

Grid Runtime Environment Solutions in KnowARC

Daniel Bayer Steffen Möller

University of Lübeck, Institute for Neuro- and Bioinformatics

EGEE'07
Budapest, October 2007



KnowARC

Outline

Static Runtime Environments

- Introduction

- The Runtime Environment Registry

- Summary

Dynamic Runtime Environments

- Introduction

- The Janitor

- Deployment

- The Catalog

Outlook



What are Runtime Environments?

- ▶ Grid jobs depend on certain programs or data being available on the worker node.
- ▶ These are called Runtime Environment (RE).
- ▶ REs are
 - ▶ specified in the Runtime Environment Registry.
 - ▶ made available on the worker nodes by administrators.
 - ▶ announced in the Grid Infosystem.
 - ▶ used for brokering.
 - ▶ enabled at the worker node on a per job basis.
 - ▶ used by the job.



The Runtime Environment Registry

- Website at <http://gridrer.csc.fi>

Runtime Environment Registry

This registry contains a list of Grid Runtime Environments (REs) and serves the members of [NorduGrid](#), [Nordic DataGrid Facility](#), [SweGrid](#) and [M-grid](#) and other Grid projects using [NorduGrid ARC middleware](#).

For more information, see:

- [Introduction to Runtime Environments and Registry](#).
- [How to use RE \(users\)](#)
- [How to install RE \(site admins\)](#)
- [How to set up and maintain RE \(RE maintainers\)](#)

The [concept](#) of the RER itself is under development. Please contribute through the nordugrid-discuss@nordugrid.org mailing list or mail directly to [RER maintainer](#).

List of Runtime Environments

The first entry for each RE is the reserved name of the RE, and the version number of the latest stable (preferred, current) release. Other available versions are documented on the RE's Homepage.

APPS/BIO/JASPAR-CORE-1.0	
Description:	JASPAR-CORE
RE Homepage:	http://www.grid.tsl.uu.se/RTES/JASPAR-CORE/
Status:	Available
Last update:	2006-09-02
APPS/BIO/LAGAN-1.2	



The Runtime Environment Registry

List of Runtime Environments

The first entry for each RE is the reserved name of the RE, and the version number release. Other available versions are documented on the RE's Homepage.

APPS/BIO/JASPAR-CORE-1.0	
Description:	JASPAR-CORE
RE Homepage:	http://www.grid.tsl.uu.se/RTEs/JASPAR-CORE/
Status:	Available
Last update:	2006-09-02
APPS/BIO/LAGAN-1.2	
Description:	LAGAN
RE Homepage:	http://www.grid.tsl.uu.se/RTEs/LAGAN/
Status:	Available
Last update:	2006-09-21
APPS/BIO/RTEs A E A	



The Runtime Environment Homepage

Name:	APPS/BIO/JASPAR-CORE-1.0
Description:	JASPAR-CORE
Status:	Available
Last update:	2006-09-02

JASPAR-CORE Runtime Environment Homepage

JASPAR CORE is a bioinformatics database.

This page describes the JASPAR-CORE runtime environment for grid systems based on the Nordugrid ARC middleware.

Interface definition

On the remote system JASPAR-CORE RE initializes the following environment variables:

JASPAR_CORE Full path to the directory of the JASPAR-CORE database files

Versions

Old versions

System administrator guide

Installation of the RE

A [tarball](#) containing the JASPAR-CORE database files.

Verification of the RE

A test for verifying the runtime environment are given below:

Jaspar Core Test The expected outcome of the test is included in the tarball.

Contact information



Summary

- ▶ traditional static REs are well suited to provide a few, stable and well-known runtime environments
- ▶ but they are not suited for a huge amount of fast changing REs as necessary for e. g. bioinformatics

⇒ Dynamic Runtime Environments



What are Dynamic Runtime Environments

- ▶ Dynamic Runtime Environments are REs prepared to be installed **automatically** and **dynamically**.
 - automated** once triggered, the installation is done without human supervision
 - dynamic** installation can be triggered by requirements of incoming jobs
- ▶ Are implemented with help of a new Service, the **Janitor**



The Janitor

- ▶ is informed by the Grid Manager about requirements of jobs and its state
 - ⇒ triggering dynamic installation
- ▶ knows which REs are available and passes this information to the Grid Information System
- ▶ removes dynamically installed DREs which are unused for a long time



Interactions of the Janitor with the middleware

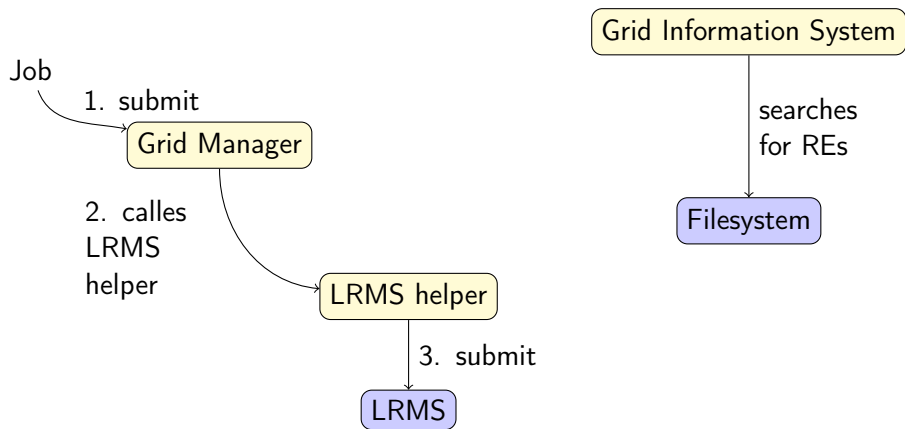


Figure: Interaction of the Janitor with the ARC 0.6 middleware



Interactions of the Janitor with the middleware

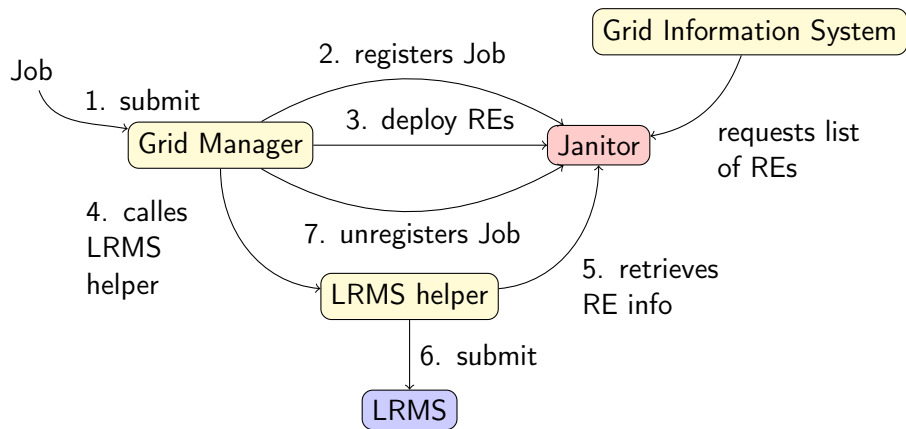


Figure: Interaction of the Janitor with the ARC 0.6 middleware



Deployment

- ▶ currently only one supported package format, TAR packages, containing:
 - ▶ the software to deploy and
 - ▶ metadata to control the deployment and usage of the software
- ▶ other kind of packages are possible, but currently not implemented
- ▶ constraints set by the site admin specify the eligibility of packages
- ▶ a shared directory is necessary



The Catalog

- ▶ describes all REs:
 - ▶ name
 - ▶ version (part of the name)
 - ▶ RE homepage
 - ▶ last change
 - ▶ set of tags
- ▶ for DRE also describes how to deploy
- ▶ currently the Catalog is a RDF file, it will be a service in the future



The RDF Schema

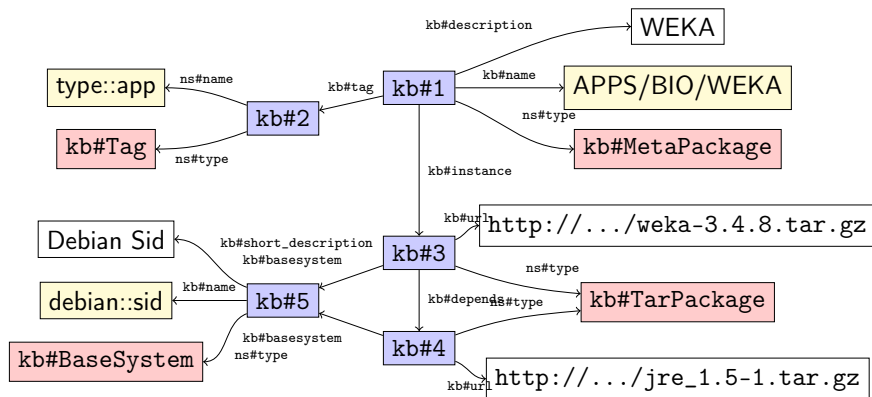


Figure: A part of the RDF graph describing the WEKA runtime environment.



Acceptance in the Community

- ▶ Current Installations:
 - ▶ Lübeck
 - ▶ Košice
 - ▶ Kopenhagen
- ▶ Upcoming Installations:
 - ▶ Helsinki and
 - ▶ the NDGF Biogrid Community



Ongoing Development

- ▶ use of virtualisation, which will
 - ▶ provide for every job a clean environment,
 - ▶ allow to easily draw software packages from Linux distributions
- ▶ the RER will be implemented with help of the Catalog
- ▶ the Catalog will become a service
- ▶ expose DRE states in the Information System

