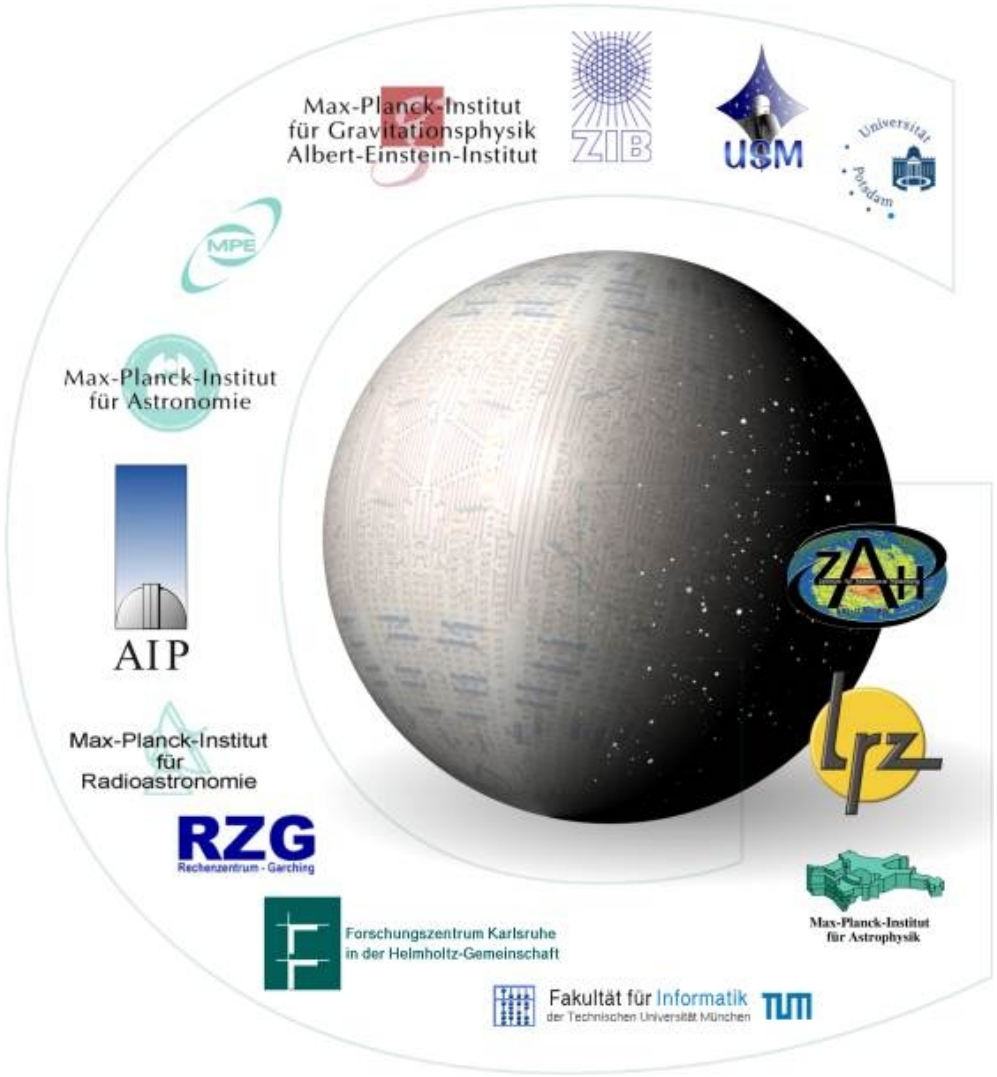
AstroGrid-D



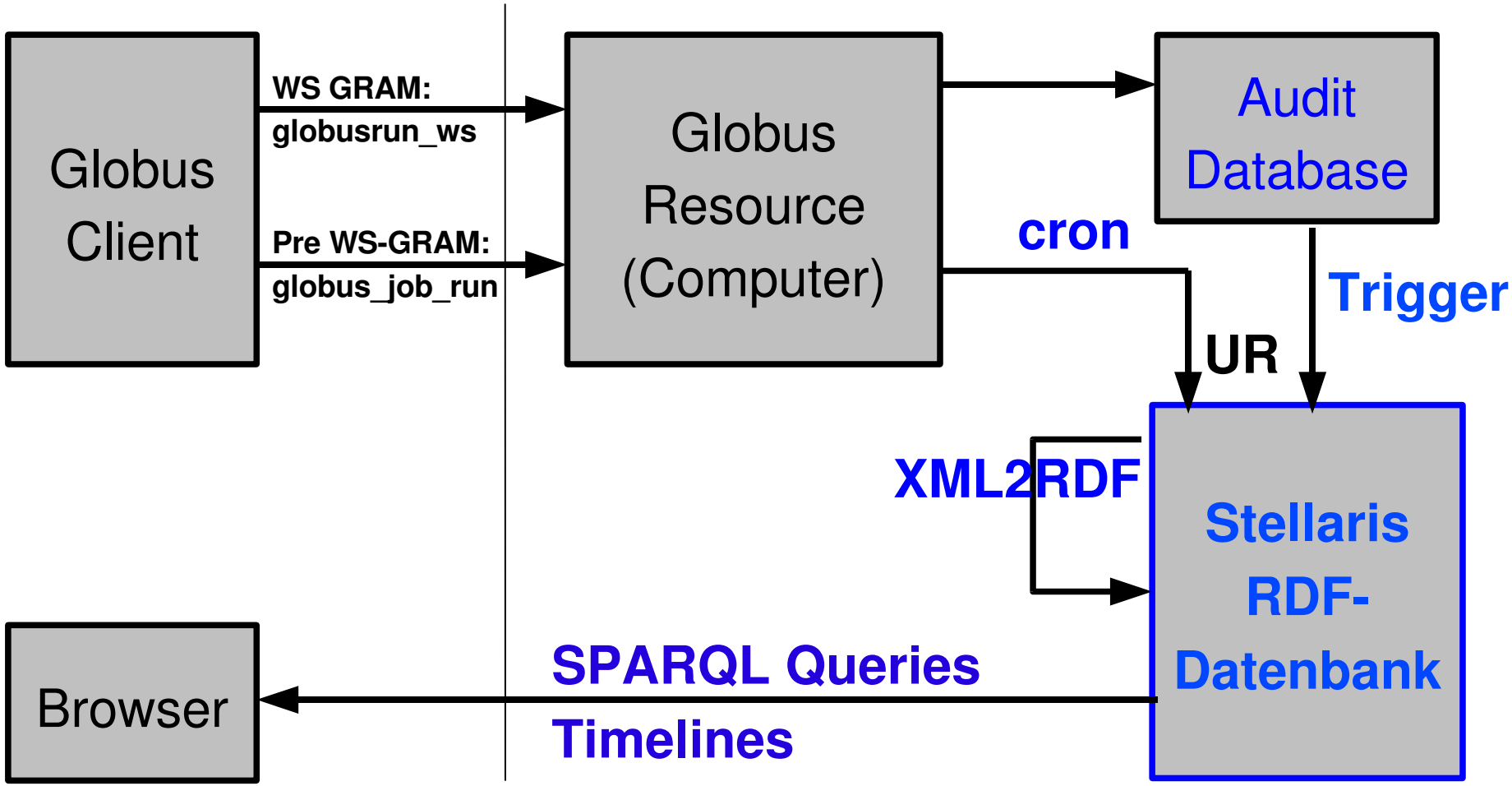
Job-Monitoring in AstroGrid-D with GRAM-audit-logging

S. Braune, F. Breitling,
H. Enke, AIP

Overview

- Audit-Logging
- AuditDatabase + Trigger
- Stellaris
- User Interface: Timeline
- RDF and SPARQL Queries
- Statistic
- Further Application: Robotic Telescopes
- Data Privacy

GRAM Audit Logging



auditDatabase Schema

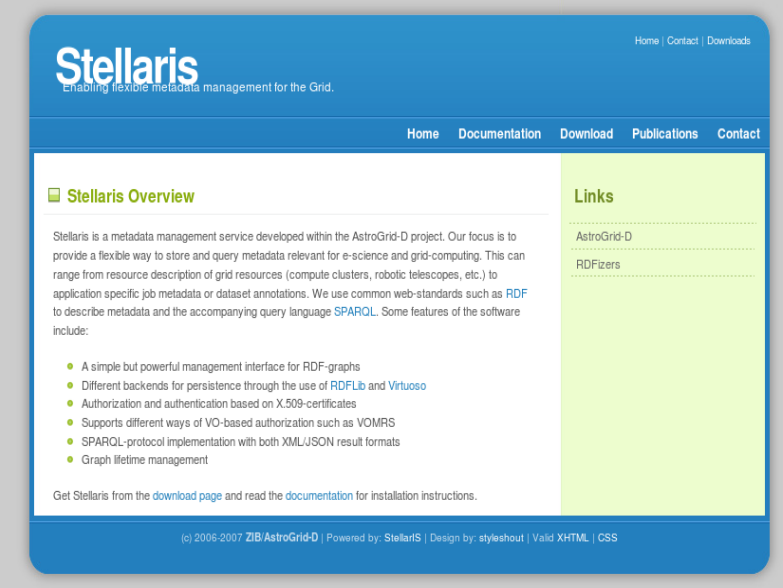
- **job_grid_id**: String representation of the resource EPR
- **local_job_id**: Job/process id generated by the scheduler
- **subject_name**: Distinguished name (DN) of the user
- **username**: Local username
- **idempotence_id**: Job id generated on the client-side
- **creation_time**: Date when the job resource is created
- **queued_time**: Date when the job is submitted to the scheduler
- **stage_in_grid_id**: String representation of the stageIn-EPR (RFT)
- **stage_out_grid_id**: String representation of the stageOut-EPR (RFT)
- **clean_up_grid_id**: String representation of the cleanUp-EPR (RFT)
- **globus_toolkit_version**: Version of the server-side GT
- **resource_manager_type**: Type of the resource manager (Fork, Condor, ...)
- **job_description**: Complete job description document
- **success_flag**: Flag that shows whether the job failed or finished successfully
- **finished_flag**: Flag that shows whether the job is already fully processed or still in progress

Requirements + Installation

- **Globus** version ≧ 4.0.5
- one **auditDatabase** (PostgreSQL or mySQL, even central) **--with-perl** (to connect Stellaris)
- Configuration files
  - WS GRAM Audit Logging:
    - $GLOBUS_LOCATION/container-log4j.properties, $GLOBUS_LOCATION/etc/gram-service/jndi-config.xml
  - Pre WS GRAM Audit Logging:
    - $GLOBUS_LOCATION/log4j.properties, $GLOBUS_LOCATION/etc/globus-job-manager.conf
- Documentation:
  - http://www.globus.org/toolkit/docs/4.0/execution/prewsgram/Pre_WS_GRAM_Audit_Logging.html
  - http://www.globus.org/toolkit/docs/4.0/execution/wsgram/WS_GRAM_Audit_Logging.html
- Documentation in AstroGrid-D (incl. Source code of the triggers):
  - http://www.gac-grid.org/project-products/grid-support/JobMonitoringAGD.html
- known Problems (alredy discussed with Globus developer team):
  - Instable connection from Globus to auditDatabase,
  - No UUID or link to job-epr-file to identify the job
  - Not enough information for real statistics (e.g. job end time, #CPUs, CPU-hours)

Stellaris as Information Service

- **Stellaris**
  - development of AstroGrid-D by ZIB (M. Högqvist, ZIB)
  - Data format: **Resource Description Framework (RDF)**
- **for Stellaris a conversion from RTML documents and UR XML into RDF required**
  - developed an XSLT (Extensible Stylesheet Language Transformation) at AIP
  - can transform arbitrary XML files into RDF/XML format (e.g. RTML)
  - usage very simple, only two command line programs „xsltproc“, and „cURL“ for uploading to Stellaris
  - available at http://www.gac-grid.org/ (Apache License)
- download: http://stellaris.zib.de/



Resource Description Framework (RDF)

- standard for storing information (part of 'semantic web')
- based on graph theory (RDF graphs), and presents information in triples (subject, predicate, object)
- different formats exist:
  - RDF/XML (W3C recommendation), and
  - Notation 3 (N3)
- (W3C specification)
- XML->RDF XSL-Transformation (e.g. for UR) was developed in AstroGrid-D
- Stellaris is a RDF database
- the **SPARQL query** language exists for retrieval of RDF information

further information: http://www.w3.org/RDF/,
AstroGrid-D examples http://stellaris.zib.de:24000/

Database Trigger

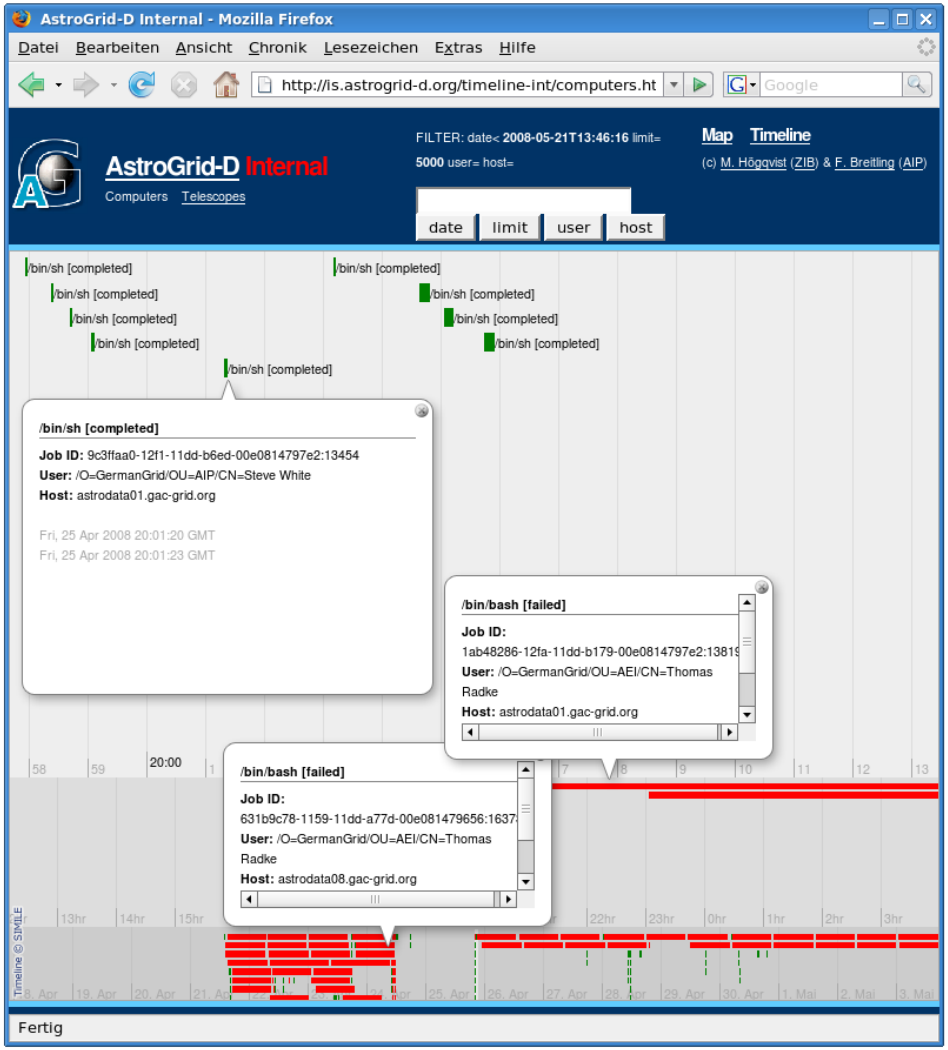
```
CREATE FUNCTION update_stellaris() RETURNS "trigger" AS $update_stellaris$
use strict; use URI; use Net::hostent; use XML::Writer; use HTTP::Request; use LWP::UserAgent;

my $job_grid_id = URI->new($_TD->{new}{job_grid_id});
my $id = unpack("H*", $job_grid_id->query()); my $host = gethost($job_grid_id->host())->name();
my $usage_record = "";
my $writer = XML::Writer->new(OUTPUT => \$usage_record, NEWLINES => 1, UNSAFE => 1);
$writer->xmlDecl("UTF-8");
$writer->startTag("JobUsageRecord", "xmlns" => "http://www.gridforum.org/2003/ur-wg#", ...);
 $writer->startTag("RecordIdentity");
  $writer->dataElement("LocalJobId", $_TD->{new}{local_job_id});   .....
 $writer->endTag("RecordIdentity");
    .....
 $writer->raw($_TD->{new}{job_description});
 $writer->dataElement("success_flag", $_TD->{new}{success_flag});
 $writer->dataElement("finished_flag", $_TD->{new}{finished_flag});
$writer->endTag("JobUsageRecord"); $writer->end();

my $req = HTTP::Request->new("PUT",
 "http://stellaris.astrogrid-d.org/files/hosts/".$host."/urs/".$id,
  HTTP::Headers->new(Content_Length => length($usage_record)), $usage_record);
my $ua = LWP::UserAgent->new(); my $res = $ua->request($req);   .....   return;
$update_stellaris$ LANGUAGE plperlu;

CREATE TRIGGER update_stellaris_trig BEFORE INSERT OR UPDATE ON gram_audit_table
 FOR EACH ROW EXECUTE PROCEDURE update_stellaris();
```

AstroGrid-D Timeline



- GUI for Stellaris
- Base on Ajcax
- user data are password protected

SPARQL Protocol and RDF Query Language (SPARQL)

SPARQL

- **is the counterpart for retrieving information from RDF** (part of 'semantic web')
- simple syntax for selected and sorted retrieval
- Perl modules exist for the integration into Perl scripts
- SPARQL web browser interface is implemented into Stellaris for a quick lookup
- Example:

```
PREFIX ur: <http://www.gridforum.org/2003/ur-wg#>
PREFIX x2r: <http://www.astrogrid-d.org/2007/08/14-xml2rdf#>

SELECT distinct(?username)
WHERE {
n1 ur:UserIdentity ?user_id .
?user_id ur:GlobalUserName ?username . }
```

gives a list of all users.

SPARQL example

Query form

SPARQL help | Collections | Query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ur: <http://www.gridforum.org/2003/ur-wg#>
PREFIX globus_job:
<http://www.globus.org/namespaces/2004/10/gram/job/>

SELECT distinct ?GlobalUserId ?executable ?SubmitHost count(*)
WHERE { graph ?g {
 ?n1 ur:RecordIdentity ?RecordIdentity .
 ?RecordIdentity ur:job_grid_id ?job_grid_id .
 ?n1 ur:UserIdentity ?UserIdentity .
 ?UserIdentity ur:GlobalUserId ?GlobalUserId .
 ?n1 ur:SubmitHost ?SubmitHost .
 ?n1 globus_job:descriptionserviceLevelAgreement ?sla .
 ?sla globus_job:descriptionjob ?job .
 OPTIONAL {
   ?job globus_job:descriptionexecutable ?executable .
   FILTER (!REGEX(?executable,"^http:"))
 }}} ORDER BY ?GlobalUserId Limit 5
```

Query Status

Indices: **query** | virtuoso

Result links:

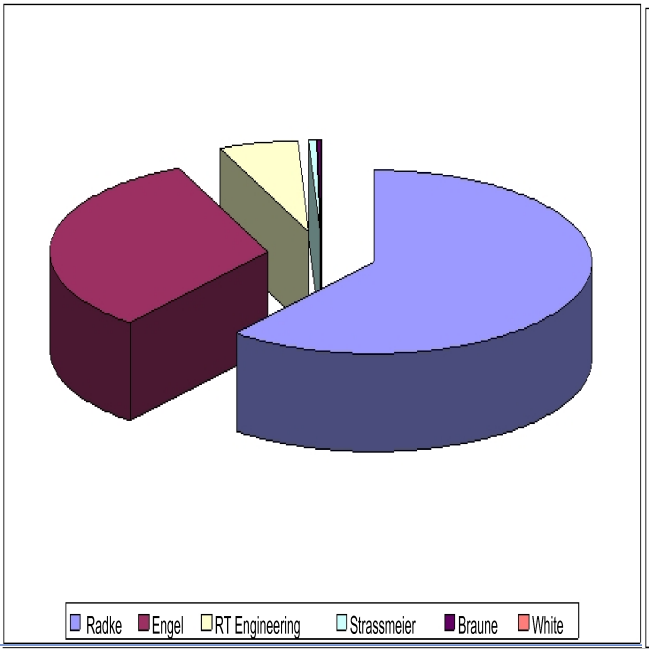
Error:

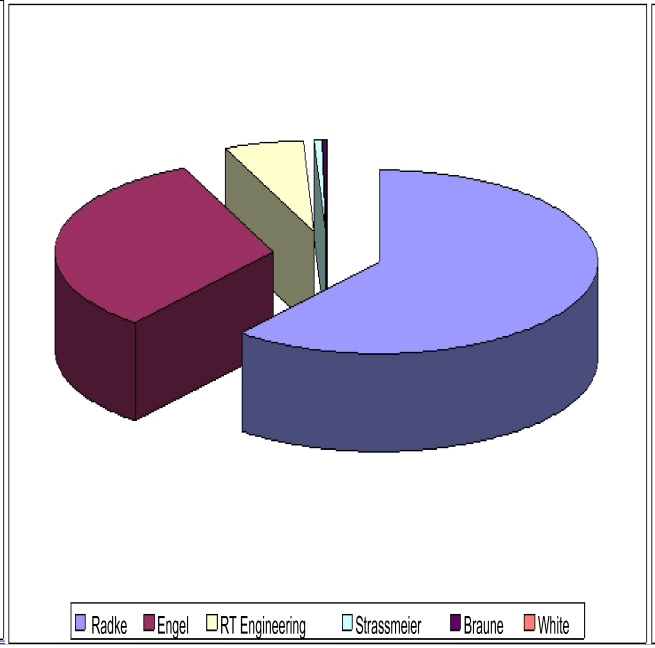
Submit query

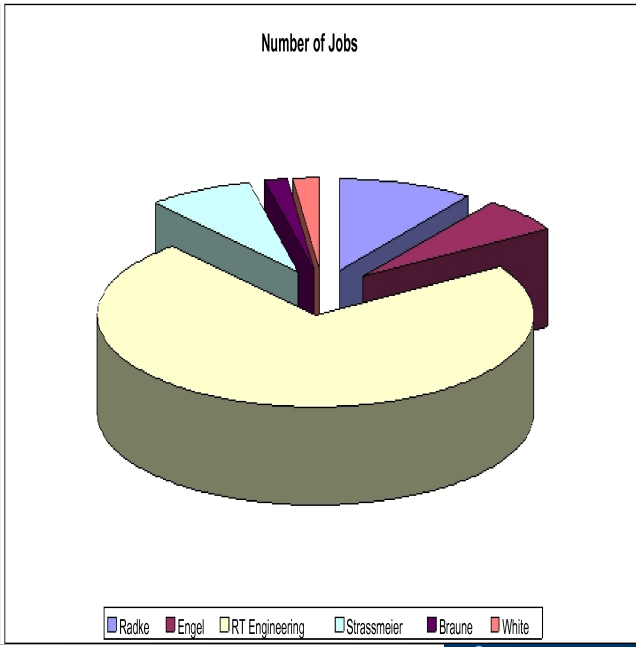
callret-0	callret-1	callret-2	callret-3
/C=DE/O=GridGermany/OU=Leibniz-Rechenzentrum/OU=TUM Informatik (I3)/CN=Benjamin Gufler	/bin/date	astrodata01.gac-grid.org	7
/C=DE/O=GridGermany/OU=Leibniz-Rechenzentrum/OU=TUM Informatik (I3)/CN=Benjamin Gufler		astrodata01.gac-grid.org	1
/O=GermanGrid/OU=AEI/CN=Oliver Wehrens	clusterfinder.sh	gavo2.aip.de	2
/O=GermanGrid/OU=AEI/CN=Robert Engel		astrodata03.gac-grid.org	2
/O=GermanGrid/OU=AEI/CN=Robert Engel		astrodata07.gac-grid.org	4

Statistic

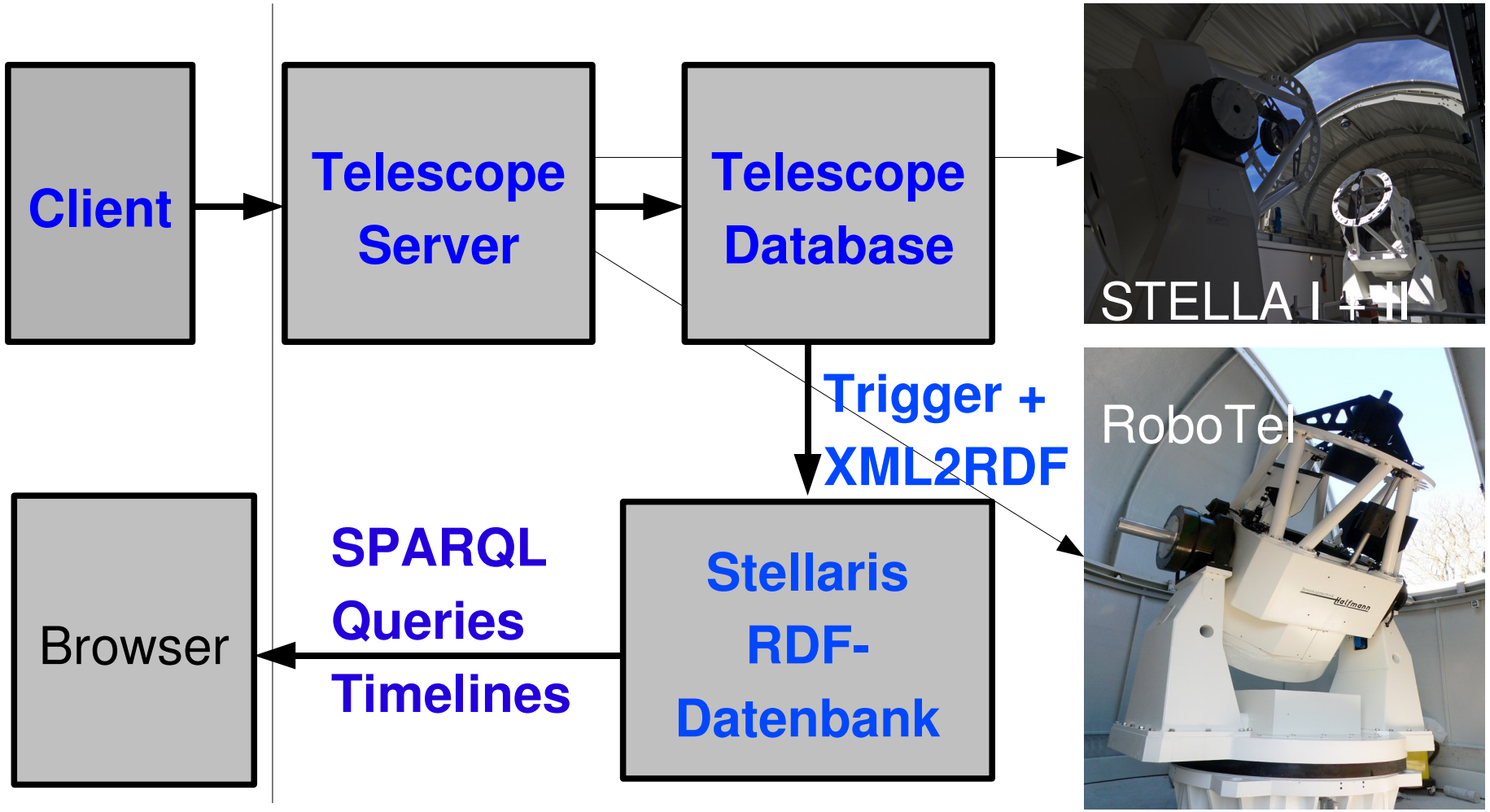
- Statistics about Users, Jobs (binary), and machine time (e.g. Monthly) can be done
- query language: SPARQL
- Examples from AstroGrid-D:



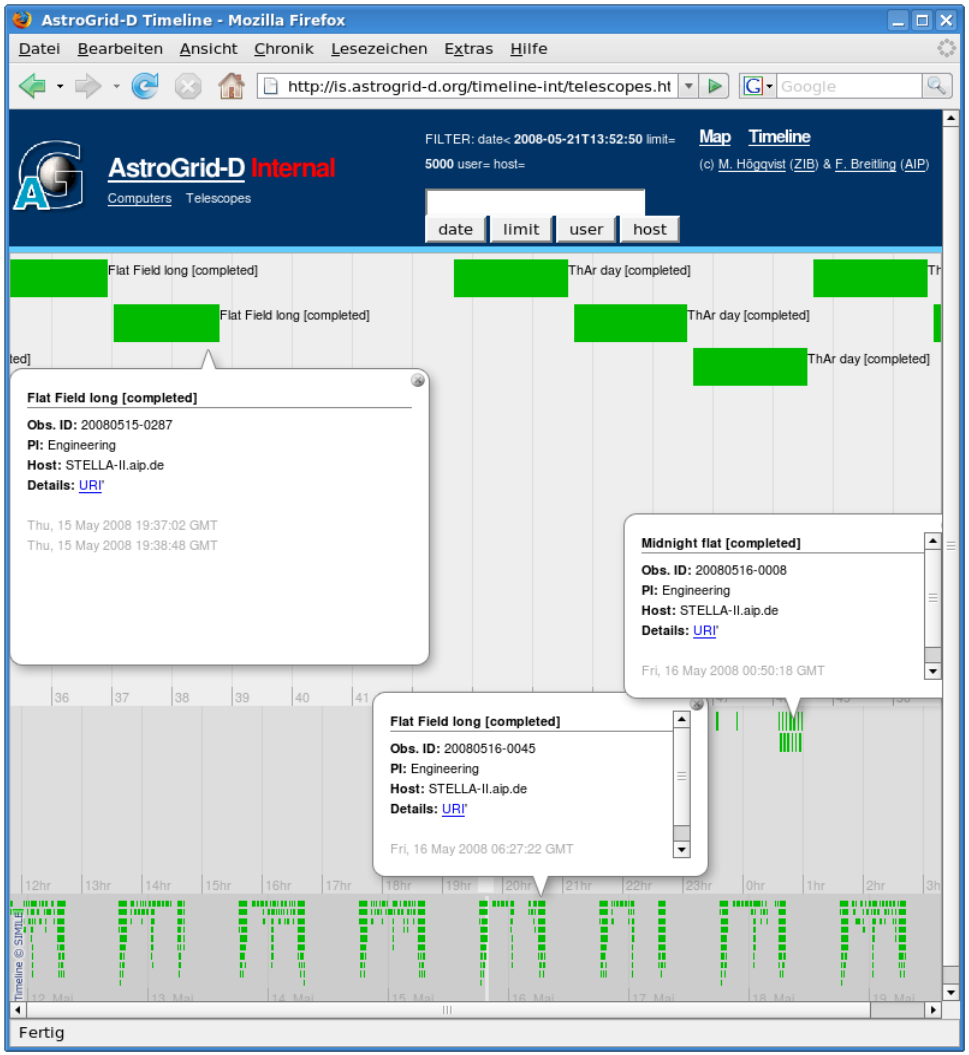


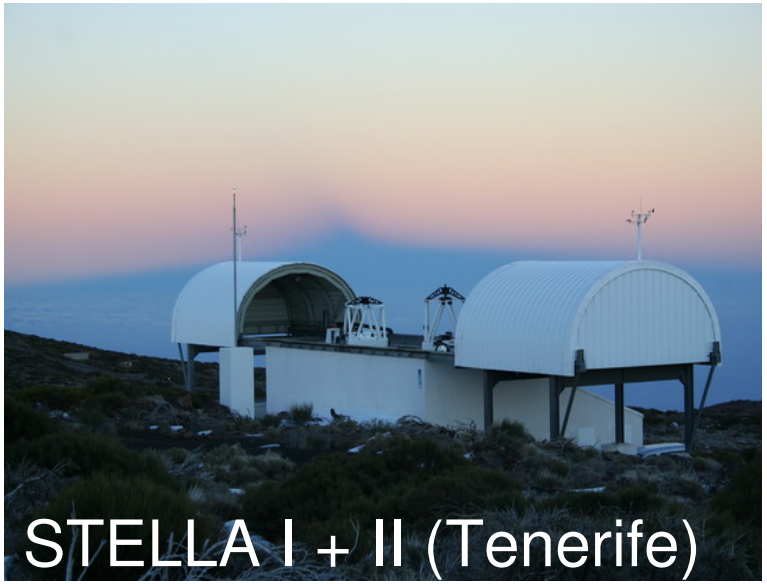


Telescope Monitoring



AstroGrid-D Telescope Timeline





- GUI for Stellaris
- Base on Ajcax
- user data are password protected

Policy for Usage Record Collection

The AstroGrid-D Steering Board accepts the policy below as part of the AUP of AstroGrid-D. (Telecon, 08.05.2008)

Policy for using Monitoring Data of Grid-Jobs

1.) AstroGrid-D collects data delivered by the Globus Audit-Module into usage records on a per job basis.

2.) The records are stored in Stellaris. Each AstroGrid-D resource runs the Globus-Audit-module and the trigger-software provided by the working group, starting in May 2008.

3.) The software collects only data for the members of VO AstroGrid.

4.) The collected data is displayed without exposing names or accounts in the GridTimeline.

5.) A protected interface allows additional queries for VO members and administrators.

Description of the Data collection:

With the Globus Audit module the following items of a grid-job are logged into a database:

Times: Start, Submit, End

Flags: Success, Failure, Running/Stopped

Job-Description: Executable and RSL-file, Job-Manager (Fork, PBS, LSF, SGE, CONDOR)

IDs : DN and local account name, execution host

Local and Global Job-ID

download: **http://www.gac-grid.org/project-overview/VirtualOrganisations/AUP/Astrogrid-Policy-UR-Collection.pdf**

Summary

- Partial installation in AstroGrid-D
- Tests have been successful and are ongoing
- Information about job execution on each grid host
- Missing information: Feature Request (Bug Tracker) Telecon Globus developers