

# Description de blocs de traitement dans les workflows appliqué au traitement d'images astronomiques dans le cadre du projet AÏDA (Astronomical Image processIng Distribution Architecture) et de l'Observatoire Virtuel

**Encadrants** : Mireille Louys<sup>1</sup>, François Bonnarel<sup>2</sup>, André Schaaff<sup>2</sup>, Eric Slezak<sup>3</sup>

<sup>1</sup> Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection

<sup>2</sup> Centre de Données astronomiques de Strasbourg – Observatoire de Strasbourg

<sup>3</sup> Observatoire de la Côte d'Azur

**Durée** : 6 mois

## Contexte

Le Centre de Données astronomiques de Strasbourg (CDS) développe des services de référence à forte valeur ajoutée, très utilisés par les astronomes du monde entier. Placé sous la responsabilité de l'Institut National des sciences de l'Univers (INSU) et de l'Université Louis Pasteur (ULP), il fait partie de l'Observatoire astronomique (UMR 7550). Le CDS participe activement, en collaboration avec de nombreux centres de recherche internationaux, aux projets d'Observatoire Virtuel Astronomique (<http://cdsweb.u-strasbg.fr/>).

## Détail

L'analyse d'images en astronomie met en oeuvre des chaînes de traitements, qui en fonction de la complexité des nouvelles données disponibles se complexifient également. Les procédures de traitements peuvent être modélisées au travers de workflows, exécutés soit sur des architectures de Calcul locales (e.g. clusters), soit sur des architectures distribuées de type 'Grille'. Pour ce faire, les workflows doivent être décrits précisément, au niveau des blocs de traitements d'une part et au niveau des données qu'ils consomment et produisent d'autre part. Cela revient à définir des méta-données décrivant l'enchaînement, les blocs et les données traitées.

Le but de ce stage est de stabiliser et tester une description de workflow complète en réutilisant principalement deux standards (ou en voie de le devenir) : *VOApplication* qui donne l'ébauche d'une description de blocs de traitements, et *Characterisation*, qui fournit une description du contenu physique des données. L'étudiant définira un schéma XML à partir des éléments disponibles et développera un validateur de documents sous forme d'application Java. Il sera encadré pour l'intégration de son travail dans l'environnement de travail AÏDA.

L'un des outils d'AÏDA permet de dessiner graphiquement un enchaînement de blocs, de construire une description XML de ces blocs et de soumettre le tout à un moteur d'exécution de workflow. Afin d'optimiser au mieux l'utilisation de cet outil, nous cherchons à introduire un contrôle dans le workflow avant que les blocs ne soient exécutés.

Cette démarche implique une description des données et la validation de quelques propriétés portant sur les données soumises. le but étant de construire une représentation souple s'appuyant sur le standard de description des metadonnées astronomiques en cours de développement, le modèle *Characterisation*, et de développer une représentation des blocs cohérente avec ces données d'une part et avec un autre modèle , celui des

blocs chaînés dans les workflows.

Le sujet demande donc de tenir compte d'un existant: les modèles de données et traitements , d'affiner les schémas XML correspondants et de prévoir une évolution souple des deux représentations.

En effet, tout nouvel instrument peut nécessiter une description plus complète des données, et par ailleurs tout producteur de logiciel de la communauté astronomique peut enrichir la description des blocs de traitements.

Le stagiaire sera formé aux aspects standards. Il devra développer un 'compilateur' de workflow plus élaboré qu'un simple parseur XML. Dans certains cas , il faut vérifier la cohérence des valeurs des métadonnées , c'est à dire traduire en règles ces conditions puis coder une application qui fait le contrôle.

Parallèlement à ce travail l'étudiant sera également amené à effectuer des évolutions de certains éléments (outil de représentation graphique des workflows, moteur d'exécution de workflows, ...) de l'architecture AIDA. Celle-ci a été développée dans le cadre du projet Masses de Données en Astronomie (ACI MD 2003). Elle se compose d'une partie cliente pour la description graphique de workflows, d'un moteur d'exécution de workflows ainsi qu'une d'une partie serveur hébergeant les applications pouvant constituer un workflow.

**Connaissances souhaitées** : pratique d'UML, programmation Java (notions de C-C++ et Perl). Bonne pratique de l'anglais. Notions d'XML.

Le stagiaire sera éventuellement amené à se déplacer pour des missions de collaboration de quelques jours en France ou à l'étranger.

#### **Références disponibles :**

les modèles d'applications disponibles dans l'Observatoire Virtuel:  
<http://www.ivoa.net/twiki/bin/view/IVOA/RegDMApplications>

un modèle pour les données 'digérées' par le workflow:  
<http://www.ivoa.net/Documents/latest/CharacterisationDM.html>