

Year 1 Internal Annual Presentation

Ketan MAHESHWARI

I3S-UNSA

November 17, 2008



Outline

1 Literature Review

- Taverna
- Triana
- Kepler

2 Enacting Cardiac Image Processing Application on MOTEUR

3 Work on Workflow Adapters

4 Work on gLite interface to Taverna v2

5 Future Work

Taverna

- SCUFL (Simple Conceptual Unified Flow Language) to express workflow.
- Advanced list processing operators.
- Not Grid Oriented.
- Good integration with Webservice invocation.
- Sophisticated User Interface.
- Excellent community support via myexperiment.org.

Triana

- Pure java based scientific workflow system.
- Uses simple WSDL-like XML to express the workflow.
- Third party add-ons for grid and P2P computation.(eg. Grid Application Toolkit)
- Text and File processing capabilities of java are exploited.
- Data and file transfer capabilities for 'HTTP', 'FTP', 'GSIFTP' and 'file:/' protocols.
- No inter-task list processing.

Kepler

- Application level dataflow
- MoML (Modeling Markup Language) to express workflow.
- Separate paradigms for handling sequential (SDF) and parallel (PN) dataflow.
- “actor-director” metaphor for enactment.
- Dedicated provenance handling framework.
- Grid Support using RSL(Resource Specification Language)-based Globus job.

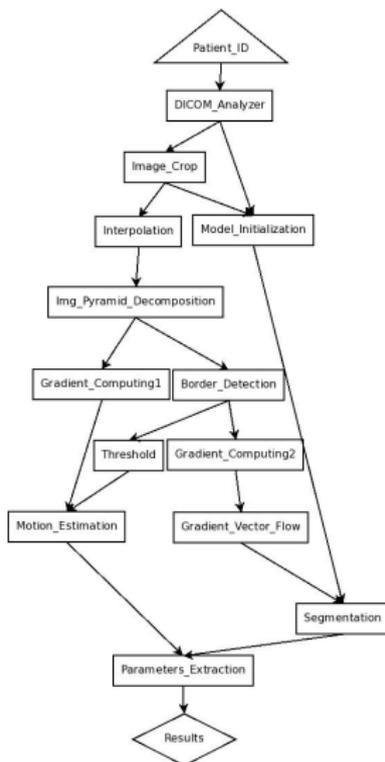
Findings

- Poor or No Support for Grid.
- Data Movement is mostly application-customized.
- Workflow engines are single strongest candidates to bridge applications and grids!
- Workflow interoperability does not exist.

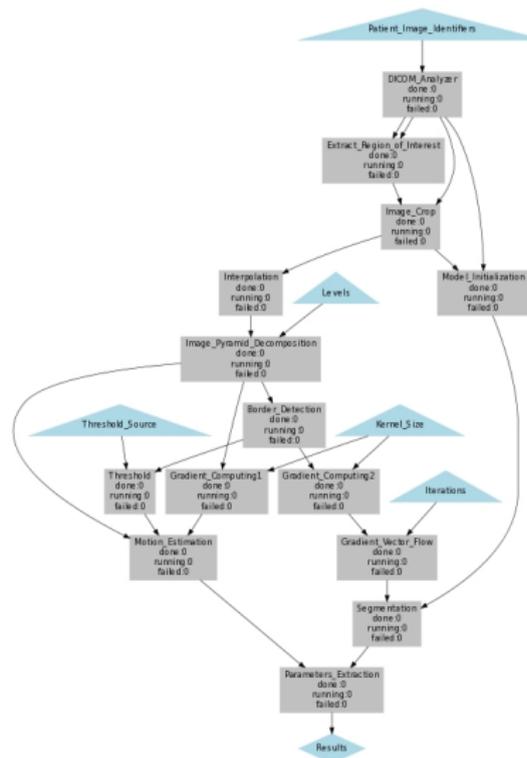
Outline

- 1 Literature Review
 - Taverna
 - Triana
 - Kepler
- 2 Enacting Cardiac Image Processing Application on MOTEUR
- 3 Work on Workflow Adapters
- 4 Work on gLite interface to Taverna v2
- 5 Future Work

Cardiac Image Application

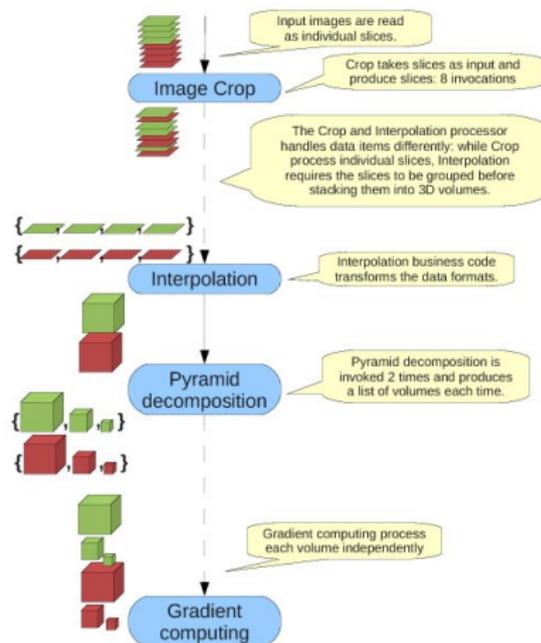


MOTEUR Enactment



Limitations of Current Enactment

- Expressibility
- Data Flow
- Notifications



1

¹Image Source: J. Montagnat

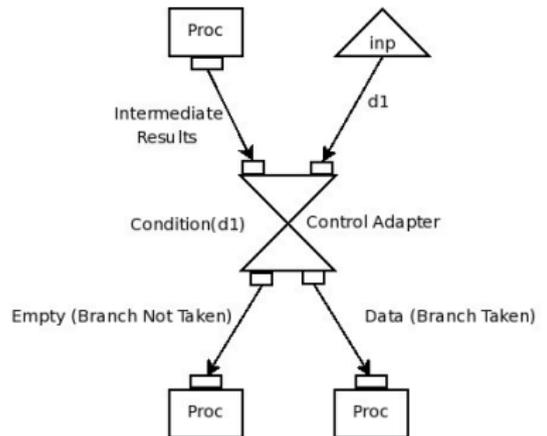
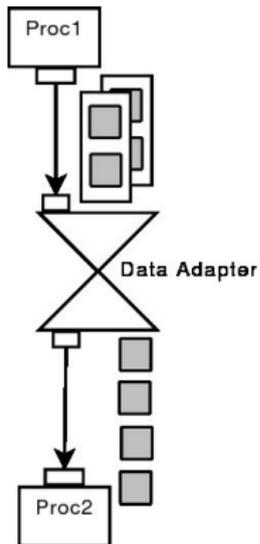
Outline

- 1 Literature Review
 - Taverna
 - Triana
 - Kepler
- 2 Enacting Cardiac Image Processing Application on MOTEUR
- 3 Work on Workflow Adapters
- 4 Work on gLite interface to Taverna v2
- 5 Future Work

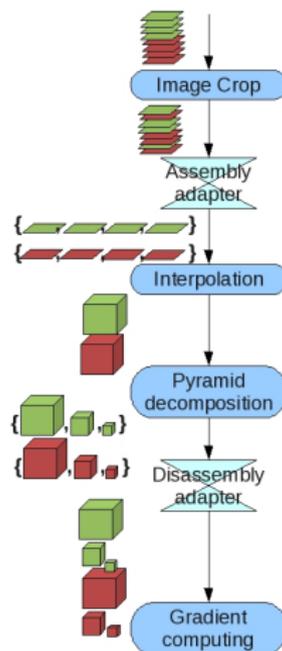
Motivation

- Correct Implementation of complex dataflow
- Implementing control flow
- Intuitive dataflow
- Exploit parallelization

Data and Control Adapters



Cardiac Workflow with Adapters



2

²Image Source: J. Montagnat

Outline

- 1 Literature Review
 - Taverna
 - Triana
 - Kepler
- 2 Enacting Cardiac Image Processing Application on MOTEUR
- 3 Work on Workflow Adapters
- 4 Work on gLite interface to Taverna v2
- 5 Future Work

Motivation

- Bring EGEE and Taverna communities closer.
- Taverna's sophisticated User Interface.
- Take inspiration from Taverna's list handling facilities.
- Mixed Mode Execution.

Screenshot

The screenshot displays the Taverna v2 interface with a workflow configuration window open. The workflow is titled "Image_Pyramid_Decomposition" and is part of a larger workflow named "dataflow2".

Workflow Configuration Window:

- Properties:** Type: Job, JobType: Normal, Nodes: 0, StdOut: stdout, StdErr: stderr.
- InputSandbox:** "Image_Pyramid_Decomposition.sh", "config.ini"
- OutputSandbox:** "stdout", "stderr"
- Executable:** Image_Pyramid_Decomposition.sh
- Arguments:** (empty)
- Job Requirements:** (duction') && RegExp("nf", other: GlueCEUniqueId)
- Retry Count:** 3
- Inputs Path:** /ketan/ManchesterWork/gliteworkflows/inputs/

Workflow Explorer:

- dataflow2
 - Inputs
 - input
 - Outputs
 - output
 - Processors
 - Image_Crop
 - Interpolation
 - Image_Pyramid_Decomposition (highlighted)
 - Gradient_Computing
 - Border_Detection
 - Motion_Estimation
 - Data links
 - input -> datain
 - dataout -> datain2
 - dataout -> output

Contextual View:

gLite Contextual View		
Input Port Name	Depth	
Output Port Name	Depth	Granular Depth

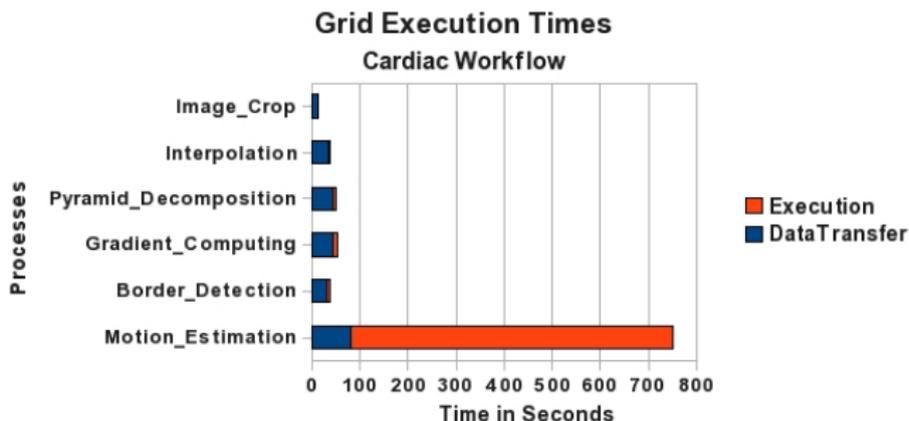
Workflow Diagram:

```

graph TD
    subgraph Workflow_Inputs
        Input[input]
    end
    Input --> Image_Crop[Image_Crop]
    Image_Crop --> Interpolation[Interpolation]
    Interpolation --> Image_Pyramid_Decomposition[Image_Pyramid_Decomposition]
    Image_Pyramid_Decomposition --> Gradient_Computing[Gradient_Computing]
    Image_Pyramid_Decomposition --> Border_Detection[Border_Detection]
    Gradient_Computing --> Motion_Estimation[Motion_Estimation]
    Border_Detection --> Motion_Estimation
    subgraph Workflow_Outputs
        Output[output]
    end
    Motion_Estimation --> Output
  
```

First Results

Patients=1, image slice sets=2 (1 fixed, 1 moving), total data size \simeq 5MB;



Current Limitations

- Needs a UI for initial data transfer.
- Polling at high frequency (I am looking into notification mechanism).
- Not yet asynchronous.

Conclusions

- Grid/Middlewares are not workflow oriented, they are batch oriented.
- Sequential execution only. Paradox??
- Plenty of manipulations required to make grid understand dataflow.

Outline

- 1 Literature Review
 - Taverna
 - Triana
 - Kepler
- 2 Enacting Cardiac Image Processing Application on MOTEUR
- 3 Work on Workflow Adapters
- 4 Work on gLite interface to Taverna v2
- 5 Future Work

- Implementing Adapters.
- Functional vs Dataflow Languages and their interoperability.
- Enactment of Drug Discovery case.
- Implementing Grid Notification.
- Writing 300 pages of Thesis!!!

Thank You! Questions and Suggestions are Welcome!!