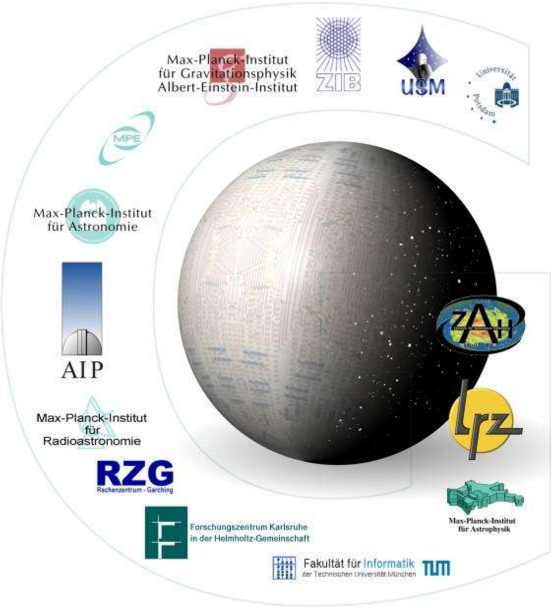


Application Monitoring & Steering in AstroGrid-D

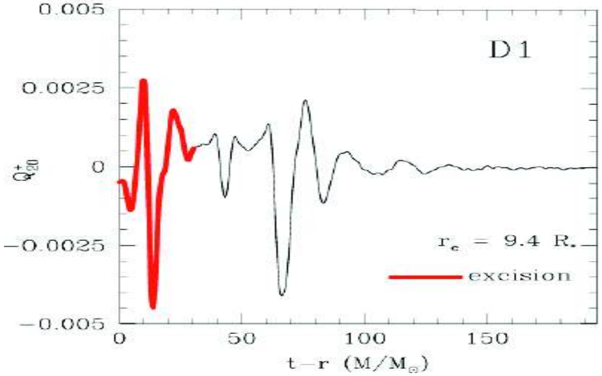
Thomas Radke
Max Planck Institute for Gravitational Physics
Albert Einstein Institute (AEI)
(with contributions from Frank Breitling
Astrophysical Institute Potsdam (AIP))

HEP Grid Monitoring Workshop
20/21 June 2007, University of Wuppertal

Gravitational Wave Astronomy @ AEI

- AEI is one of the leading research institutes in the international GW astronomy community and a member of LIGO, the *Laser Interferometer GW Observatory*

- GW group is trying to detect and measure waves using interferometers
- NumRel group is solving the EE numerically to predict waveform templates for cosmic events which are likely sources of GW emission



- *Cactus* is used as the computational framework for numerical simulations

AstroGrid-D Use Cases for Cactus

- Cactus is one of the astrophysical applications which are to be grid-enabled in AstroGrid-D
- the *Cactus Computational Toolkit* serves as a framework to develop application-specific numerical simulation codes
- ~20 physicists at AEI, with collaborators in Italy, UK and USA
- most users also contribute to the code, turning distributed code management into a non-trivial task
- two Cactus use cases have been identified with specific requirements on job / application monitoring:
  I. Automated Integration Tests
  II. Online Simulation Monitoring & Interactive Steering

Cactus Use Case I:
Automated Integration Tests

Motivation: provide a service for Cactus programmers
to monitor results from code validation test runs

1. Definition and flexible description of application-specific simulation monitoring data
2. Collection of Cactus metadata from remote testing/validation procedures such as nightly integration tests
3. Metadata management by an information service with persistent datastore backend and query engine
4. Querying, retrieval, and presentation of monitoring metadata to the end-user through a portal user interface

AstroGrid-D Metadata Management

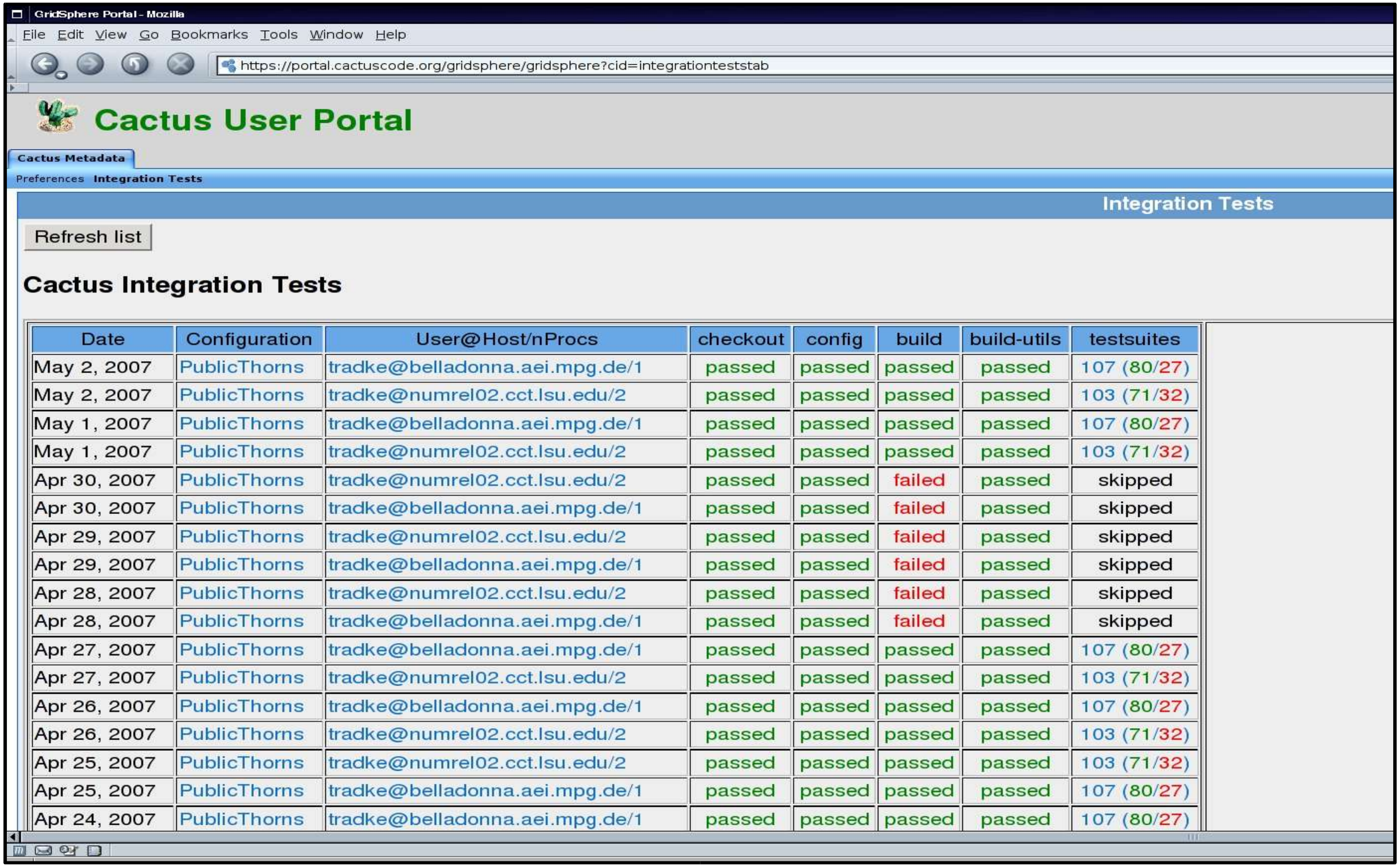
- in AstroGrid-D we use RDF as a common data model to describe metadata about both resources and applications
- ZIB has developed *Stellaris*, an AstroGrid-D Information Service with an RDF storage and a SPARQL query engine
- an RDF schema defines the vocabulary for our monitoring data
- monitoring data are collected by a Perl script which is scheduled to run as a nightly cron job on selected production machines
  1) triggers the execution of individual Cactus unit tests
     `checkout, configure, build, build-utils, testsuites`
  2) analyses the logfile(s) output of each test
  3) generates an intermediate RDF/XML document from it
  4) sends it off to an external *Stellaris* information service

Monitoring User Interface

A *Cactus Portal* serves as a standard web-based user interface to query and retrieve monitoring data from Cactus integration tests.

- implemented with the *GridSphere portal framework*
- an RDF portlet was written for GridSphere which
  - dynamically generates SPARQL queries to search in metadata
  - retrieves the query results from *Stellaris*
  - presents them in different interlinked views
- users can also refine their queries via preferences
  - search for simulations by configuration name, user, host
  - limit the list of returned query results to the first N entries
- access to monitoring data gathered since December 2006

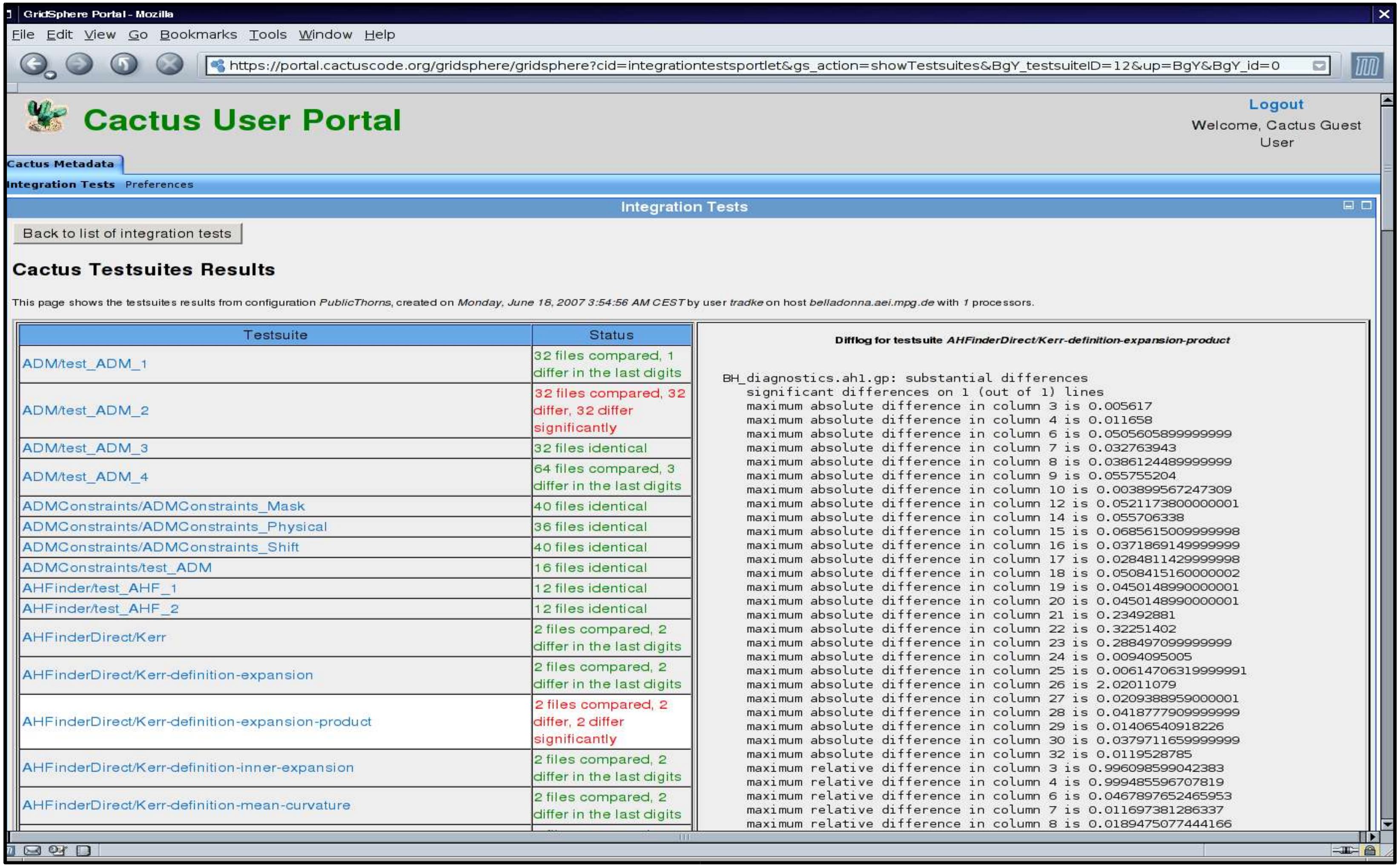
Automated Cactus Integration Tests


GridSphere Portal - Mozilla
File Edit View Go Bookmarks Tools Window Help
https://portal.cactuscode.org/gridsphere/gridsphere?cid=integrationteststab
Cactus User Portal
Cactus Metadata
Preferences Integration Tests
Integration Tests
Refresh list

Cactus Integration Tests

Date	Configuration	User@Host/nProcs	checkout	config	build	build-utils	testsuites
May 2, 2007	PublicThorns	tradke@belladonna.aei.mpg.de/1	passed	passed	passed	passed	107 (80/27)
May 2, 2007	PublicThorns	tradke@numrel02.cct.lsu.edu/2	passed	passed	passed	passed	103 (71/32)
May 1, 2007	PublicThorns	tradke@belladonna.aei.mpg.de/1	passed	passed	passed	passed	107 (80/27)
May 1, 2007	PublicThorns	tradke@numrel02.cct.lsu.edu/2	passed	passed	passed	passed	103 (71/32)
Apr 30, 2007	PublicThorns	tradke@numrel02.cct.lsu.edu/2	passed	passed	failed	passed	skipped
Apr 30, 2007	PublicThorns	tradke@belladonna.aei.mpg.de/1	passed	passed	failed	passed	skipped
Apr 29, 2007	PublicThorns	tradke@numrel02.cct.lsu.edu/2	passed	passed	failed	passed	skipped
Apr 29, 2007	PublicThorns	tradke@belladonna.aei.mpg.de/1	passed	passed	failed	passed	skipped
Apr 28, 2007	PublicThorns	tradke@numrel02.cct.lsu.edu/2	passed	passed	failed	passed	skipped
Apr 28, 2007	PublicThorns	tradke@belladonna.aei.mpg.de/1	passed	passed	failed	passed	skipped
Apr 27, 2007	PublicThorns	tradke@belladonna.aei.mpg.de/1	passed	passed	passed	passed	107 (80/27)
Apr 27, 2007	PublicThorns	tradke@numrel02.cct.lsu.edu/2	passed	passed	passed	passed	103 (71/32)
Apr 26, 2007	PublicThorns	tradke@belladonna.aei.mpg.de/1	passed	passed	passed	passed	107 (80/27)
Apr 26, 2007	PublicThorns	tradke@numrel02.cct.lsu.edu/2	passed	passed	passed	passed	103 (71/32)
Apr 25, 2007	PublicThorns	tradke@numrel02.cct.lsu.edu/2	passed	passed	passed	passed	103 (71/32)
Apr 25, 2007	PublicThorns	tradke@belladonna.aei.mpg.de/1	passed	passed	passed	passed	107 (80/27)
Apr 24, 2007	PublicThorns	tradke@belladonna.aei.mpg.de/1	passed	passed	passed	passed	107 (80/27)

Automated Cactus Integration Tests


GridSphere Portal - Mozilla
File Edit View Go Bookmarks Tools Window Help
https://portal.cactuscode.org/gridsphere/gridsphere?cid=integrationtestsportlet&gs_action=showTestsuites&BgY_testsuiteID=12&up=BgY&BgY_id=0
Cactus User Portal
Logout
Welcome, Cactus Guest User
Cactus Metadata
Integration Tests
Preferences
Integration Tests
Back to list of integration tests

Cactus Testsuites Results

This page shows the testsuites results from configuration *PublicThorns*, created on *Monday, June 18, 2007 3:54:56 AM CEST* by user *tradke* on host *belladonna.aei.mpg.de* with *1* processors.

Testsuite	Status
ADM/test_ADM_1	32 files compared, 1 differ in the last digits
ADM/test_ADM_2	32 files compared, 32 differ, 32 differ significantly
ADM/test_ADM_3	32 files identical
ADM/test_ADM_4	64 files compared, 3 differ in the last digits
ADMConstraints/ADMConstraints_Mask	40 files identical
ADMConstraints/ADMConstraints_Physical	36 files identical
ADMConstraints/ADMConstraints_Shift	40 files identical
ADMConstraints/test_ADM	16 files identical
AHFinder/test_AHF_1	12 files identical
AHFinder/test_AHF_2	12 files identical
AHFinderDirect/Kerr	2 files compared, 2 differ in the last digits
AHFinderDirect/Kerr-definition-expansion	2 files compared, 2 differ in the last digits
AHFinderDirect/Kerr-definition-expansion-product	2 files compared, 2 differ, 2 differ significantly
AHFinderDirect/Kerr-definition-inner-expansion	2 files compared, 2 differ in the last digits
AHFinderDirect/Kerr-definition-mean-curvature	2 files compared, 2 differ in the last digits

Difflog for testsuite *AHFinderDirect/Kerr-definition-expansion-product*

```
BH_diagnostics.ah1.gp: substantial differences
   significant differences on 1 (out of 1) lines
   maximum absolute difference in column 3 is 0.005617
   maximum absolute difference in column 4 is 0.011658
   maximum absolute difference in column 6 is 0.0505605899999999
   maximum absolute difference in column 7 is 0.032763943
   maximum absolute difference in column 8 is 0.0386124489999999
   maximum absolute difference in column 9 is 0.055755204
   maximum absolute difference in column 10 is 0.003899567247309
   maximum absolute difference in column 12 is 0.0521173800000001
   maximum absolute difference in column 14 is 0.055706338
   maximum absolute difference in column 15 is 0.0685615009999998
   maximum absolute difference in column 16 is 0.0371869149999999
   maximum absolute difference in column 17 is 0.0284811429999998
   maximum absolute difference in column 18 is 0.0508415160000002
   maximum absolute difference in column 19 is 0.0450148990000001
   maximum absolute difference in column 20 is 0.0450148990000001
   maximum absolute difference in column 21 is 0.23492881
   maximum absolute difference in column 22 is 0.32251402
   maximum absolute difference in column 23 is 0.288497099999999
   maximum absolute difference in column 24 is 0.0094095005
   maximum absolute difference in column 25 is 0.00614706319999991
   maximum absolute difference in column 26 is 2.02011079
   maximum absolute difference in column 27 is 0.0209388959000001
   maximum absolute difference in column 28 is 0.0418777909999999
   maximum absolute difference in column 29 is 0.01406540918226
   maximum absolute difference in column 30 is 0.0379711659999999
   maximum absolute difference in column 32 is 0.0119528785
   maximum relative difference in column 3 is 0.996098599042383
   maximum relative difference in column 4 is 0.999485596707819
   maximum relative difference in column 6 is 0.0467897652465953
   maximum relative difference in column 7 is 0.011697381286337
   maximum relative difference in column 8 is 0.0189475077444166
```

Cactus Use Case II: Online Simulation Monitoring & Steering

Motivation for users:

- to be able to monitor the progress of long-running simulations
  - current physical simulation time, iteration number
  - initial and current parameter settings
  - location(s) and list of data files generated
  - pre-visualisation of intermediate results
- have the possibility to directly interact with the simulation
  - change steerable parameters at runtime
  - initiate a checkpoint
  - pause and debug the simulation
  - trigger safe termination of a Cactus run

Cactus Monitoring Thorn HTTPS

- thorn HTTPS is a web server built into the application code
- at simulation startup: opens a port on processor 0
  to listen for HTTP client requests
- registers a worker routine with the Cactus flesh
  which is periodically scheduled during evolution
  to accept and serve any waiting client requests
- generates dynamic HTML depending on the requests' URI
- provides an API for other thorns to plug in
  their own HTTP request handlers

User Authentication & Authorisation

- users demand a secure simulation access control mechanism
- thorn HTTPS implements authentication on both the server and the client side, based on X.509 certificates
- server certificates are automatically generated and self-signed
  - *Can we use something like a Grid service certificate here ?*
- clients can be requested to present a user certificate
- user certificate validation is done based on
  - (steerable) simulation parameters specified in the parfile

    ```
    String HTTPS::authorised_users       = “”
    String HTTPS::authorised_superusers  = “*”
    ```
  - a list of known CA certificates
(either built-in or read from a trusted CA directory)

Monitoring & Steering Functionality implemented on top of HTTPS

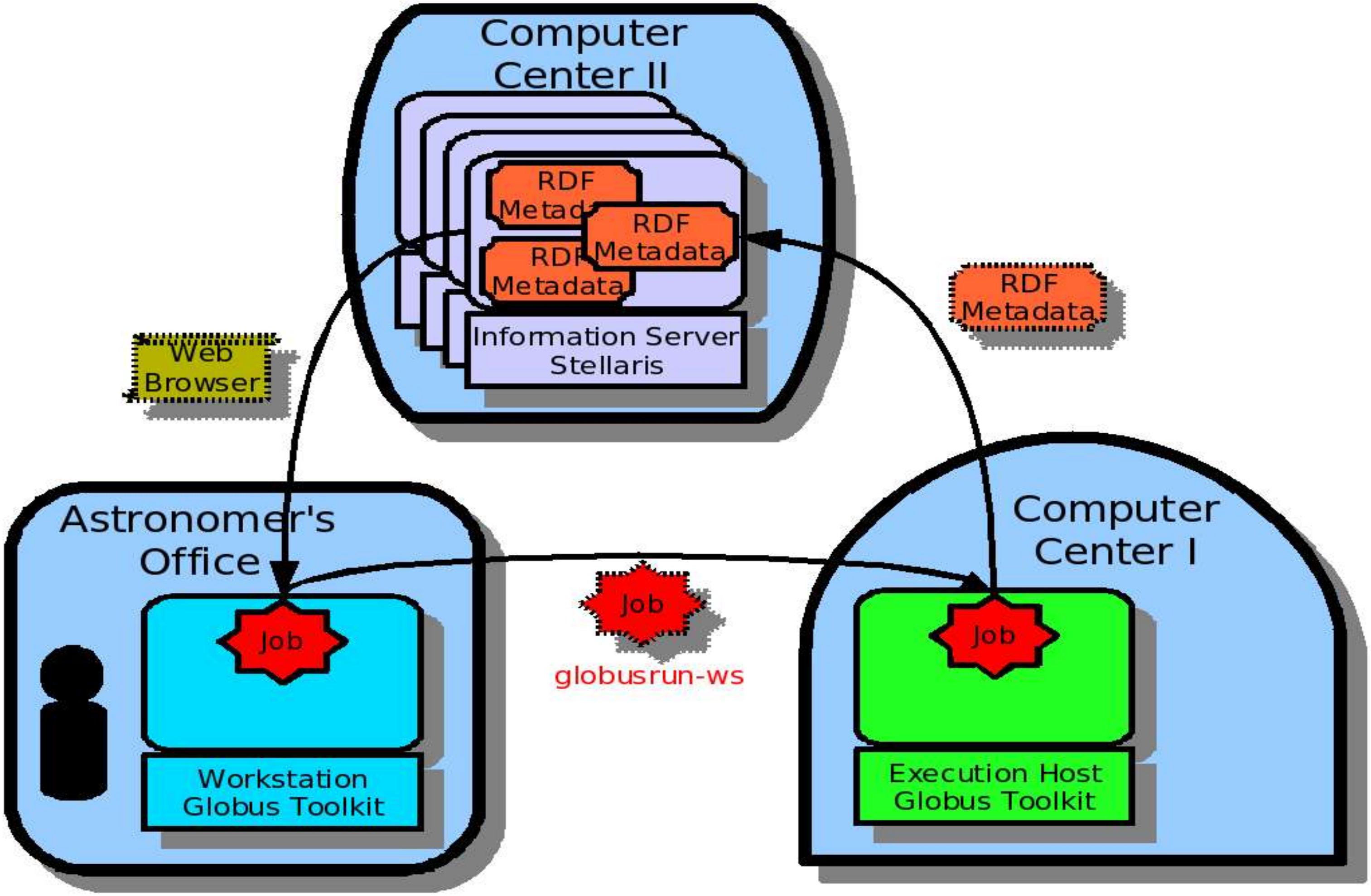
live examples of Cactus-specific online monitoring & interactive steering methods implemented so far

- simulation homepage
- online visualisation of intermediate simulation data
- interactive steering of simulation parameters

You can also check out our online demo example:

- perpetual wavetoy simulation
  `http://cactus.cct.lsu.edu:5555/`

AIP Use Case: AMIGA Job Monitoring



RDF/XML Template for Monitoring

```
<?ml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:gacgjobs='http://www.gac-grid.org/schema/jobs#'
  xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
>
  <gacgjobs:Job rdf:about="http://www.gac-grid.org/jobs/REPLACEJOBID">
    <gacgjobs:currentState>
      <rdf:Description rdf:about="http://www.gac-grid.org/REPLACEJOBID/state">
        <gacgjobs:timeStamp>RECPLACETIMESTAMP</gacgjobs:timeStamp>
        <gacgjobs:state>REPLACESTATE</gacgjobs:state>
      </rdf:Description>
    </gacgjobs:currentState>
    <gacgjobs:jobname>REPLACEJOBNAME</gacgjobs:jobname>
    <gacgjobs:site rdf:resource="http://REPLACEHOST"/>
    <gacgjobs:stageintime>REPLACESTAGEINTIME</gacgjobs:stageintime>
    <gacgjobs:stageouttime>REPLACESTAGEOUTTIME</gacgjobs:stageouttime>
    <gacgjobs:activetime>REPLACEACTIVETIME</gacgjobs:activetime>
    <gacgjobs:user>REPLACEUSER</gacgjobs:user>
    <gacgjobs:starttime>REPLACESTARTTIME</gacgjobs:starttime>
  </gacgjobs:Job>
</rdf:RDF>
```

in a Perl script

- This **RDF template** is updated with the current job states provided by the Globus Toolkit commands

```
#: globusrun-ws -status -job-epr-file ~/.epr/69ef6110-1dc1-11dc-bac0-0017313ead04.epr
#: Current job state: StageIn
```

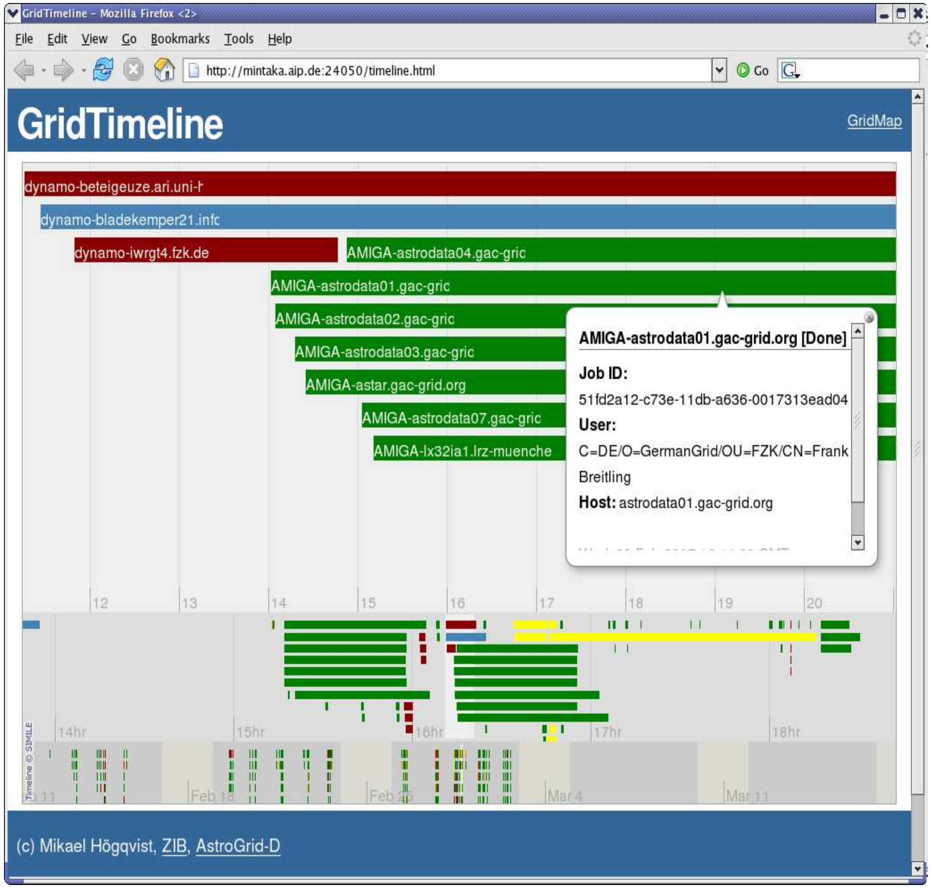
- and is uploaded to the AstroGrid-D Information Service **Stellaris**:

```
#: curl -d \@$RDFFile $InfoService/context/gridmap/jobs?action=update
```

Web Browser User Interface

for the visualization

- Developed in AstroGrid-D by ZIB (M. Högqvist)
- currently used to display job-state metadata
- Based on „Simile“ Timeline
    - like Google Maps for time-based information
    - a DHTML-based AJAXy widget for visualizing time-based events.
    - http://simile.mit.edu/timeline/


GridTimeline – Mozilla Firefox <2>
File Edit View Go Bookmarks Tools Help
Go
GridTimeline
GridMap
dynamo-beteigeuze.ari.uni-h
dynamo-bladekemper21.infc
dynamo-iwrgt4.fzk.de
AMIGA-astrodata04.gac-gric
AMIGA-astrodata01.gac-gric
AMIGA-astrodata02.gac-gric
AMIGA-astrodata03.gac-gric
AMIGA-astar.gac-grid.org
AMIGA-astrodata07.gac-gric
AMIGA-lx32ia1.lrz-muenche
12
13
14
15
16
17
18
19
20
14hr
15hr
16hr
17hr
18hr
Timeline © SIMILE
Feb 18
Feb 25
Mar 4
Mar 11

http://mintaka.aip.de:24050/timeline.html

AMIGA-astrodata01.gac-grid.org [Done]

Job ID:

51fd2a12-c73e-11db-a636-0017313ead04

User:

C=DE/O=GermanGrid/OU=FZK/CN=Frank Breitling

Host: astrodata01.gac-grid.org

(c) Mikael Högqvist, ZIB, AstroGrid-D

- available at http://www.gac-grid.org/project-products/ (Apache License)

Summary and Outlook

- We implemented different mechanisms to monitor and steer AstroGrid-D simulations. Although application-specific, their concepts can be adopted by other numerical codes.
- We make use of standard Web and Grid technologies
  - AstroGrid-D's *Stellaris* information service
  - the *GridSphere* portal framework
  - X.509 certificate based user authentication

  but hide most of their complexity from the end user.

- support for simulations running in closed environments
- integration of monitoring/steering methods in the portal
  - simulation tracking and metadata searches

Links to Further Information

- Cactus portal installations
  - public user portal with guest user account
    `https://portal.cactuscode.org`
  - production portal for the NumRel groups at AEI and CCT
    `https://portal.aei.mpg.de`
- Cactus homepage: `http://www.cactuscode.org`
- GridSphere homepage: `http://www.gridsphere.org`
- AstroGrid-D homepage: `http://www.gac-grid.org`
- AstroGrid-D software products:
  `http://www.gac-grid.org/project-products.html`

Q/A & Discussion

Suggestions to initiate a discussion

- Is there interest in the HEP community in using some of AstroGrid-D's monitoring concepts/tools/services ?
  I can put you in contact with the developers.
- Do you have monitoring solutions we could use ?
  - job/application monitoring in closed environments
  - job statistics, resource/service quality evaluation
  - user-friendly interfaces for presenting monitoring data
- Can we identify synergies to develop common solutions ?
  - request support from DGI-2 (eg. D-MON)
  - prepare D-Grid 3 call