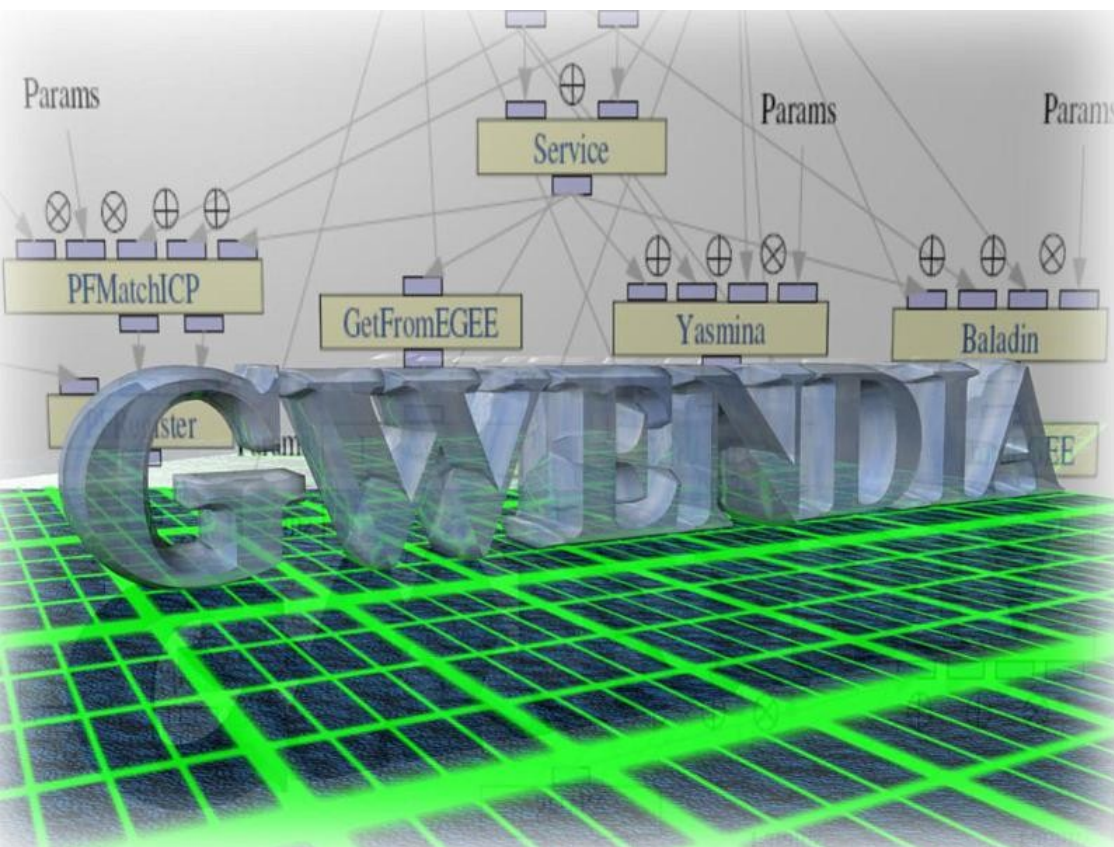




Grid Workflow Efficient Enactment for Data Intensive Applications

GWENDIA workflows for cardiac image analysis



*E. Caron, C. Casta, P. Clarysse,
T. Glatard, B. Isnard,
K. Maheshwari, J. Montagnat*
CNRS

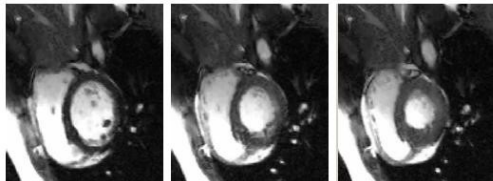


*EGEE User Forum
Uppsala, Sweden
April 13, 2010*

Financé par
ANR

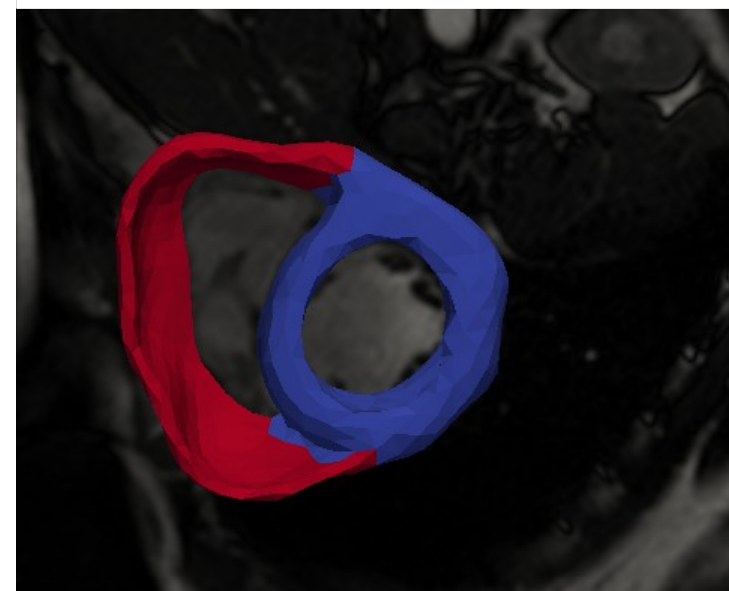
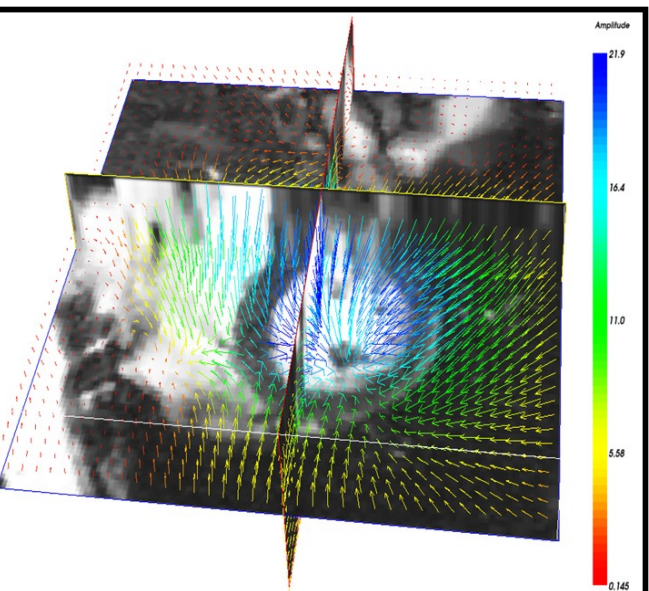
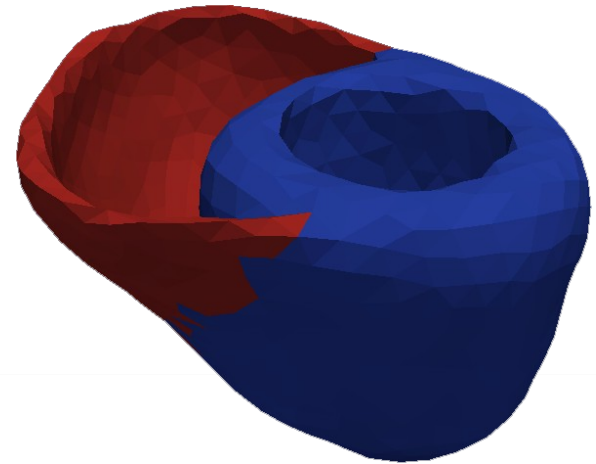
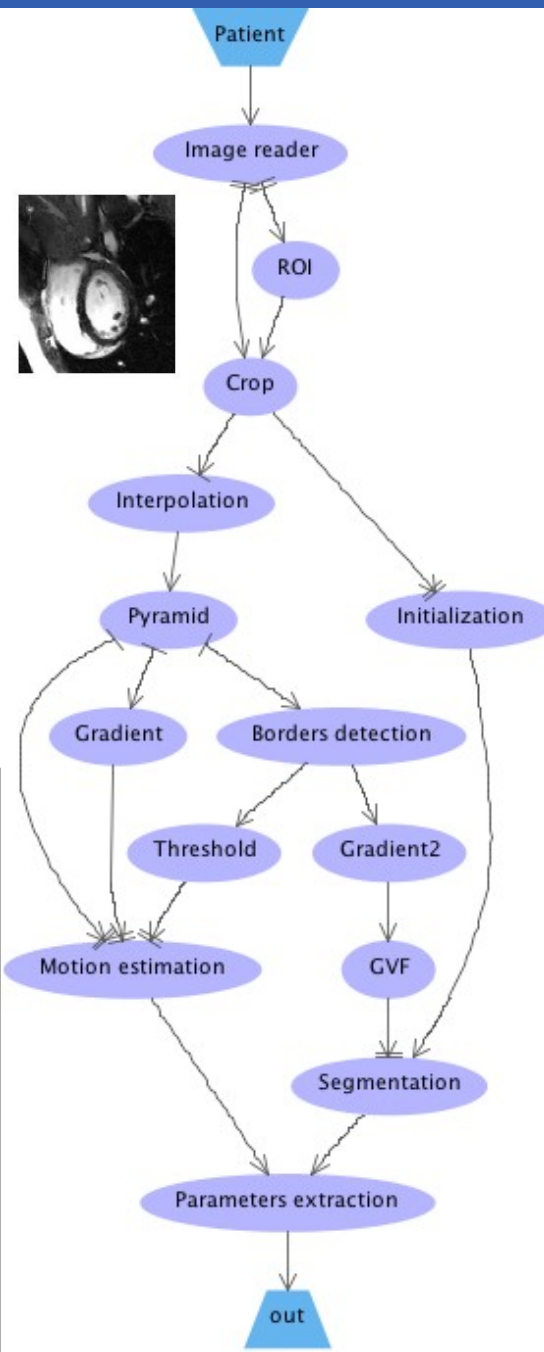


Cardiac MR sequences analysis workflow(s)



Grid World

Convs





- **Complex data flow representation**

- Data parallel processing
- Workflow variants (segmentation branch, motion estimation branch, parameter sweep for algorithms tuning)

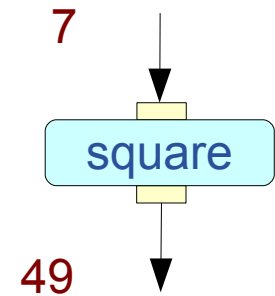
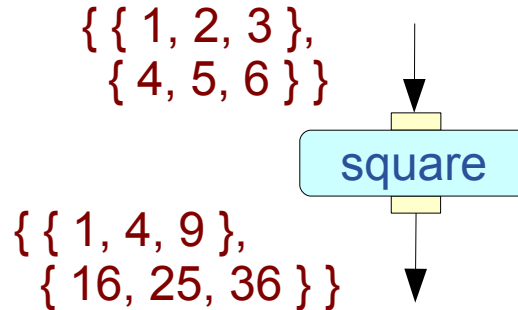
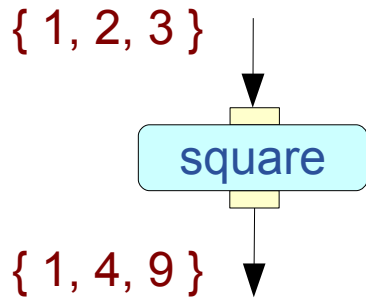
➡ GWENDIA language, use of array programming principles

- **Different execution environments**

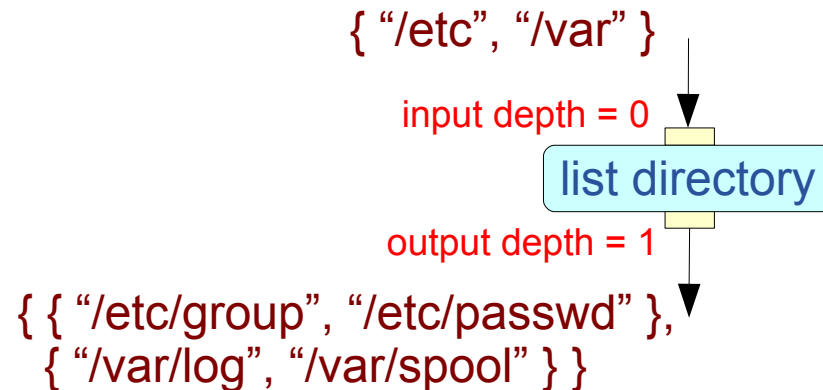
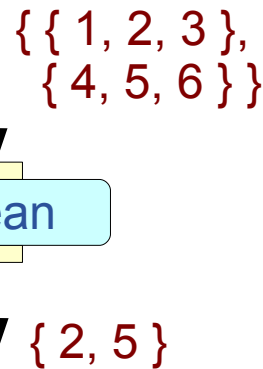
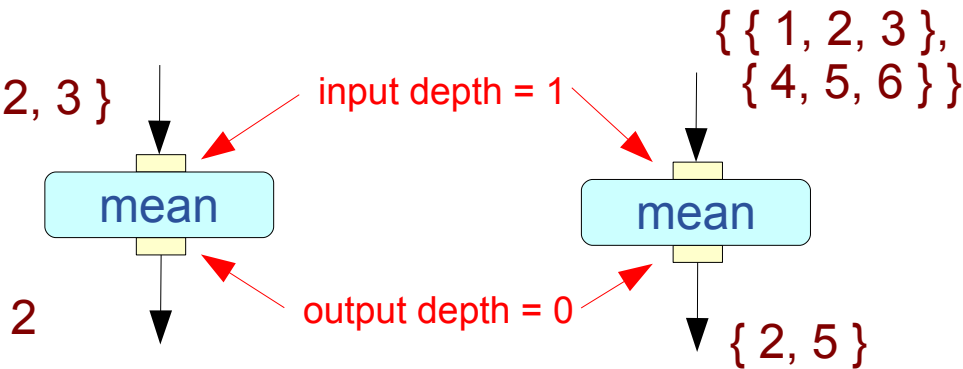
- EGEE production infrastructure
- Grid'5000 French research infrastructure

➡ Workflow interoperability techniques

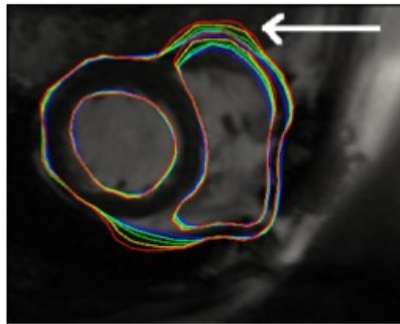
- **Data structures are (nested) arrays**
 - A scalar is a particular case of 0-dimension array



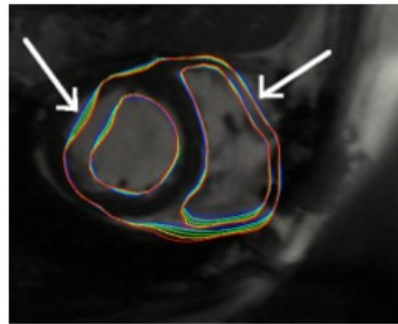
- **Each port is associated a port depth**
 - Input arrays nesting levels are combined with activities ports depth



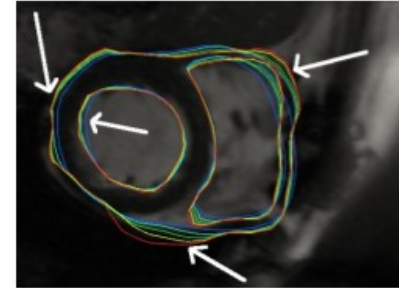
- Optimize algorithms parameters



init 0



init 1

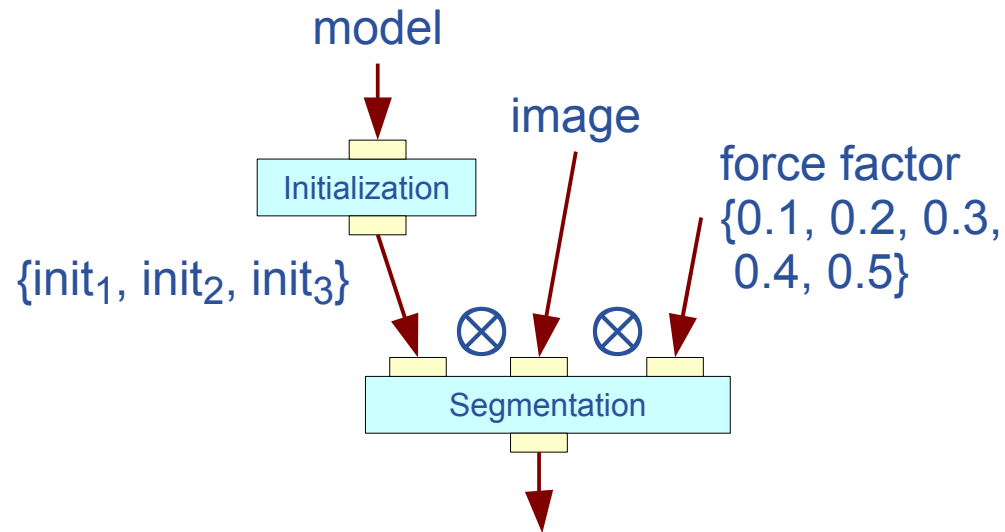
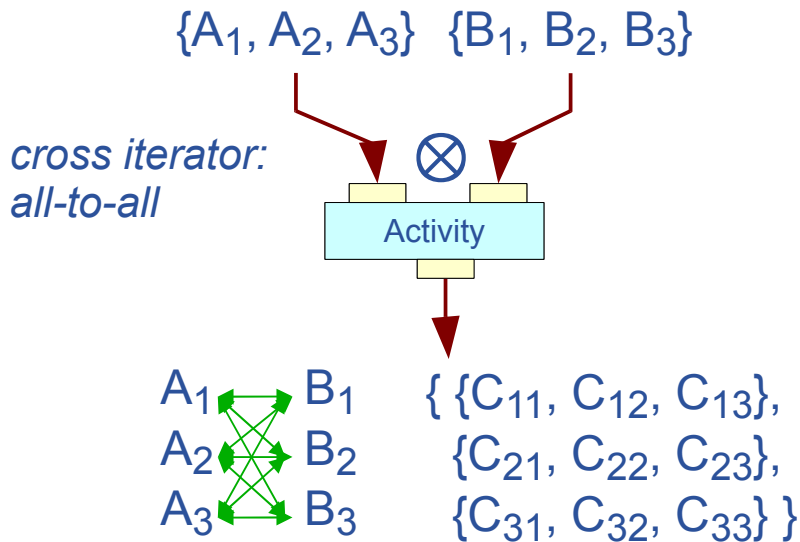


init 2

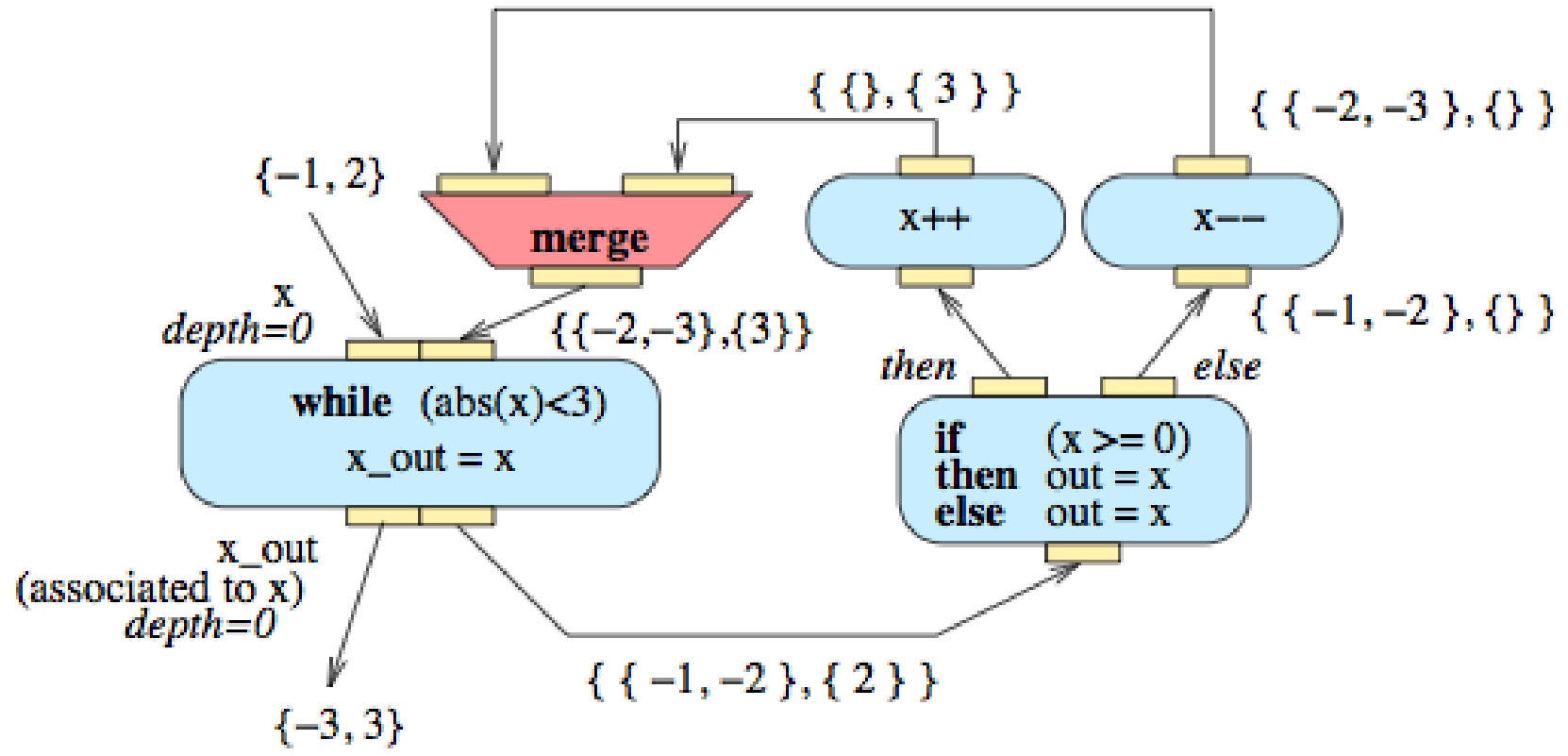
- █ force factor = 0.1
- █ force factor = 0.2
- █ force factor = 0.3
- █ force factor = 0.4
- █ force factor = 0.5

[Maheshwari et al, HG'09]

- Use of iteration strategies

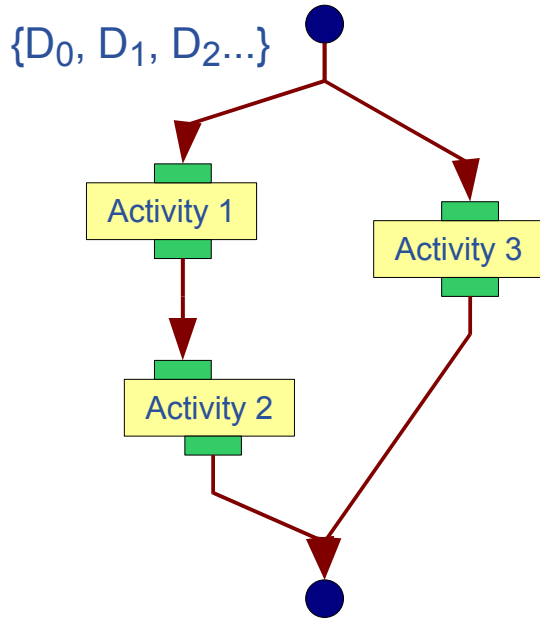


- **Iteration strategies**
 - dot, cross, flat-cross, match
- **Conditional and loops**

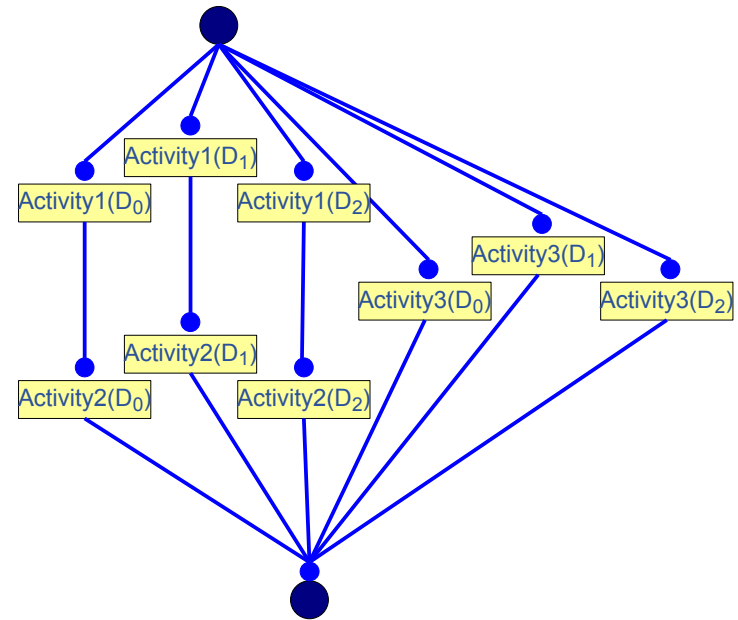


[Montagnat et al, WORKS'09]

- **Activities seen as services**



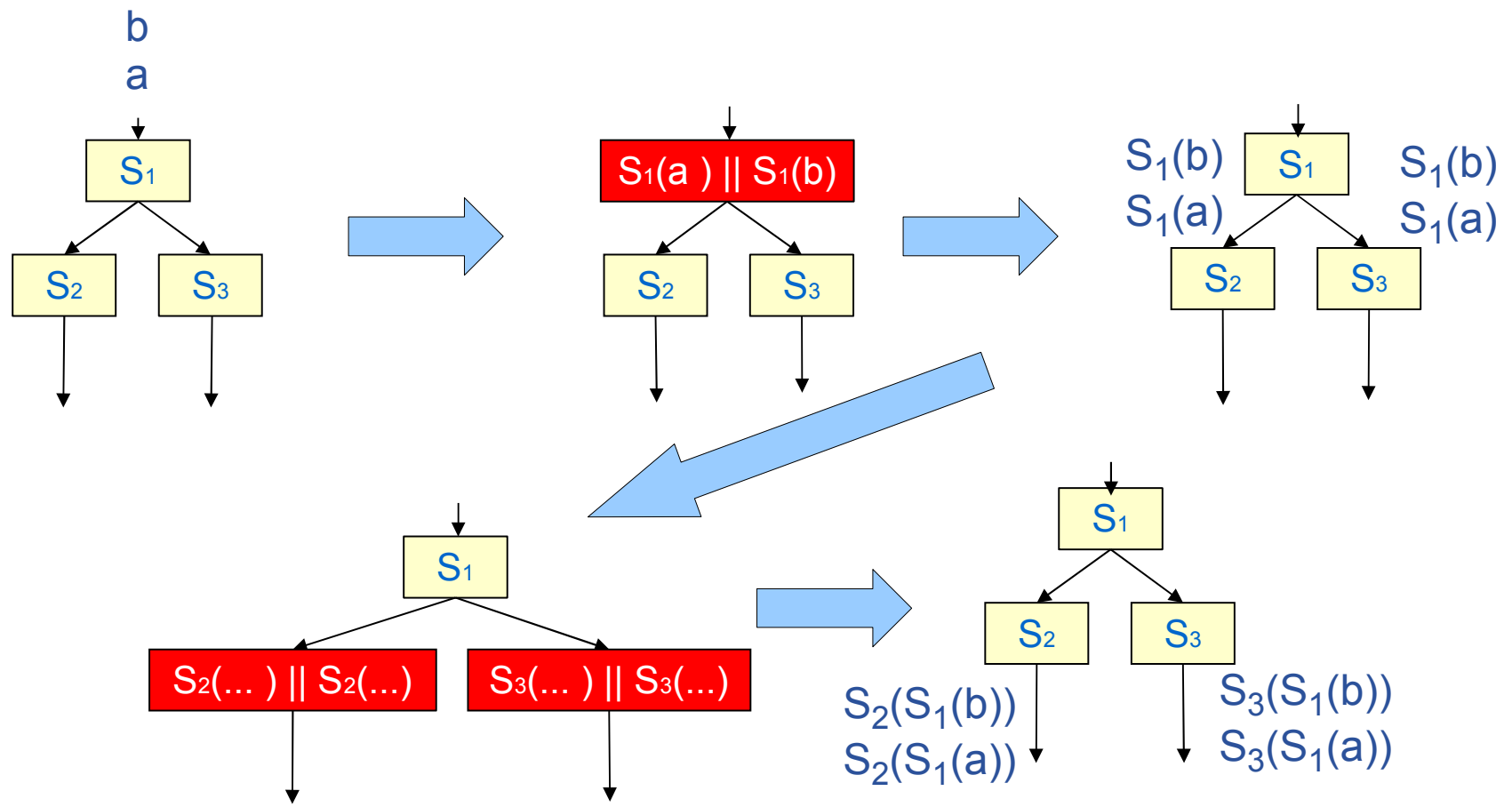
- **Activities seen as tasks**



- **Different enactment strategies**

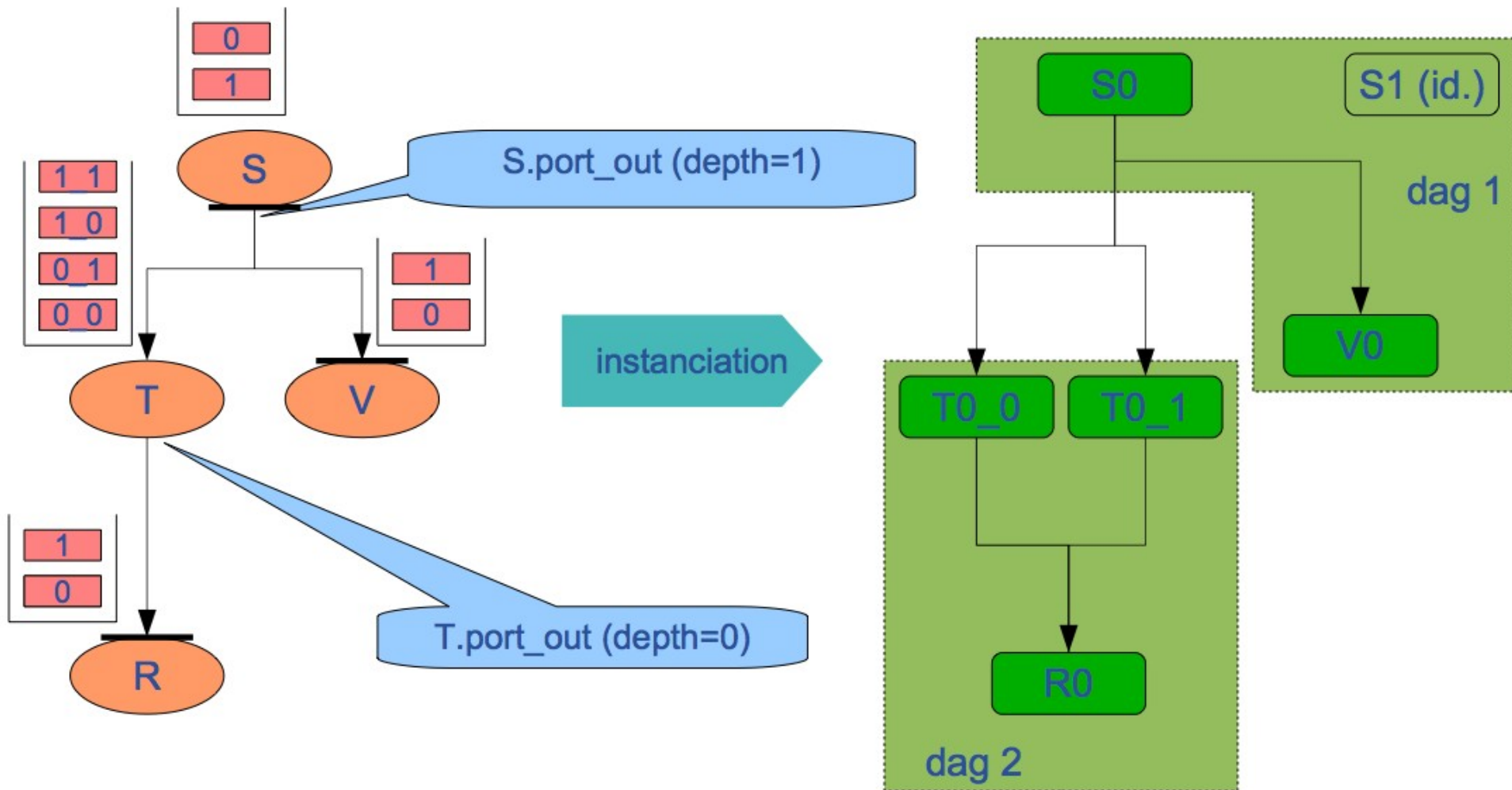
- Data-driven dynamic flow composition
- DAG generation

- Data flow resolved dynamically
- Application services invoked multiple times



Dynamic sub-dags generation

Grid Workflow Efficient Enactment for Data Intensive Applications



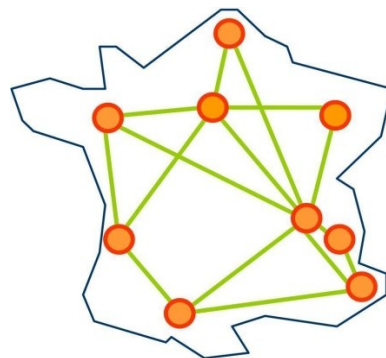
- **EGEE infrastructure**
 - Production
 - Batch-oriented
 - Homogeneous (SL4, gLite)



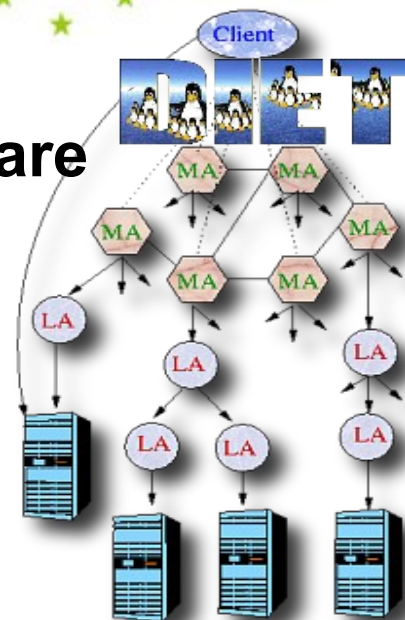
- **gLite middleware**
 - WMS jobs management system
 - MOTEUR workflow engine

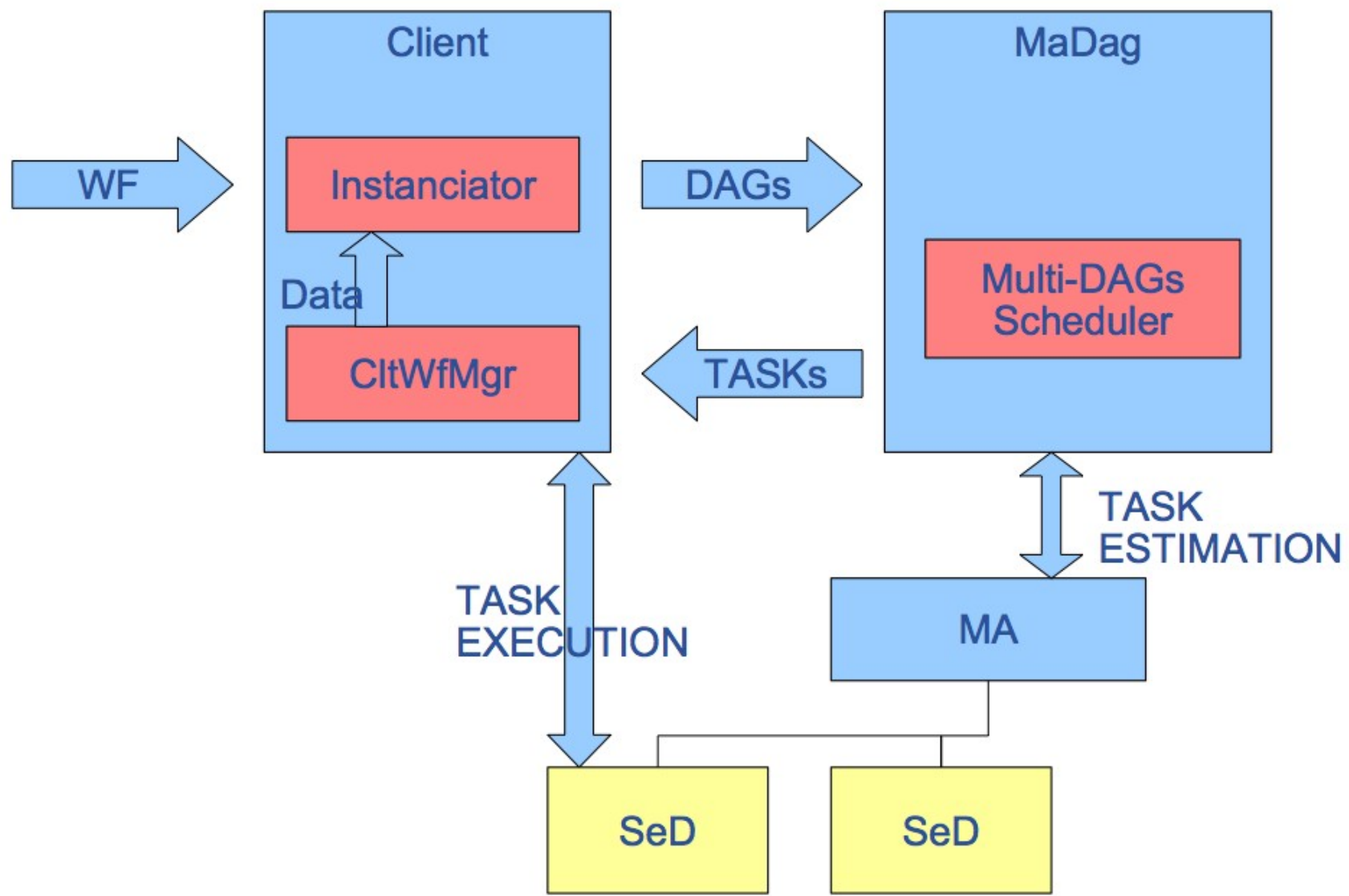


- **Grid'5000 infrastructure**
 - Research
 - Customizable
 - Middleware agnostic

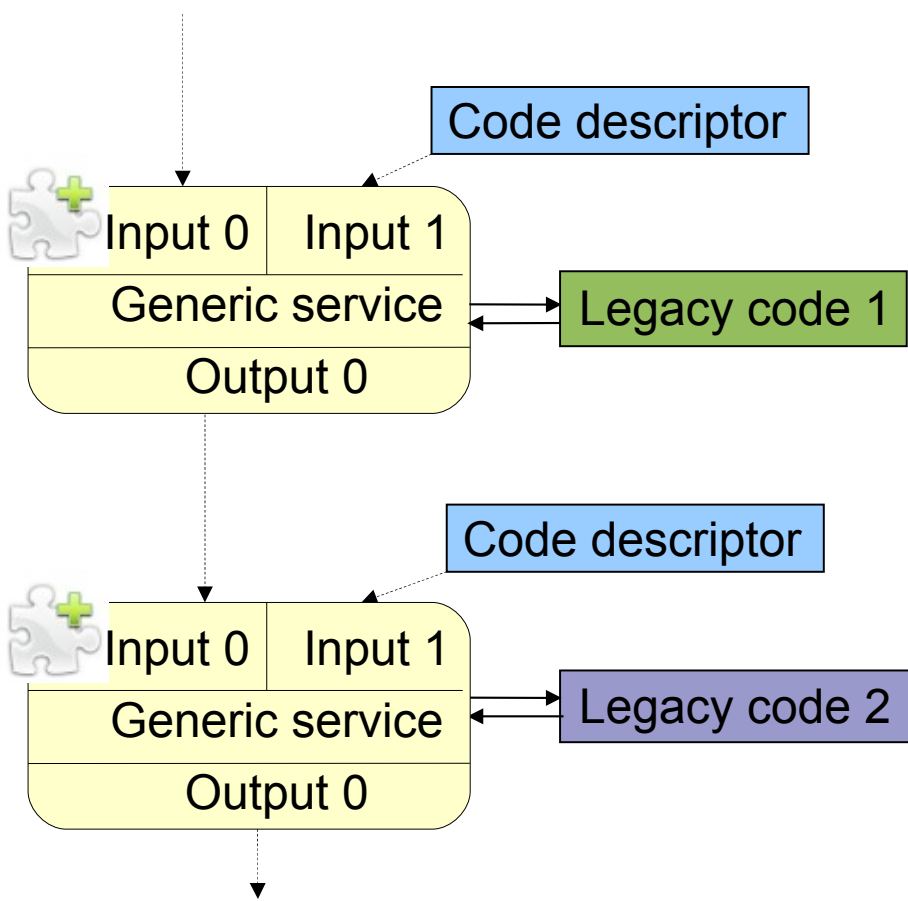


- **DIET middleware**
 - advanced schedulers
 - MA DAG workflow engine
 - SeD services





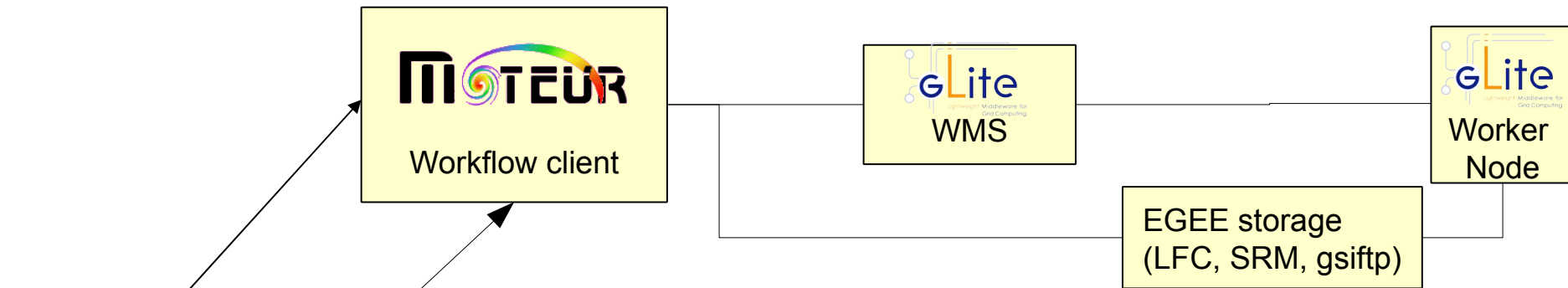
- Provide service wrapper to non instrumented code
- Handle data transfers (references to grid data)



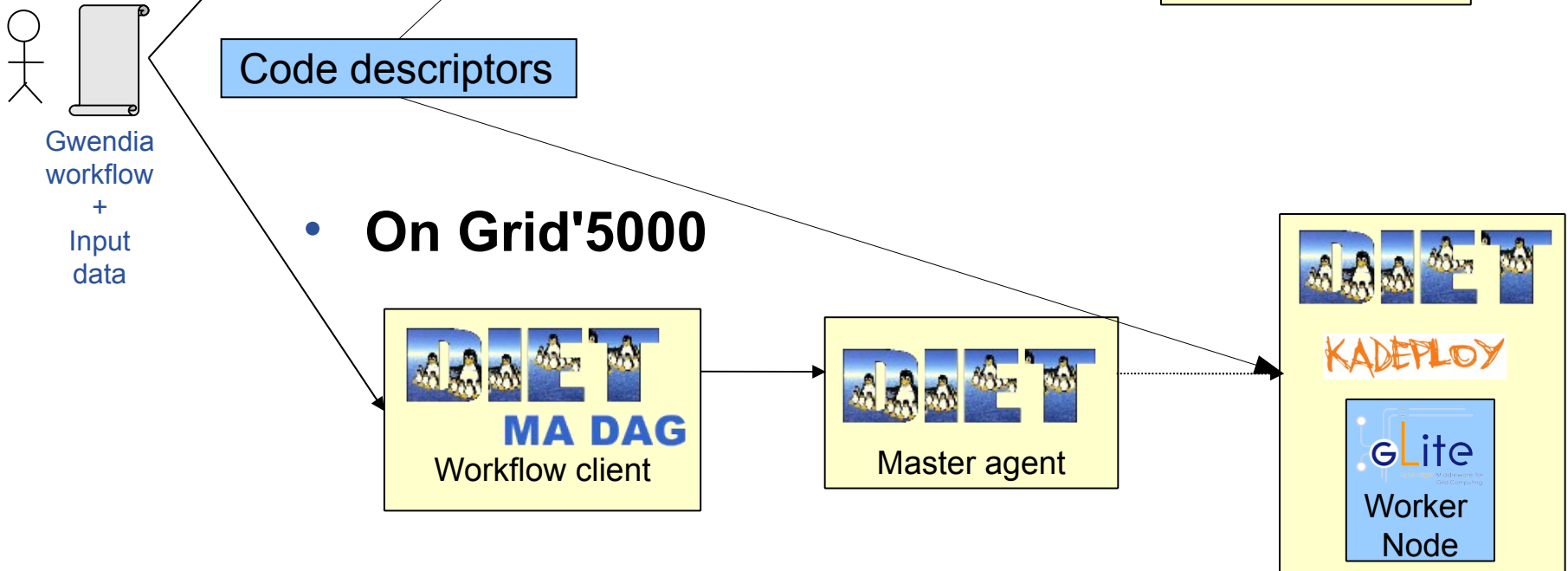
```

<description>
  <executable name="CrestLines.pl">
    <access type="URL">
      <path value="http://colors.unice.fr:80/" />
    </access>
    <value value="CrestLines.pl" />
  </executable>
  <input name="image" option="-im1">
    <access type="LFN" />
  </input>
  <input name="scale" option="-s" />
  <output name="crest_lines" option="-c2">
    <access type="LFN" />
  </output>
  <sandbox name="convert8bits">
    <access type="URL">
      <path value="http://colors.unice.fr:80/" />
    </access>
    <value value="Convert8bits.pl" />
  </sandbox>
</executable>
</description>
  
```

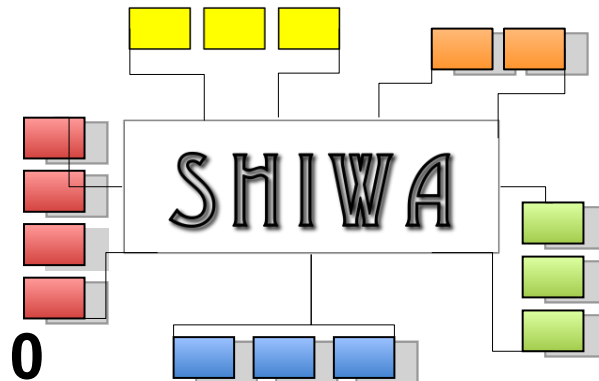
- On the EGEE grid



- On Grid'5000



- **Cardiac image analysis**
 - Complex use case with advanced data flows composition and enactment requirements
- **GWENDIA language**
 - Data-driven language based on array programming principles
 - Different enactors
 - MOTEUR (service-oriented), <http://modalis.polytech.unice.fr/software>
 - DIET MA DAG (DAG scheduler), <http://graal.ens-lyon.fr/~diet>
- **Grid interoperability at workflow level**
 - Language interoperability
 - Application services



➔ **SHIWA EU project, starting July 1st 2010**

- Application communities with need for wf interoperability wanted
- Open positions (research engineer, postdoc)