

# RINGrid

## Evaluation of Remote Instrumentation Infrastructures

Markus Baumgartner  
on behalf of  
Thomas Prokosch

GUP, Joh. Kepler University Linz, Austria

EGEE 2007

Funded by the EC under 6FP  
(contract number 031891).



**RINGrid**



# Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil

# Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil
- challenges
  - travel costs

# Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil
- challenges
  - travel costs
  - time used for travelling

# Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil
- challenges
  - travel costs
  - time used for travelling
  - experiments may fail

# Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil
- challenges
  - travel costs
  - time used for travelling
  - experiments may fail
  - limited expertise on instrument

# Remote Instrumentation (RI)

- experiments on large instruments
  - Southern Astrophysical Research Telescope (optical, 4.1 m): Chile
  - electron spectrometer: Bulgaria
  - satellite network (24 stations, mesh): Italy
  - synchrotron light source (11 beam lines): Brazil
- challenges
  - travel costs
  - time used for travelling
  - experiments may fail
  - limited expertise on instrument
- idea: use these instruments over the Internet
- focus on LA countries (Chile, Brazil)



# RI Advantages

- resources are shared
- **instruments are better utilized**
- expertise is increased in LA countries
- new communities formed
- expenses are justified
- **better instruments will be bought**
- new opportunities created



# Problem Definition

What does an experiment look like?

- raw data
- data processing
  
- filtered data (results)

# Problem Definition

What does an experiment look like?

- raw data (storage on the grid)
- data processing (on the grid)
  
- filtered data (results, stored on the grid)

# Problem Definition

What does an experiment look like?

- data retrieval
- raw data (storage on the grid)
- data processing (on the grid)
  
- filtered data (results, stored on the grid)

# Problem Definition

What does an experiment look like?

- data retrieval (instrument as **grid component**)
- raw data (storage on the grid)
- data processing (on the grid)
  
- filtered data (results, stored on the grid)

# Problem Definition

What does an experiment look like?

- data retrieval (instrument as **grid component**)
- raw data (storage on the grid)
- data processing (on the grid)
- collaboration (**needs grid support**, timezones)
- filtered data (results, stored on the grid)

# Problem Definition

What does an experiment look like?

- data retrieval (instrument as **grid component**)
- raw data (storage on the grid)
- data processing (on the grid)
- collaboration (**needs grid support**, timezones)
- filtered data (results, stored on the grid)

workflow management (in need of **grid support**)

# Research Approach

Traditional approach:

- What is there?
- What is missing?
- Develop an architecture.
- Verification: Test the architecture.

# Research Approach

Traditional approach:

- What is there?
- What is missing?
- Develop an architecture.
- Verification: Test the architecture.





# Middleware Components for RI

Experiments conducted remotely require:

- 1 workflow management
- 2 interactive experiment steering
- 3 data transmission
- 4 data storage
- 5 collaboration

# Workflow Management

“Where can data for the next step in the experiment be retrieved?”

- dependencies inherent within an experimental setup
- for automating processes, i.e. experiments

data from instruments have to be **stored near** the instrument  
data needs to be **forwarded** to other hosts for processing  
need to **keep track** of data



# Workflow Management Solutions

We had a look at:

- **g-Eclipse**  
workflow capability cannot be used independently
- **Yet Another Workflow Language (YAWL)**  
execution engine tied to graphical editor
- **XML Process Definition Language (XPDL)**  
create diagrams based on the XML files
- **Business Process Execution Language (BPEL)**  
long-running applications  
span multiple organizational entities

# Middleware Components for RI

Experiments conducted remotely require:

- 1 workflow management
- 2 interactive experiment steering
- 3 data transmission
- 4 data storage
- 5 collaboration

# Interactive Experiment Steering

utilize rapid turnaround

try several variations of parameters → better results



# Interactive Experiment Steering

utilize rapid turnaround

try several variations of parameters → better results

## Requirements:

- **display of** (preliminary) **experimental data** (simple/complex)  
data rate, evtl. latency
- **show control elements for instrument**  
latency

# Visualization Components

## glogin

- bi-directional channels
- forward data securely (X11, TCP, VPNs, shell)
- no account needed

## GVid

video rendering on any node, complex tasks split

- transmission of video data to user's desktop
- interaction elements communicated back

**missing:** adapt components for RI (latency, etc.)  
encoding of instrument controls





# Middleware Components for RI

Experiments conducted remotely require:

- 1 workflow management
- 2 interactive experiment steering
- 3 data transmission**
- 4 data storage
- 5 collaboration

# Data Transmission

Transmission qualities:

- high connection bandwidth
- reasonable round trip time
- small jitter
- no packet loss

# Data Transmission

Transmission qualities:

- high connection bandwidth
- reasonable round trip time
- small jitter
- no packet loss

Luckily, not all applications demand all these qualities.

**Classify applications** by the requirements and choose underlying network connectivity appropriately.



# Data Transmission Gaps

We have already done a first classification of RI applications.

**Missing:**

Verification of classification by conducting experiments.

# Middleware Components for RI

Experiments conducted remotely require:

- 1 workflow management
- 2 interactive experiment steering
- 3 data transmission
- 4 data storage**
- 5 collaboration

# Data Storage

huge amounts of data (think LHC)  
not a problem for a single experiment  
data archive/digital library

# Data Storage

huge amounts of data (think LHC)  
not a problem for a single experiment  
data archive/digital library

- 1 reduce data to a reasonable amount  
data reduction, compression, preliminary analysis

# Data Storage

huge amounts of data (think LHC)  
not a problem for a single experiment  
data archive/digital library

- 1 reduce data to a reasonable amount  
data reduction, compression, preliminary analysis
- 2 distribute data among several grid nodes



# Data Storage

huge amounts of data (think LHC)  
not a problem for a single experiment  
data archive/digital library

- 1 reduce data to a reasonable amount  
data reduction, compression, preliminary analysis
- 2 distribute data among several grid nodes
- 3 retrieval of data with data **management service**

# Data Storage: Pandora's Box

- **data movement**  
done: GridFTP/RFT (Reliable File Transfer)
- **data replication and access**  
good management of data metadata
- **data consistency**  
data locking, lazy-copy (consistency problems)
- **movement planning**  
access patterns, statistical methods
- **replica management**  
finding an optimal place for replicas is  
NP-complete (Wolfson, Milo)

# Middleware Components for RI

Experiments conducted remotely require:

- 1 workflow management
- 2 interactive experiment steering
- 3 data transmission
- 4 data storage
- 5 collaboration

# Collaboration

good thing: no hard technical restrictions

## ■ communication

- instant messaging software like Skype, Gadu-Gadu
- custom software, like GridCC's "Multipurpose Collaboration Environment" (MCE)

# Collaboration

good thing: no hard technical restrictions

## ■ communication

- instant messaging software like Skype, Gadu-Gadu
- custom software, like GridCC's "Multipurpose Collaboration Environment" (MCE)

## ■ common file system

- Storage Resource Broker (SRB by San Diego Supercomputer Center)

# Collaboration

good thing: no hard technical restrictions

## ■ communication

- instant messaging software like Skype, Gadu-Gadu
- custom software, like GridCC's "Multipurpose Collaboration Environment" (MCE)

## ■ common file system

- Storage Resource Broker (SRB by San Diego Supercomputer Center)

**Problem:** Missing integration.



# Summary

- 1 adaptation of a workflow management solution for RI
- 2 adaptation of visualization components
- 3 verify application classification in terms of data transmission
- 4 data storage: replica management