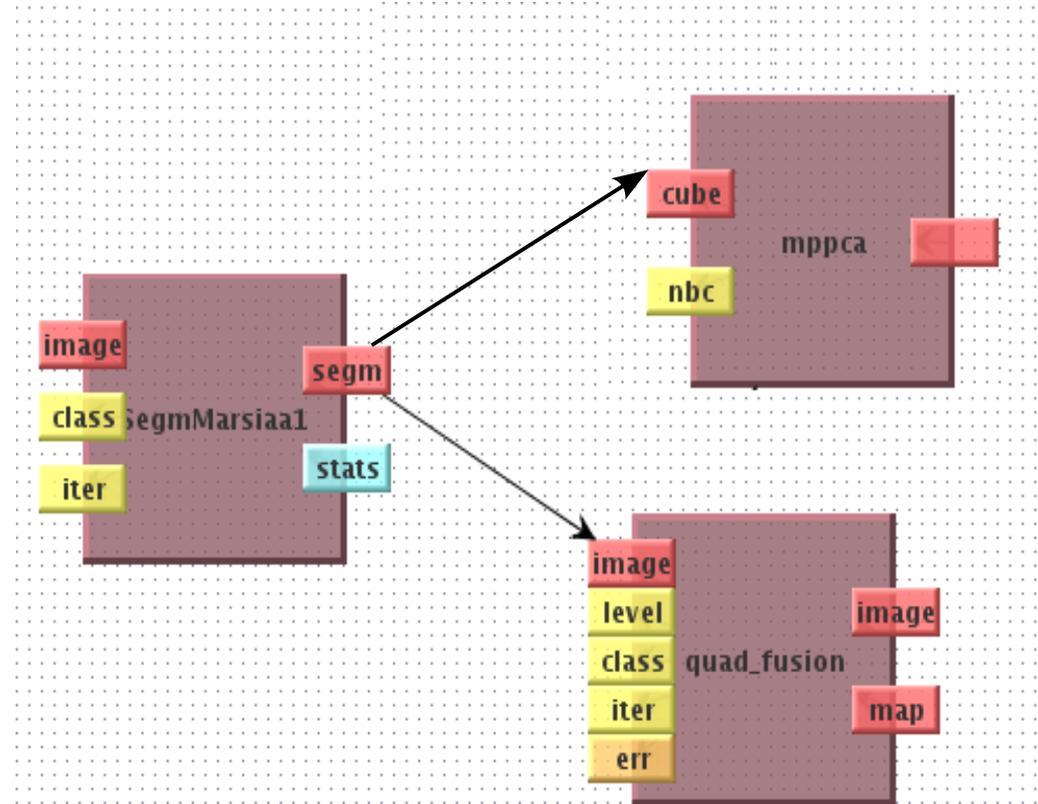
Focus : introduction d'un standard IVOA dans le traitement de workflows

- **Caractérisation** des entrées/sorties de chaque outil.
- **Permettre la vérification de la cohérence entre les images candidates et celles que l'outil attend en entrée**
 - **Démarche** : il faut traduire le fichier « **Characterization** » d'une entrée de l'outil en une liste de contraintes. Celles-ci permettront alors de vérifier s'il est possible d'appliquer le traitement sur les images candidates.
 - **Amélioration de l'efficacité du workflow => une sorte de compilation du workflow.**
- **Des retombées**
 - **Optimisation, normalisation des images en entrées**
 - **Applications à d'autres types de données**

Focus : une optimisation des workflows

■ Bénéfices

- Une vérification est effectuée au niveau du client avant la soumission au moteur de workflow
- Si la vérification est négative : minimise l'utilisation des ressources
- Gain de temps pour le scientifique
- ...



Groupe de travail Workflow d'OV France

■ Initié en 2005 et soutenu depuis par OV France

- <http://www.france-ov.org/twiki/bin/view/GROUPEStravail/Workflow>

■ Motivations

- De nombreux services sont amenés à interagir dans le cadre de l'Observatoire Virtuel (Registry, data services, Web Services, computing and Grid services, ...).
- L'utilisation et la coordination de ces services est possible au travers de workflows
 - Evolution de l'exécution "isolée" d'un service vers une exécution combinée de services (échange de données, synchronisation, choix dynamique des services, ...)
 - Questions : Localisation et vérification des services (temps d'exécution, tests, résultats, ...) ? Comment prendre en compte les outils existants dans des workflows ? Quelles contraintes au niveau des futurs développements ? Comment y intégrer des standards VO ?

Groupe de travail Workflow d'OVF (2)

■ Objectifs initiaux

- Définition de cas d'utilisation d'intérêt général dans divers domaines
- Proposer des solutions pour définir et implémenter ces workflows
- Identification de structures de workflow simples autorisant la portabilité
- Définition de briques élémentaires
- Collaboration STIC

■ Actuellement

- Implémentation des cas d'utilisations et utilisation de standards IVOA (cf. démo avec Characterization), collaboration STIC (J. Montagnat, CNRS Sophia)
- Lien avec les Grilles
- Utilisation d'AÏDA 2.0 pour prototyper une partie des travaux



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*

N.B. ce n'est qu'une partie d'un cas d'utilisation

Cas d'utilisation : Simulation

1 - PDR code

modélisation of diffuse clouds
gives theoretical absorption spectra

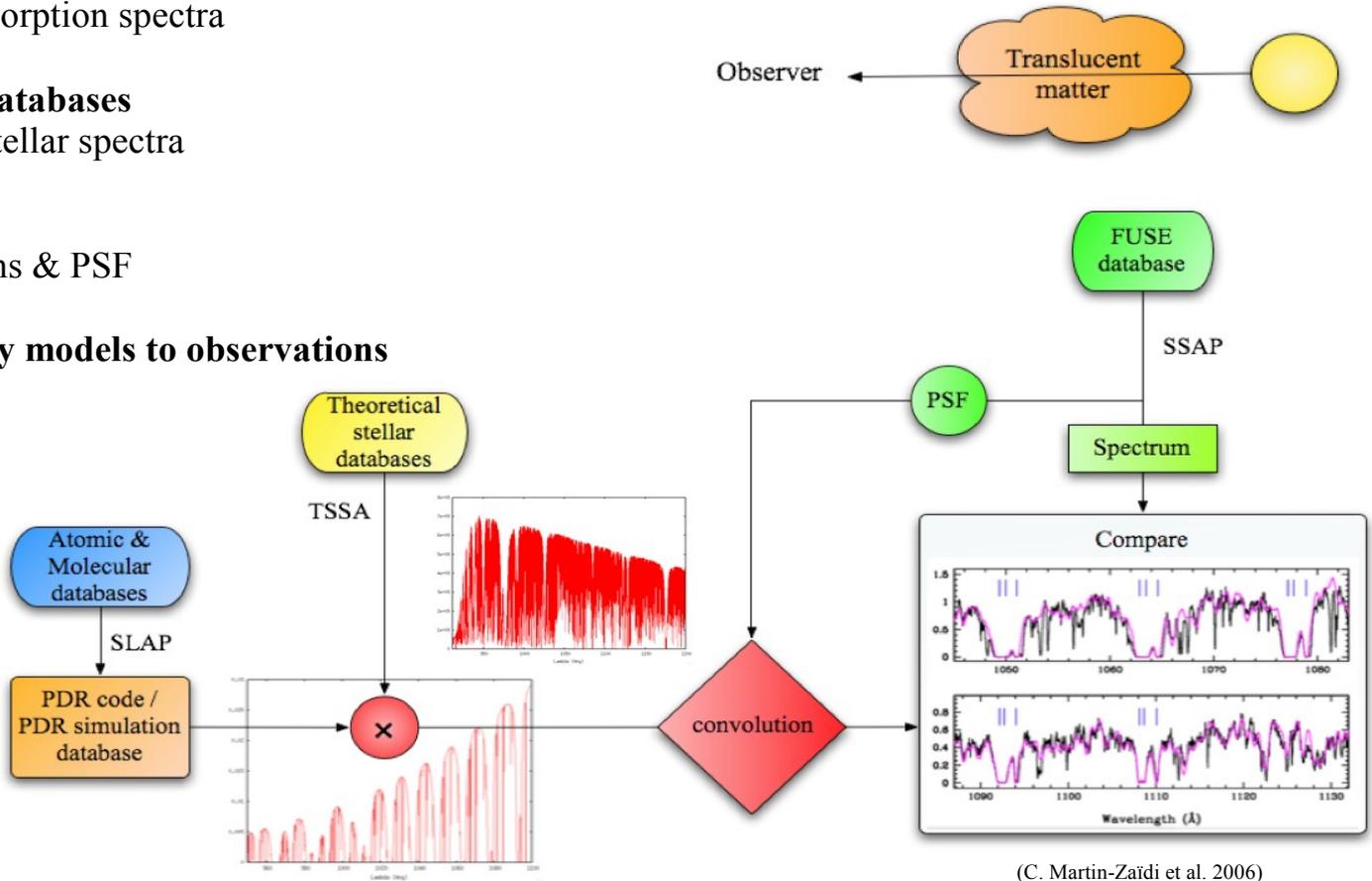
2 - Stellar spectra databases

provides synthetic stellar spectra

3 - FUSE database

provides observations & PSF

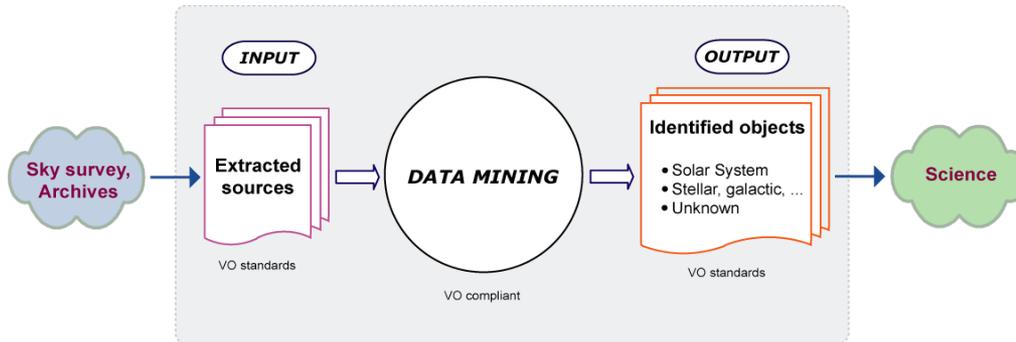
4 - Compare directly models to observations



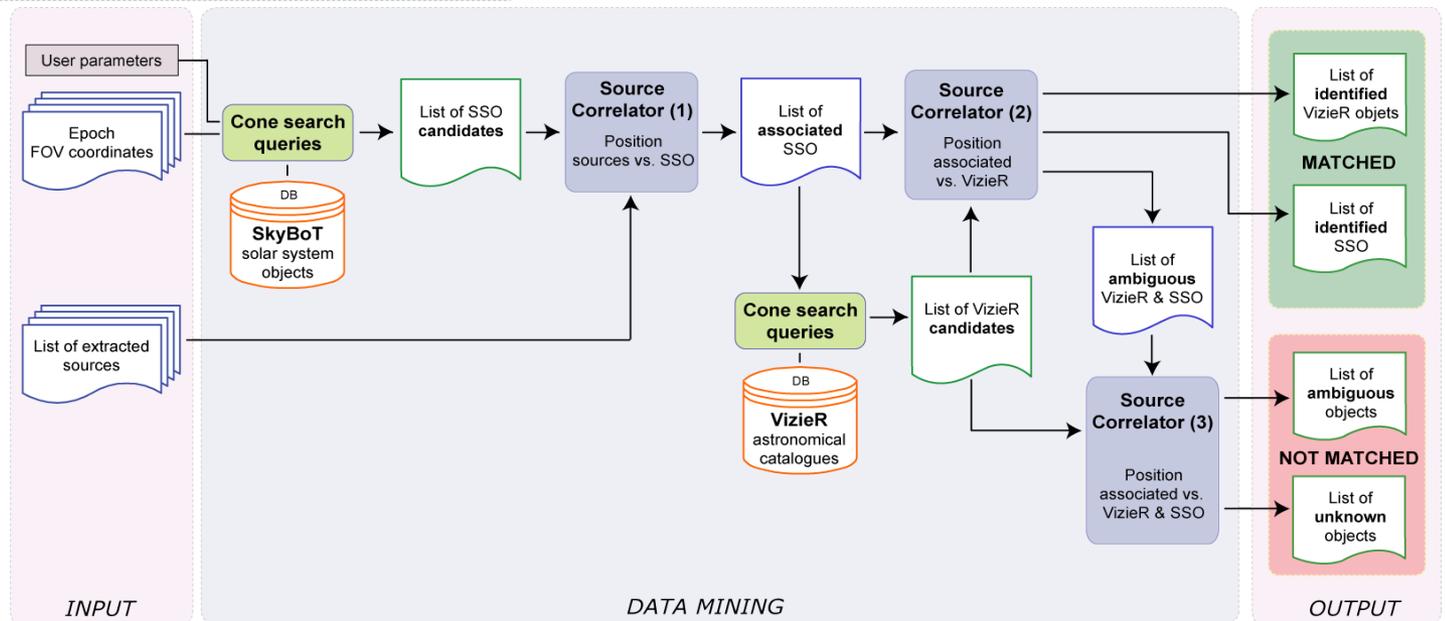
(C. Martin-Zaïdi et al. 2006)

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

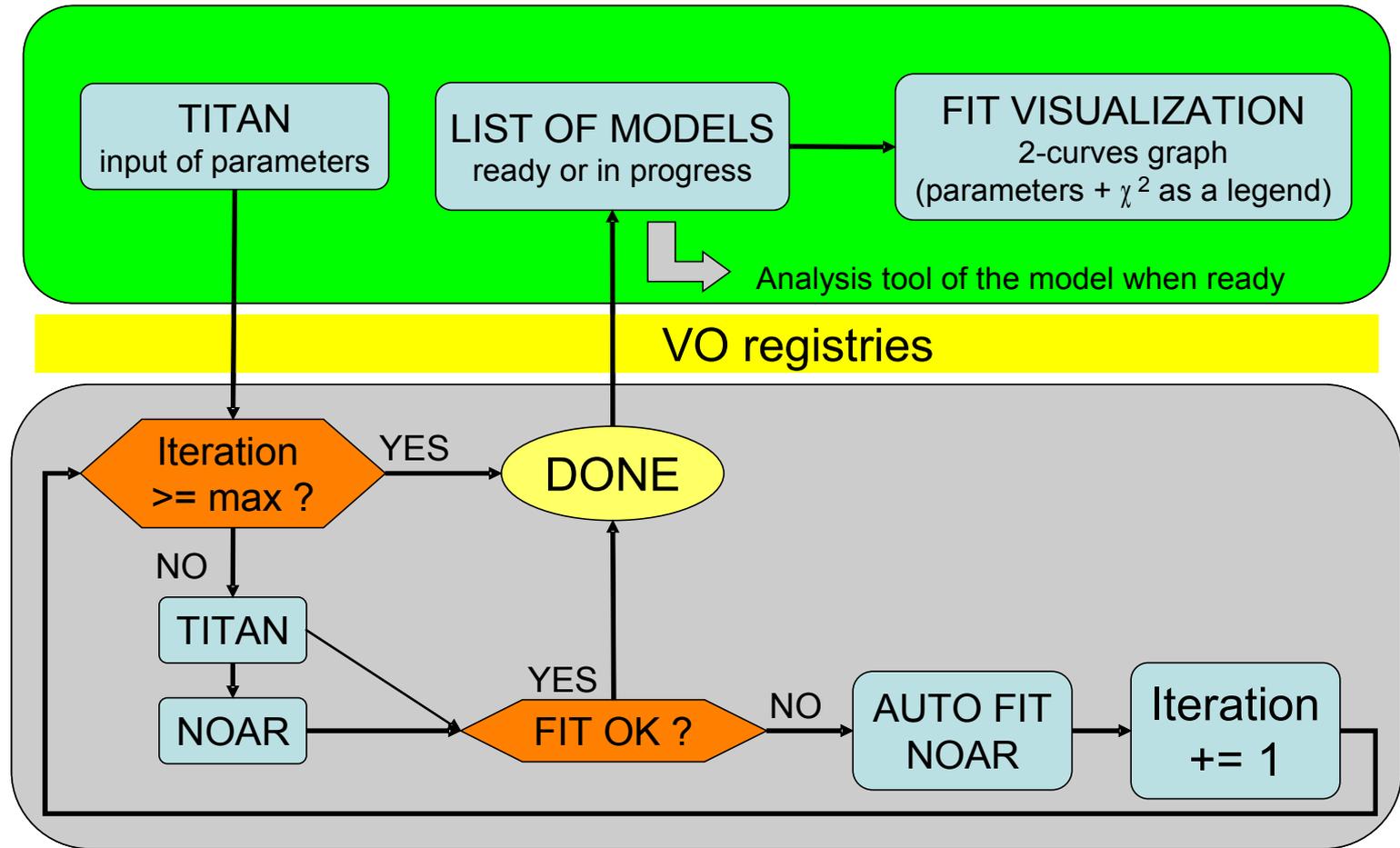
Cas d'utilisation : Data Mining



J. Berthier et al.,
IMCCE



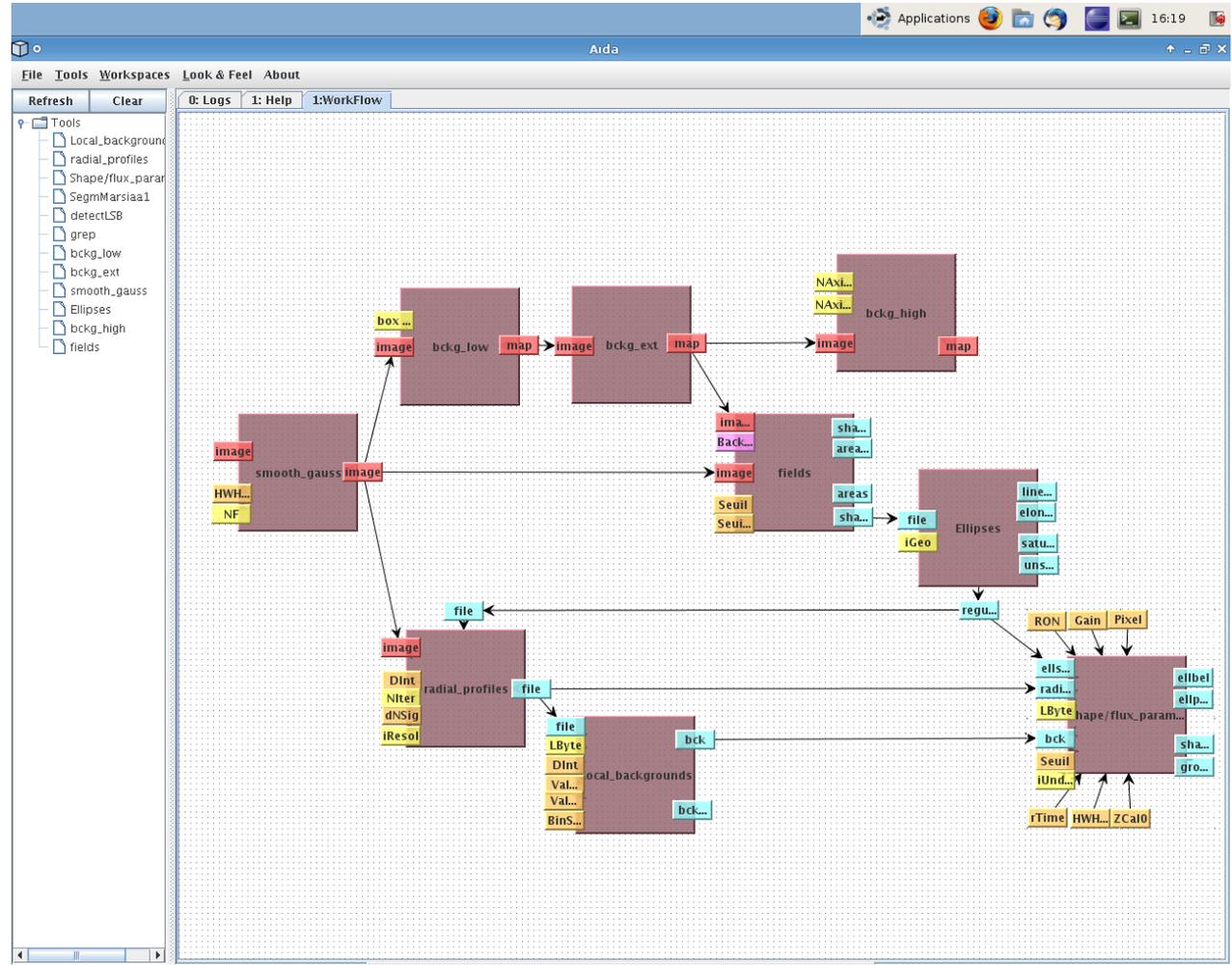
Cas d'utilisation pour des codes TITAN et NOAR



L. Chevallier
Obs. Paris
Meudon

Exemple d'implémentation du cas d'utilisation images

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



Workflow builder plugin dans Aladin

Une alternative
à la création de
scripts

C. Pestel
T. Boch
CDS

Aladin v4.0 *** BETA VERSION (based on v4.013) ***

Aladin workflow builder v. 0.3 - Using JLOW library

```

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

Remarques

- La discussion des Cas d'utilisation dans divers domaines (image, spectroscopie, data mining, simulation, ...) ainsi que leur implémentation est une approche intéressante de l'utilisation des workflows dans le cadre du VO
- L'utilisation d'outils constitue une aide réelle pour les astronomes
 - Souplesse importante par rapport à du chaînage par simple script ou des exécutions successives de différents outils « à la main »
- Bon entraînement avant de passer éventuellement à d'autres outils
- ...

Synthèse Workflow OVF

■ Historique

- 07/2003 - 12/2006, “Masses de Données en Astronomie” projet MDA de l’ACI MD : développement de l’architecture AïDA
- 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 AïDA, présentation des travaux du groupe de travail OVF, collaborations avec autres équipes VO, ...

■ Publications :

- Image Processing and Scientific Workflows in the Virtual Observatory Context, E. Slezak et al., IAU 2006
- Implementing Astronomical Image analysis pipelines using VO Standards, M. Louys et al., IAU 2006
- Workflow in Astronomy, the VO France Workflow Working Group experience, A. Schaaff et al., ADASS 2007

■ Participants and contributeurs aux réunions du groupe de travail 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-LSIIT, 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, Eric Slezak, Obs. Nice, Christian Surace – LAM, Françoise Tran Minh - LERMA Obs. Paris-Meudon, Frédéric Vachier – IMCCE, Bruno Voisin - NUI Galway, ...

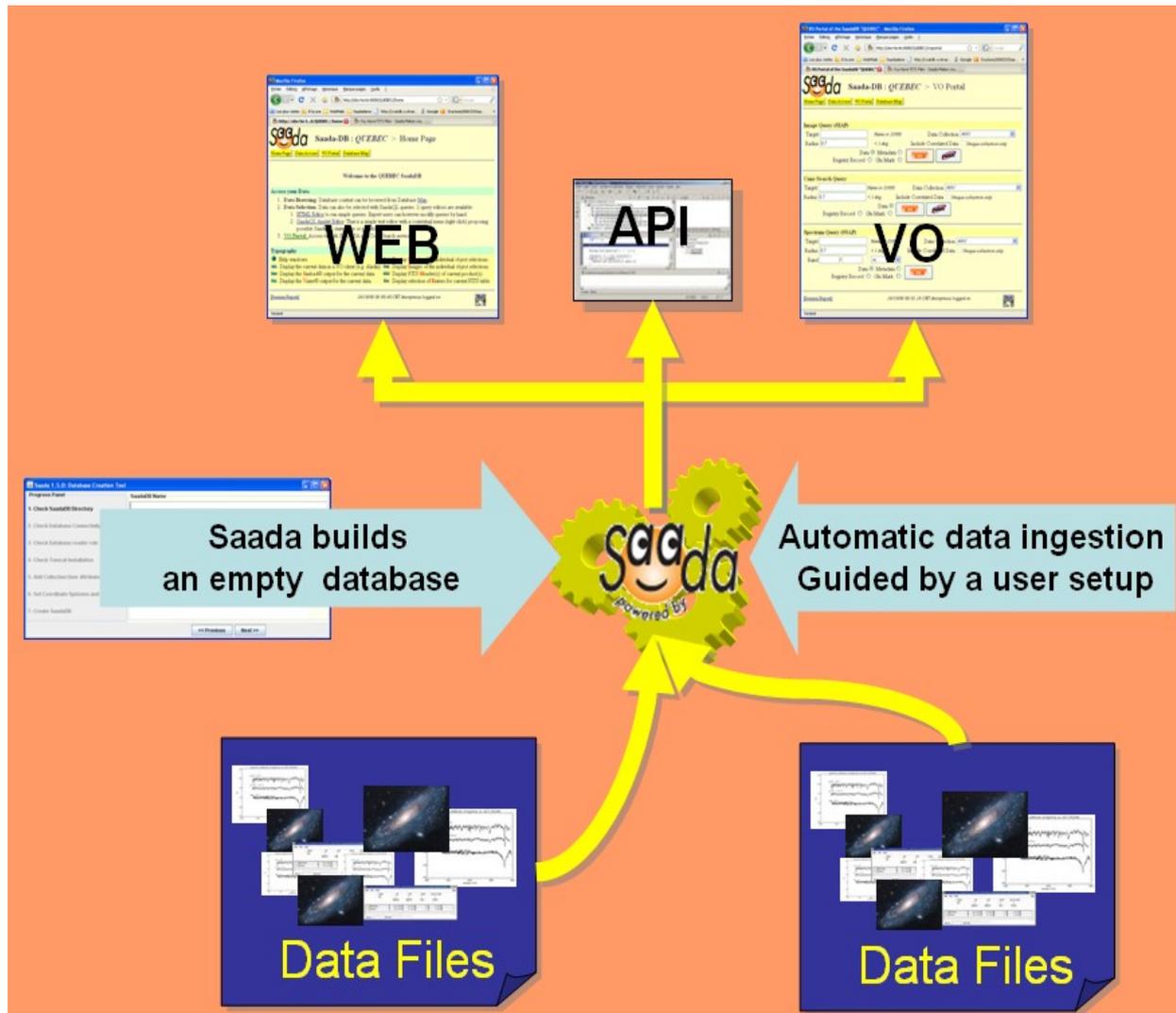
Autres outils et applications de l'OV

- Il existe maintenant de nombreux outils et applications
- Le groupe de travail IVOA Applications maintient une liste sur la page
 - <http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaApplications>
- Illustration
 - Saada (création et garnissage d'une base de données)
 - Astrogrid desktop

S a a d a

- Saada transforme un ensemble hétérogène de fichiers FITS ou de différentes catégories (images, tables, spectre....) en une base de données déployée sur le Web et fourni des « Web Services » implémentant des standards VO. La base de données (SaadaDB) est localisée chez vous et reste indépendante des serveurs extérieurs. Cette démarche ne requière aucune écriture de code.

Saada (2)



Saada (3)

■ D emos vid eo

AstroGrid Desktop

- **VODesktop est une suite d'outils interoperables : VOExplorer, Astroscope/Helioscope, File Explorer, Task Runner, et Query Builder.**

- **Principales fonctionnalités**
 - Recherche de ressources et de données dans l'OV
 - Marque-page des favoris
 - Images, spectres et catalogues;
 - Exécution de requêtes sur des bases de données
 - Sauvegarde et partage de fichiers dans un VOspace
 - Invocation d'applications distantes
 - VO Desktop comporte également un "Astro Runtime" qui permet de gérer des tâches de fond

AstroGrid Desktop (2)

■ Démo live

Les tutoriels de l'OV

- Dans le cadre des projets européens (AVO, VOTECH, Euro-VO DCA,... et maintenant Euro-VO AIDA)
- Cf.
 - <http://www.france-ov.org/twiki/bin/view/PresentationsOV/PresentationsNice>
- Liens et informations concernant ces tutoriels

Questions / réponses

