

# Rapid Prototyping of Science Gateways in the Brazilian National HPC Network

Bruno Fernandes Bastos (bfbastos@lncc.br)  
Brazilian National System for HPC - SINAPAD

Vinícius de Macedo Moreira (vmacedo@lncc.br)  
Brazilian National Laboratory for Scientific Computing - LNCC

Antônio Tadeu Azevedo Gomes (aatagomes@lncc.br)  
LNCC / SINAPAD

# Agenda

- Motivation
- Overview of the CSGrid Middleware
- PortEngin Tool
  - File Sharing
  - Data Provenance Tracking
  - Restricted Anonymous Access
- PortEditor Tool
- Command Line Interface and the CORE API
- Conclusions

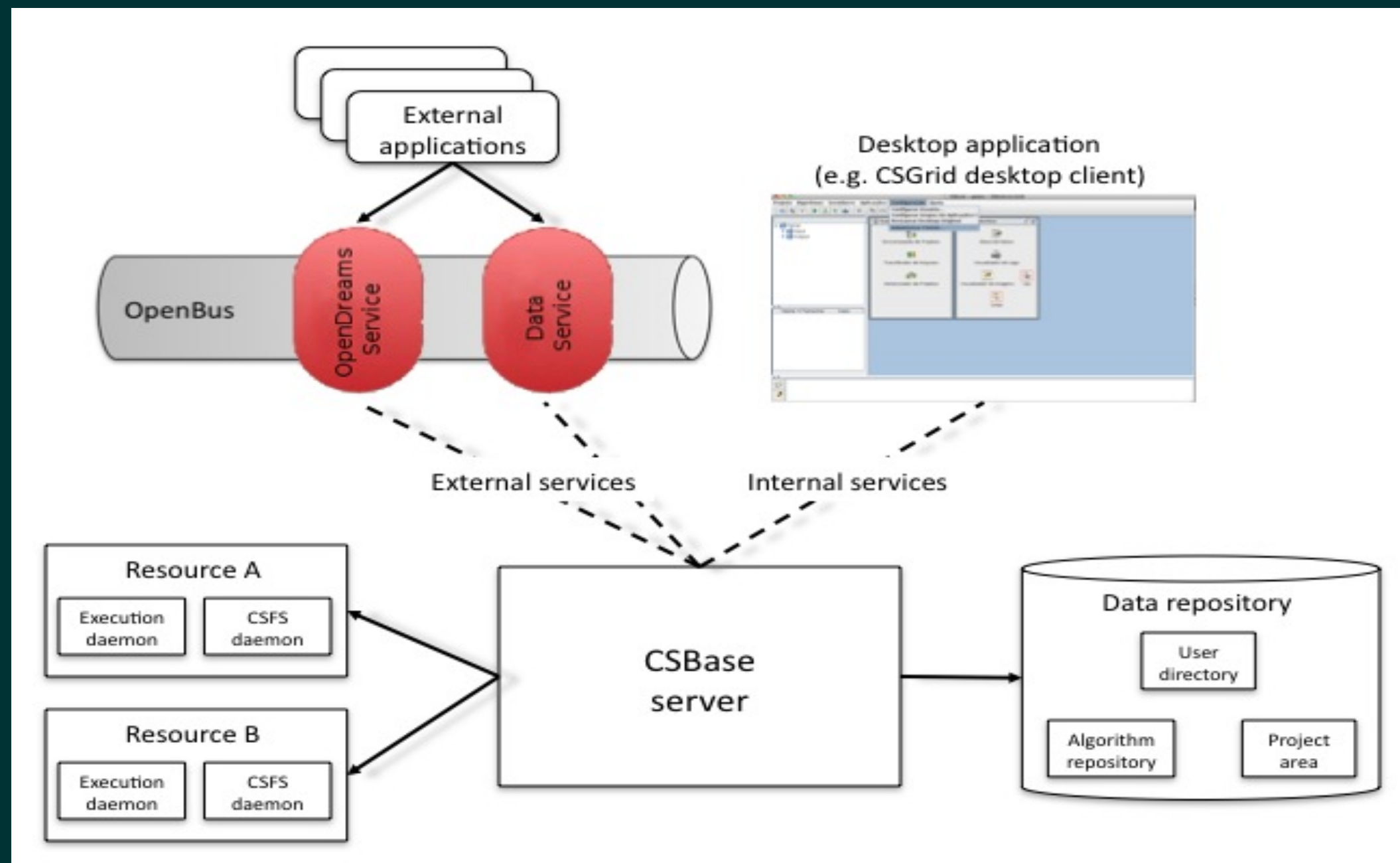
# Motivation

- SINAPAD Resources are geographically distributed in some institutions.
- Usually accessed through SSH.
- Job submission for any of the HPC resources of the SINAPAD network though a single access point (Web site).
- Centralized data provenance.
- Online and off-line job monitoring.

# Overview of the CSGrid Middleware

- Allows users to save data into a private project area.
- Sends e-mail containing the status of job completion.
- All resources can be accessed from a single point.
  - Automatically stages data into and out of the HPC resources.
- Developed as a Desktop application.
- Any (external) application that uses CSGrid must be previously certified in a service bus (OpenBus).

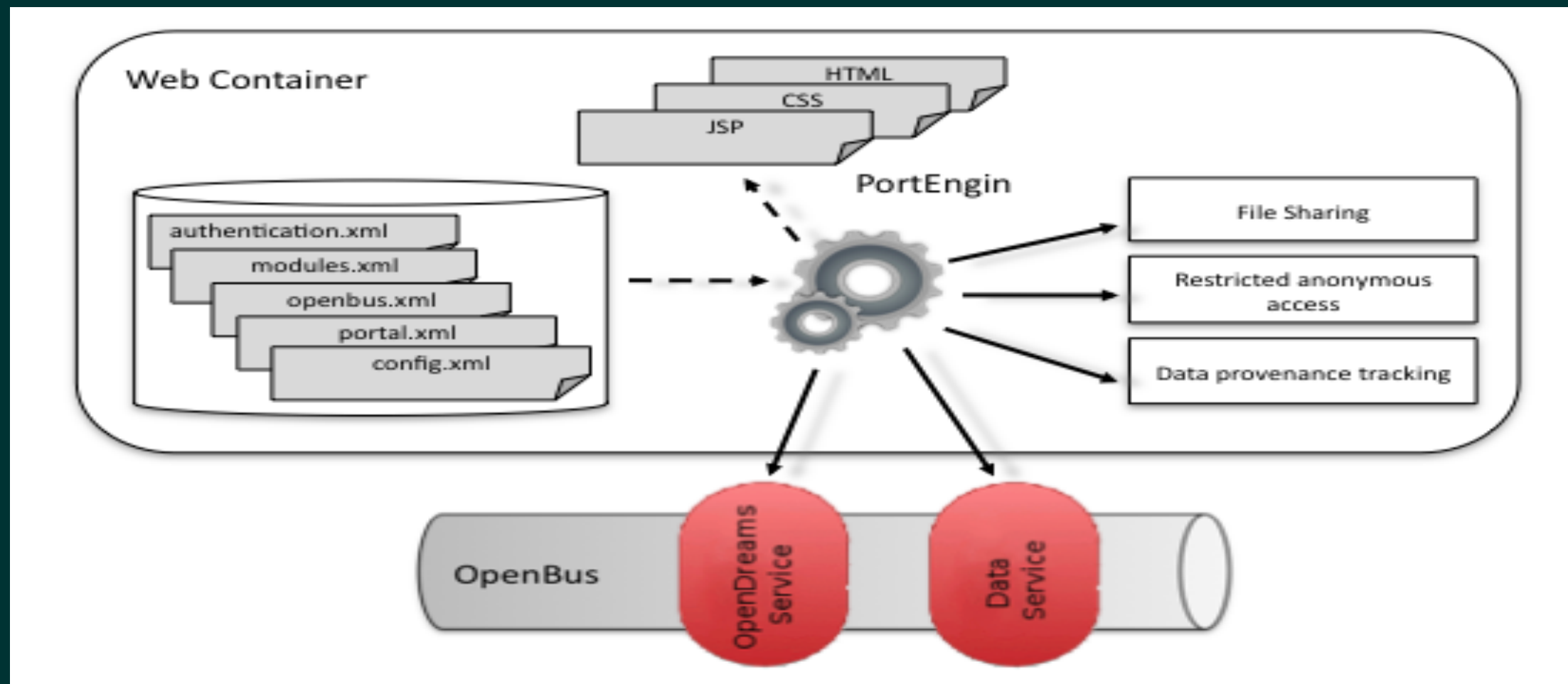
# CSGrid Middleware Architecture



## PortEngine Tool

- Gateway creation by configuring a few XML files.
- Easy to enable / disable features.
- Changes to the default layout can be done solely in the presentation layer (JSP, CSS, HTML).
- New features and bug corrections are done in a single place and can be offered as an update package.
- A CORE API is offered for advanced programming.

# PortEngin Architecture



# PortEngin Configuration

- LDAP authentication service.
- Access to CSGrid components.
- Modules:
  - File sharing.
  - Restrict anonymous access.
  - Data provenance tracking.



## File Sharing

- Allows a user to share a file between every registered gateway user.
- Users can also publish a file or directory through a (tiny) URL.
- Shared files cannot be edited or directly used in a job submission. If the user wants to use the file for such purpose, he must copy it to his private project area.

## Restricted Anonymous Access

- Allows a guest user to access the gateway without username and password.
- Requires the user to provide a valid e-mail address and pass through a captcha validation before submitting a job.
- May constrain job parameters, such as the size of an input file on an specific gateway;
- Doesn't offer the user a private project area nor file sharing/publishing capabilities.

# Data Provenance Tracking


- Stores input and output data of a given job submission.
- Keeps such data immutable, no matter if the user modifies their versions in his project area.
- Has all data used for a job, so it's easy to reproduce a previously executed experiment.
- Also keeps software provenance (version of the executable used for the submission).




# PortEditor Tool


- Simplifies configuration of XML files through a web interface.
- Gateway developer does not need to know the syntax of XML files.
- Automatically registers all digital certificates needed for PortEngin in the OpenBus service bus to access CSGrid services.
- After configuration of a new gateway, manual intervention is still needed for specifying which user can access such gateway.

# PortEditor Web Interface

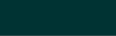


**LNCC**  
 Ministério da Ciência e Tecnologia







**SINAPAD**

**SISTEMA NACIONAL DE PROCESSAMENTO DE ALTO DESEMPENHO**  


---

Home

New Portal

User: Bruno Bastos

Logout

---

## Portal Configurator and Editor



Welcome to the Portal Configurator Bruno Bastos. Here you can configure, edit and download your already made portals.

You have 3 Portals.

- CAM.war [Download](#)
- DANCE.war [Download](#)
- Profrager.war [Download](#)



Edit


New

**HISTÓRICO** - ORGANIZAÇÃO - FALE CONOSCO - MAPA - CRÉDITOS
 

:: RESTRITO




Ministério da Ciência e Tecnologia

SINAPAD

SISTEMA NACIONAL DE PROCESSAMENTO DE ALTO DESEMPENHO

pad.brasil



[Home](#)
[New Portal](#)

User: Bruno Bastos
[Logout](#)

## Portal Configurator and Editor

### Portal Configuration

Portal Full Name	Portal
Portal Acronym Name	DANCE
Portal Description	<div> Closeness Centrality is a traditional and important network centrality measure which ranks the nodes by how close each node is to all other nodes in the network. This ranking shows the nodes that are best positioned for efficient diffusion processes, such as information or goods distribution, disease or rumor spreading, and so </div>
Algorithm Name	Dance
Algorithm Default Version	1.0.0
Allow Multiple Versions	<input type="checkbox"/>
Automatically Generate Interface	<input checked="" type="checkbox"/>

### LDAP Authentication Configuration

Enable SSL	<input type="checkbox"/>
Port	389



### Modules Configuration

Shared Files	<input type="checkbox"/>
Public Files	<input type="checkbox"/>
Algorithm History LifeTime in Days ( -1 equals undefined time. )	-1
Guest Access	<input type="checkbox"/>

### Error Report E-Mail Configuration

SMTP Host	smtp.sinapad.lnoc.br
Email From	sinapad-sup@lnoc.br
Email to	sinapad-sup@lnoc.br

[Create Portal](#)

HISTÓRICO

ORGANIZAÇÃO

FALE CONOSCO

MAPA

CRÉDITOS

RESTRIÇÃO

# Command Line Interface

- Allows job submission and monitoring through a (ssh) terminal.
- Can be used to automate sets of submissions through scripting.



# Command Line Interface Example

```
bfbastos@sinapadcli:~$ run --project DANCE --help
job parameters (for version 1.1.0):
-FILE=INPUT FILE
Input File
-RADIUS=INTEGER
Radius
-CLASSIFIER=SELECTION
Classifier (1 - DACCER, 2 - EGO-BETWEENS, 3 - LBC, 4 - WEIGHED VOLUME)
-OUTPUT=OUTPUT DIRECTORY
Work directory
```

```
bfbastos@sinapadcli:~$ run --project DANCE --email bfbastos@lncc.br -FILE=CAIDA.txt -RADIUS=2 -CLASSIFIER=1 -OUTPUT=output
Your job bfba bfba SBH3QX27E5 has been submitted
bfbastos@sinapadcli:~$
```

```
bfbastos@sinapadcli:~$ stats --project DANCE --status DONE
```

JOB-ID	QUEUE	SUBMISSION TIME	STATUS
bfba@bfba.SBH22Q00JK	ufc-padufc0_longsinapad	20130517214925	DONE


```
bfbastos@sinapadcli:~$
```



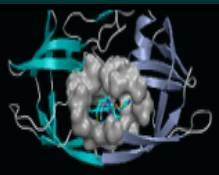
## Core API

- Offers primitives for user authentication, data management and job submission.
- Developed in Java but also offered as a web service.
- Used by PortEngin and CLI (and also by PortEditor for collecting information about the executables made available through the CSGrid middleware)
- Developers can create their own gateways with highly customized interfaces while still using most of PortEngin functionalities.

# Core API Examples



A RECEPTOR-LIGAND DOCKING PROGRAM



Home
Docking
References
About
Support

Protein
Ligand
Cofactors
Docking
Results and Analyzes

1) Submit your protein file (.pdb or .in):  
Protein file uploaded: 1XM6.pdb

Upload

2) Prepare the protein file: [\(View 3D\)](#)  

Reprepare

Attention: it is only possible to change the protonation states and visualize the protein structure for .pdb files. All the hydrogen atoms in the original files will be removed. To consider cofactor molecules during docking, treat them separately in the Cofactor tab.

3) (Optional) Change the protonation state of the amino acids. After, click on the Reprepare button.  


Residue protonation options

Chain A









Chain B


4) Send the protein to docking:  

NEXT



Download prepared files



SISTEMA NACIONAL DE PROCESSAMENTO DE ALTO DESEMPENHO

pad.brasil

Welcome Stanley Lima!

Add Page

Create a Job

Dashboard

BRAMS Input Data

Dashboard

Local ID	Remote ID	City	Creation Date	Submission Date	Job Status	Actions
227	stan@stan.SBHVTJFJ64	Sao Paulo	2013-03-14 12:54	2013-03-14 01:10	READY	
226	stan@stan.SBHVSFFF4	Default Ramsin	2013-03-13 02:44	2013-03-13 02:58	READY	
225	stan@stan.SBHVP42AVT	Default Ramsin	2013-03-12 11:32	2013-03-12 12:01	FAILED	
224	stan@stan.SBHVNQNMNJ	Default Ramsin	2013-03-11 05:01	2013-03-11 07:53	DONE	
218	stan@stan.SBHVCMSWYGP	Default Ramsin	2013-03-07 04:58	2013-03-07 05:14	DONE	
217	stan@stan.SBHVCMSRCS	Default Ramsin	2013-03-07 11:59	2013-03-07 12:14	FAILED	
216	stan@stan.SBHVAVG4ZF	Default Ramsin	2013-03-06 07:51	2013-03-06 08:07	RUNNING	
214		Default Ramsin	2013-03-05 05:37		CREATED	
213	stan@stan.SBHUSM22BS	Default Ramsin	2013-03-05 01:26	2013-03-05 01:43	READY	
210	stan@stan.SBHUBBJD6	Default Ramsin	2013-02-21 12:50	2013-02-22 01:18	READY	

137 jobs found, displaying 10 job(s), from 1 to 10. Page 1 / 14.

Visualization

Map

Output Directories

View Output Directories

227

227

A

H

INITIAL

MAKEFILE

Visualization

Map

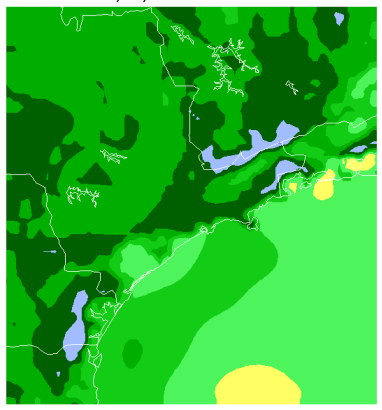
Output Directories

Information about Job with PID [ stan@stan.SBHVS4WQ56 ]

RAMSIN for city [ Sao Paulo ]

RELATIVE HUMIDITY [pct]

Date: 2010/04/21 Hour: 01:00 Level: 57m



Variables

Relative Humidity

Levels [m]

57

Initial Date

2010-04-21 00:00:00.0

Forecast:

Date:

2010/04/21

Time:

01

IWSG 2013 ã Zurich, CH

http://www.lncc.br/sinapad

## Conclusions

- Once the executable is installed in the HPC resources of the SINAPAD network, it's possible to configure a new gateway in a few hours with the PortEditor and PortEngin tools.
- The anonymous access offers a good way for publicizing an e-Science tool.
- Both input and output files can be made public through a (tiny) URL, which may be used e.g. in a paper.
- Data provenance helps users to reproduce and validate experiments.

# Thank you!

sinapad-sup@lncc.br