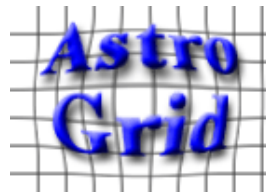


Présentation des Workflows Astrogrid



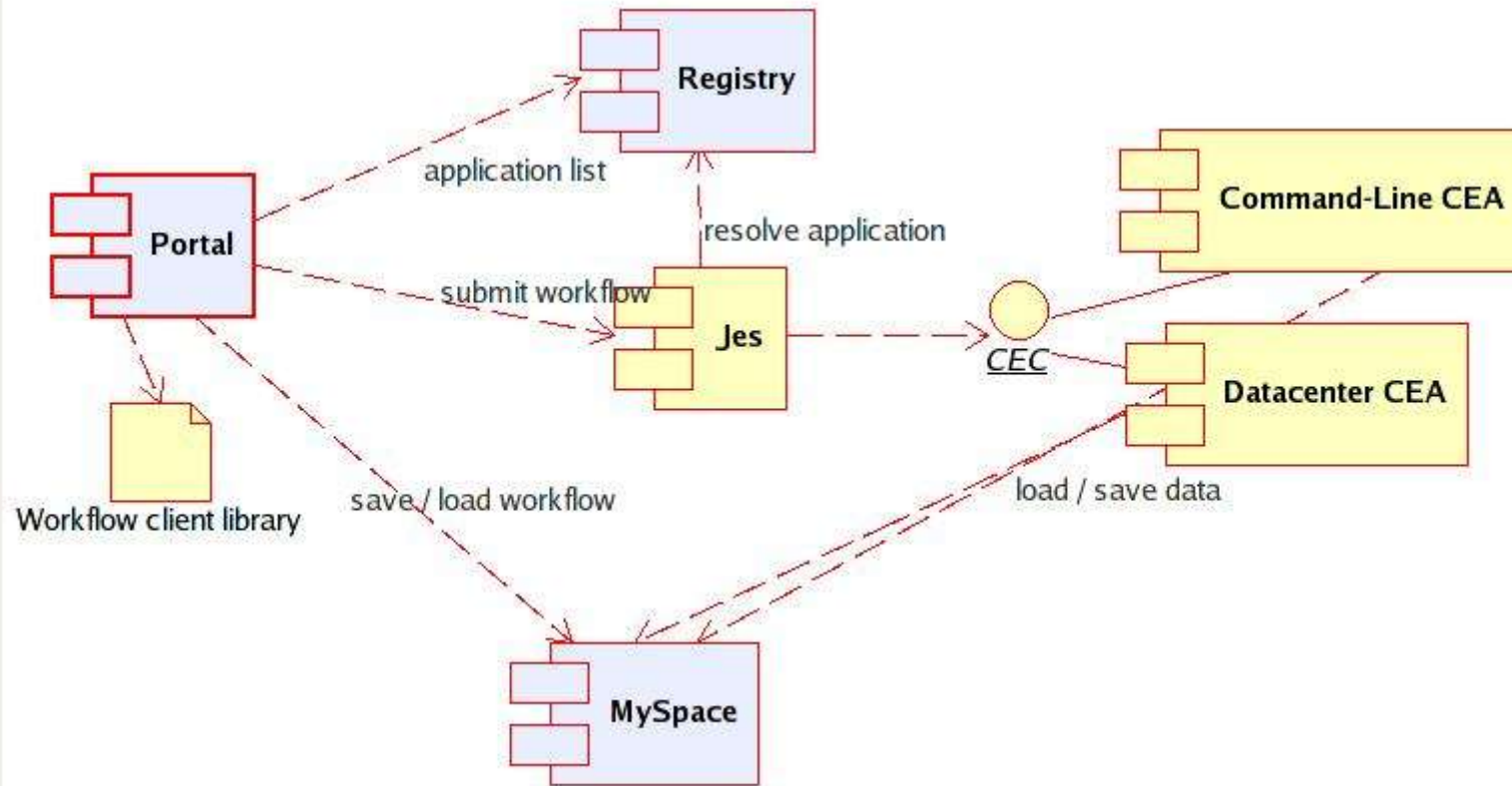
Thomas Boch [CDS]



Groupe de travail Workflow OV-France –
Réunion du 10/11/2005



Architecture



Glossaire

- Portal : Interface Web permettant l'accès aux services Astrogrid
- Workbench : application cliente fournissant
 - les mêmes fonctionnalités que le portail Web
 - + accès programmatique aux services
- CEA : **C**ommon **E**xecution **A**rchitecture
 - Standard de description des applications (paramètres d'entrée, paramètres de sortie, méthodes d'accès, ...)
 - Permet une description et un accès uniforme aux différentes applications composant un workflow



Glossaire (2)

- JES : Job Execution System
 - Moteur d'exécution du workflow
 - contrôle l'exécution du workflow dans sa globalité
- MySpace : espace de stockage distribué
 - Stockage des résultats intermédiaires et finaux
 - Fichiers référencés par leur IVORN (IVO Resource Name), eg
<ivo://uk.ac.le.star/ThomasBoch#OV-France/2mass.xml>
- Registry : annuaire listant (entre autres) les applications CEA disponibles
- CEC : Common Execution Controller
 - Contrôle l'exécution d'une tâche précise
 - Avertit JES de la fin d'une tâche



Mécanisme de persistance

- Structure d'un workflow décrite par un XML Schema
- Chaque description de workflow est sauvegardée dans un fichier XML (dans MySpace)
- Exemple :

```
- <Activity xsi:type="sequence">
  <Activity var="twomassadql" value="" xsi:type="set"/>
  - <Activity xsi:type="script">
    <description>open twomass query template</description>
    - <body>
      ev = astrogrid.ioHelper.getExternalValue("http://www.merlin.ac.uk/astrogrid/cctwomass.q") twomassadql = astrogrid.ioHelper.getContents(ev)
    </body>
  </Activity>
  - <Activity xsi:type="script">
    - <description>
      Insert sky area, 2MASS bands to select on, and colour limit, into query
    </description>
    - <body>
      import groovy.text.Template import groovy.text.SimpleTemplateEngine binding = ["ramin":nra[0], "ramax":nra[1], "raaltmin":nraalt[0],
      "raaltmax":nraalt[1], "decmin":ndec[0], "decmax":ndec[1], "twomassb1":twomass[0], "twomassb2":twomass[1], "twomasslim":twomass[2]] engine =
      new SimpleTemplateEngine() template = engine.createTemplate(twomassadql).make(binding) twomassadql = template.toString()
    </body>
  </Activity>
- <Activity name="IR query to 2MASS" joinCondition="any" xsi:type="step">
  <description>IR query with colour cut</description>
  - <tool name="roe.ac.uk/DSA_2MASS/ceaApplication" interface="adql">
    - <input>
      - <parameter name="Query" indirect="false">
        <ns1:value>$(twomassadql)</ns1:value>
      </parameter>
      - <parameter name="Format" indirect="false">
        <ns2:value>VOTABLE</ns2:value>
    </input>
  </tool>
</Activity>
```



Éléments d'un workflow

- Comparable à un langage de programmation
- **Step** : Élément appelant une application CEA
 - Requêtes à des centres de données (ConeSearch, SIA, ADQL, ...)
 - Tâches de traitement (SExtractor, HyperZ, CrossMatcher, ...)
- **Set** : Définit une nouvelle variable ou met à jour la valeur d'une variable existante
- Structures de contrôle
- Scripts Groovy



Structures de contrôle

- **Flow** : ensemble d'opérations exécutées en parallèle
- **Sequence** : ensemble d'opérations exécutées séquentiellement
- **If** : exécution conditionnelle
- **ForLoop** : boucle itérant sur un ensemble
- **ParallelLoop** : boucle dont chaque itération est indépendante des autres --> parallélisme
- **WhileLoop** : boucle avec test à chaque itération



Groovy

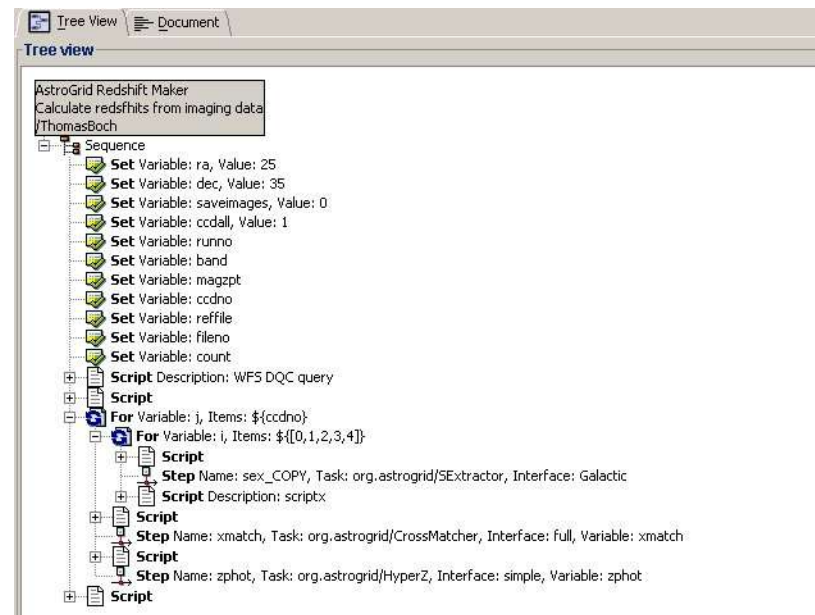
- Langage de script pour Java
 - Super-ensemble de Java
 - Sucre syntaxique pour faciliter certaines opérations
- Utilisé comme “glue” dans les workflows :
 - Post-traitement sur le résultat d'une appli. CEA (eg: extraction de la valeur d'un champ)
 - Adaptation des paramètres d'entrée aux paramètres attendues par l'application
 - Accéder aux données MySpace

```
col = ( 0 ... table.columnCount ).find{ table.getColumnInfo  
( it ).getUCD() == 'VOX:Image_AccessReference' }  
urls = table.columnIterator(col).collect{it}
```



Création/visualisation d'un workflow

- Via le portail Web
 - Visualisation & création
- Via le Workbench
 - Visualisation uniquement pour le moment
 - Création de workflows : **coming soon**



Accès programmatique aux workflows

- ACR : Astronomy Client Runtime
 - API offrant une “façade” permettant un accès “programmatique” aux services Astrogrid
 - Multi-plateformes & multi-langages
- Au niveau “workflow”, ACR permet de
 - Soumettre un workflow
 - Avoir des informations sur un job en cours d'exécution
 - Obtenir des composants graphiques :
 - ◆ Workflow viewer
 - ◆ Job monitor



Liens

- AstroGrid Release 1.1 :
<http://software.astrogrid.org/>
- Documentation workflow Astrogrid :
<http://software.astrogrid.org/components/jes/userguide-architecture.html>
- JES scripting FAQ :
<http://wiki.astrogrid.org/bin/view/Astrogrid/JesScriptingFAQ>
- Différence Portal/Workbench :
<http://software.astrogrid.org/userdocs/index.html>
- ACR :
<http://www.astrogrid.org/maven/docs/HEAD/desktop/>

