

Masses de Données en Astronomie
Réunion finale
Strasbourg, 10-11 avril 2006

**Architecture et moyens de traitements :
travaux et perspectives**

Groupes Calcul et Image

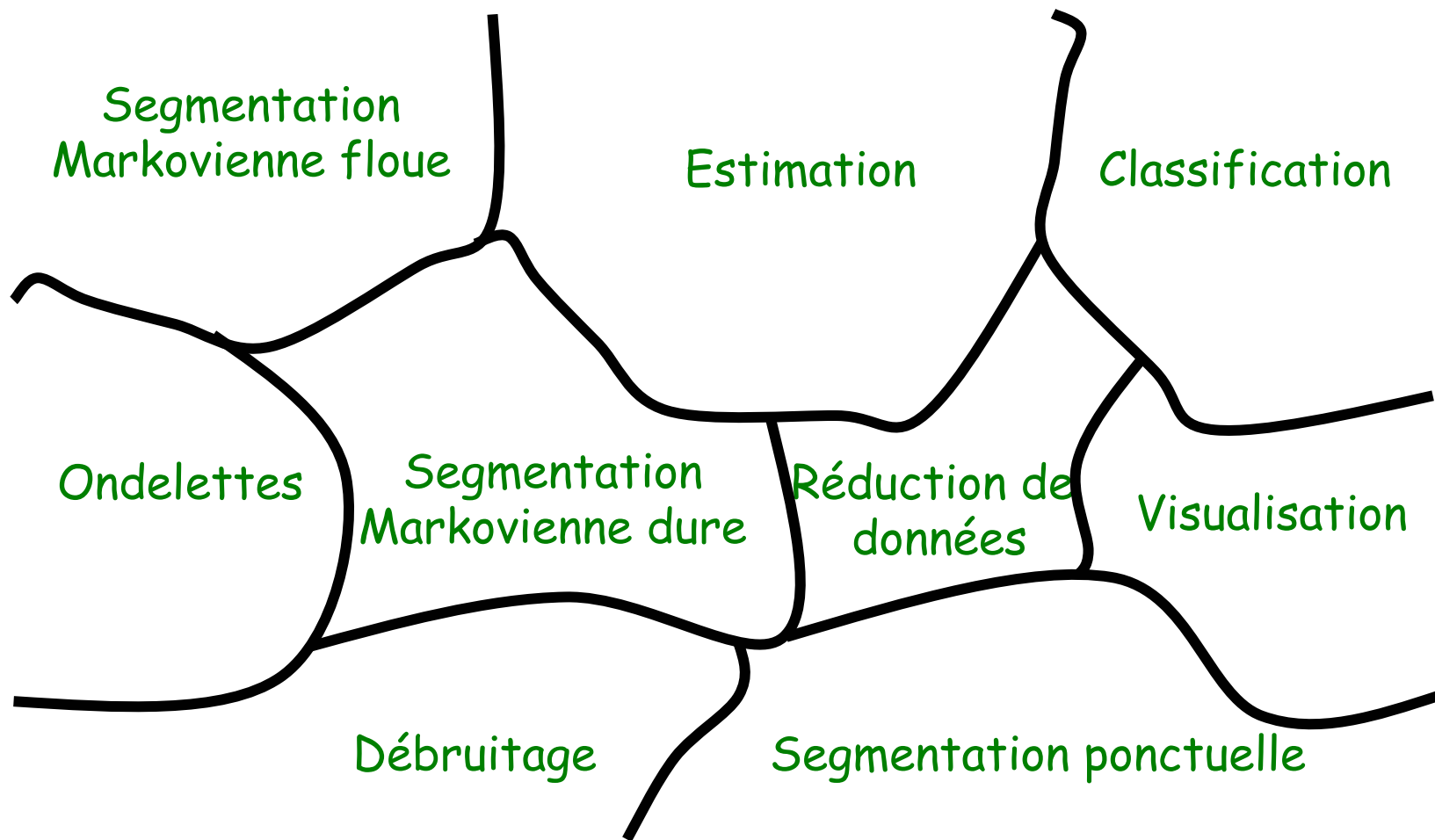


Réunion finale MDA
Strasbourg – 10 et 11 avril 2006

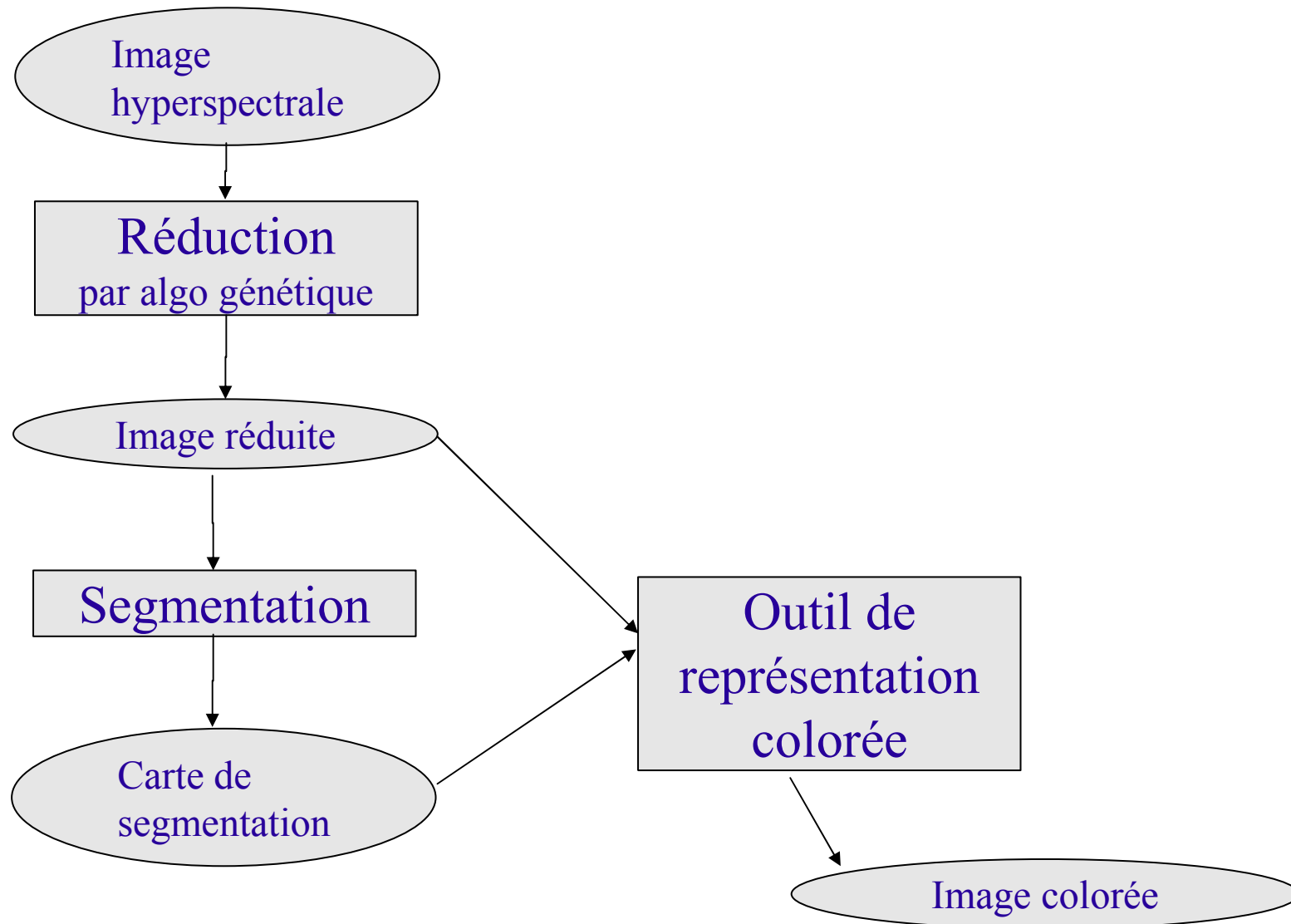
Rappel (1) : la problématique

- **Utilisation du traitement d'image en Astronomie**
- **Outils existants**
 - **Différentes techniques de traitement implémentées :**
 - essentiellement en Matlab (contraintes : exécution locale et licence propriétaire)
 - dans un contexte de recherche
 - **Utilisation au cas par cas :**
 - par un public restreint (nouveaux algorithmes de traitement et cas d'étude astronomique)
 - en optimisant principalement les algorithmes mais les temps d'exécution restent élevés (particulièrement pénalisant en phase de test)

Rappel (2) : des programmes de traitement d'images



Un exemple d'utilisation



Objectifs initiaux

- **Mise à disposition uniformisée des outils**
 - Interface utilisateur
 - Ouverture vers d'autres outils (Aladin, ...) par des méthodes d'accès génériques
 - Intégration facilitée de nouveaux outils

- **Amélioration à divers niveaux**
 - Performances
 - Enchaînement de tâches
 - Diminution de la complexité et de l'hétérogénéité

- **Utilisation élargie à un plus vaste public**

Mise à disposition uniformisée d'outils

■ Au niveau de l'utilisateur

- Masquer les langages utilisés
- Proposer des traitements sous forme de « boîte noire » avec des entrées et des sorties
 - Référentiel des outils disponibles, sélection d'un outil, génération d'une interface utilisateur spécifique, ...

■ Au niveau de l'intégration des outils

- Apporter à des outils écrits dans divers langages (MATLAB, C, ..) une interface standard (cgi-bin, Web Service)
 - Utilisation d'un fichier « descripteur »
 - Intégration dans un serveur Web
 - ▶ induit l'ouverture vers d'autres outils (exemple : Aladin)
 - ▶ permet d'envisager la mise en œuvre d'enchaînements de tâches (Workflow)

Améliorations...

■ Première piste

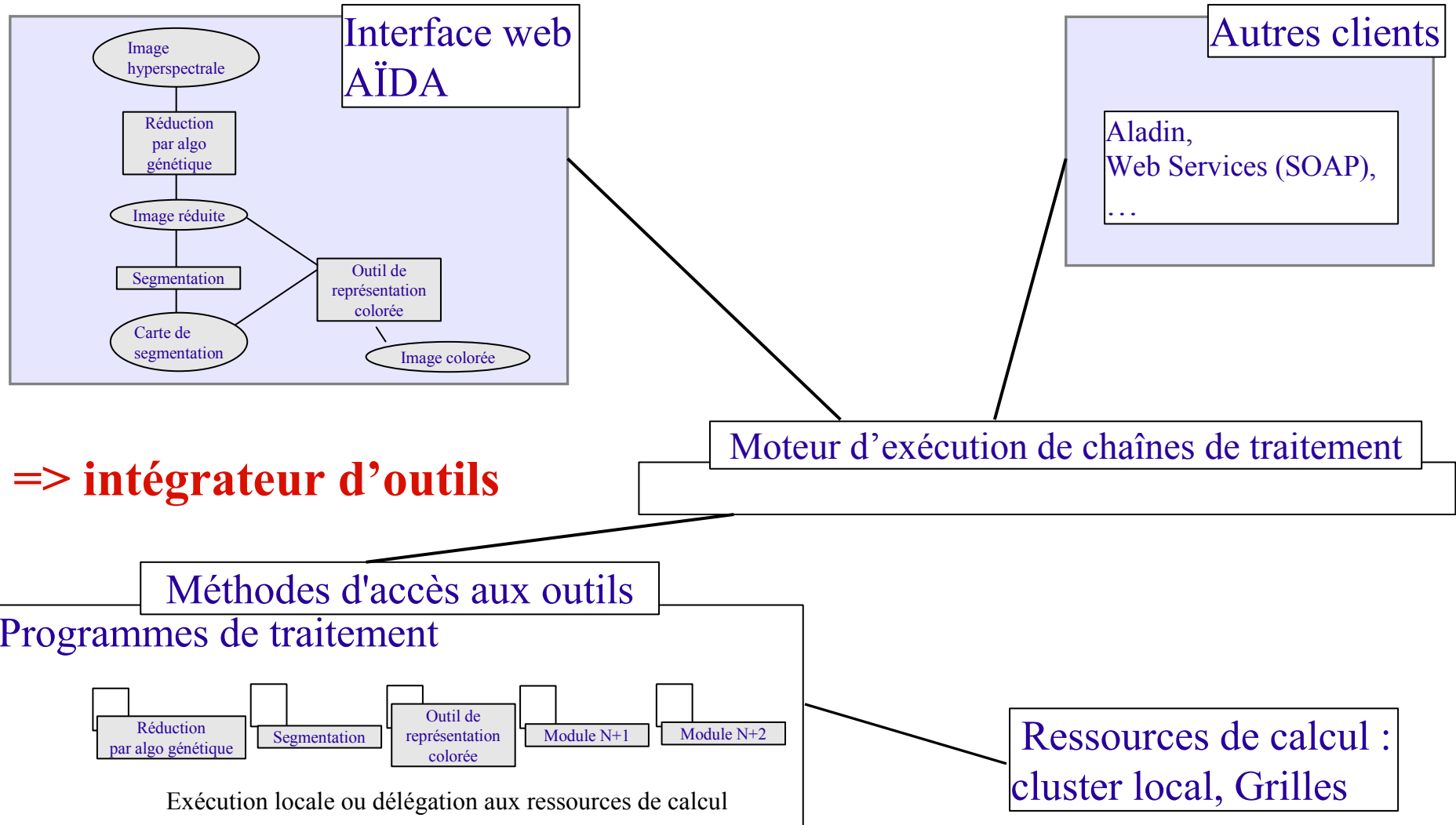
■ Modification des sources C générés à partir d'utilitaires MATLAB

- Parallélisation de certaines parties des sources, ...
- Abandonné pour diverses raisons : travail à reproduire après modification des programmes écrits en MATLAB, sources masqués dans les dernières versions de MATLAB, ...

■ Solution actuelle

- Prise en compte en amont (dans MATLAB) du découpage (dans la mesure du possible) des traitements
- Intégration de la notion de « Workflow » dès le développement initial
 - Par rapport aux traitements
 - Par rapport aux données (découpage d'une image en n parties)
- Exploration des techniques de Grilles

AIDA : Astronomical Image processing Distribution Architecture



=> intégrateur d'outils

Ressources de calcul (1) : le cluster local

■ Au début du projet MDA

- Utilisation de ressources existantes pour les premières expérimentations

- Pas d'achat afin d'éviter une obsolescence du matériel en fin de projet

■ En milieu de projet

- Achat d'un premier lot de 4 biprocesseurs Opteron 64 bits

- Tests du passage à une architecture 64 bits, ...

■ Actuellement

- Nouvel achat de 6 biprocesseurs Opteron 64 bits pour finaliser le cluster

- Travail d'intégration en cours, en partie dans le cadre d'un stage MDA de 6 mois (Cyril Pestel, Master2)

Ressources de calcul (2) : exploration des techniques des Grilles

10

■ Collaboration avec l'ESSI Sophia

- Premiers tests d'accès aux grilles Grid5000 et EGEE
- Accès aux grilles au travers de l'outil « MOTEUR » développé par l'ESSI

■ Utilisation des grilles à bon escient

- Intéressante lorsque le traitement est lourd et qu'il peut être découpé
 - ▶ Cas asynchrone
- Inutilisable dans le cas de traitements interactifs (problème de maîtrise de l'ordonnancement, ...)

■ Autre : Kerrighed en cours de test mais ne sera pas utilisable en production avant environ 18 mois

Ressources de calcul (3) : vers une solution mixte

11

- **Un cluster local est indispensable**
 - Pour les traitements pas trop coûteux en ressources de calcul
 - Pour les traitements coûteux mais non parallélisables
 - Pour garantir de bonnes performances lors de l'utilisation interactive d'outils

- **Le recours aux grilles peut se justifier dans certains cas :**
 - **Traitements parallélisables et très coûteux en ressources de calcul**
 - Risque de dégradation des performances du cluster au détriment des autres tâches
 - Adapté aux grilles qui permettent un changement d'échelle au niveau du nombre de nœuds utilisables

Résultats

- Amélioration notable de l'ergonomie et des performances
- Aide à l'élaboration de « workflow »
- Proposition d'outils pointus à des astronomes non spécialistes en traitement d'image
- ...

Perspectives

- **Architecture plus ouverte :**
 - **Possibilité d'ajout de programmes de calcul délocalisés**
 - Vers une notion de nœud AIDA
 - Création de « workflow » comportant des traitements répartis sur plusieurs nœuds
 - **Travail en cours :**
 - Installation d'AIDA à l'Observatoire de Nice
 - Etude de nouveaux cas d'utilisation de traitements délocalisés
- **Amélioration de la robustesse dans la perspective d'un outil utilisable en production post MDA**