



Dziedzinowo zorientowane  
usługi i zasoby infrastruktury  
PL-Grid dla wspomagania  
Polskiej Nauki w Europejskiej  
Przestrzeni Badawczej

# DataNet – GridSpace Data Management Framework

Daniel Haręźlak<sup>1</sup>, Eryk Ciepiela<sup>1</sup>, Marek Kasztelnik<sup>1</sup>,  
Bartosz Wilk<sup>1</sup>, Marian Bubak<sup>1,2</sup>

<sup>1</sup>ACC Cyfronet AGH,

<sup>2</sup>AGH University of Science and Technology, Institute of Computer  
Science AGH

Cracow Grid Workshop '12, October 22-24, 2012, Krakow, Poland



- Introduction to GridSpace
- Motivation behind DataNet
- Metadata Management Requirements
- Architecture Description
- Future Work

- GridSpace - a virtual laboratory framework
  - Enables researchers to conduct virtual experiments on Grid-based resources and other HPC infrastructures
  - Facilitates exploratory development of experiments
  - Provides a convenient web-based interface for collaborative research
- GridSpace – a publication platform
  - GridSpace experiments can be published in private or public mode
  - Individual elements of a GridSpace experiment, such as code snippets or data items, are embeddable on any web page
  - Published content can be executed and verified by readers

GridSpace  
@GridSpace

- **Experiment** – an organized collection of resources such as scripts, libraries, input and output data items
- Data items are processed on HPC infrastructures straight from the web interface (cross-site execution is supported)
- Both local and Grid access are supported




# GridSpace Introduction – Deployments



- CYFRONET Deployment – <https://gs2.cyfronet.pl>
  - Cutting edge experimental release
  - CYFRONET HPC resources available
  - Publication platform based on WordPress – <http://gs2.cyfronet.pl/epapers>
- PL-Grid Production Deployment – <https://gs2.plgrid.pl>
  - HPC resources of Poland's top five computing centers available
- Executable Papers for 3D Object Retrieval Deployment - <https://collage.elsevier.com>
  - Dedicated computing node utilized

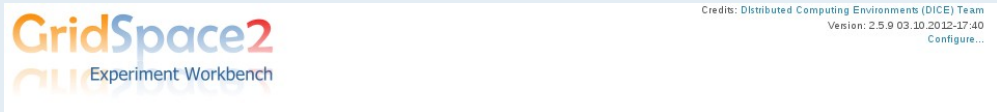
Call for Proposals:  
Special Issue on  
Executable Papers  
for 3D Object  
Retrieval

[Click for details!](#)

powered by   
**Collage**  
Authoring Environment

The banner features a vertical gradient background transitioning from dark blue at the top to light blue at the bottom. It contains text about a call for proposals for a special issue on executable papers for 3D object retrieval. A button labeled 'Click for details!' is positioned below the text. At the bottom, it mentions 'powered by Collage Authoring Environment' with a small logo.

## Let's have a look at GridSpace web interfaces...

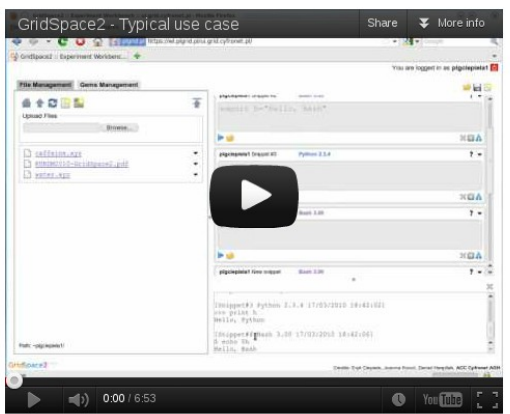


GridSpace is a novel virtual laboratory framework enabling researchers to conduct virtual experiments on Grid-based resources and other HPC infrastructures. GridSpace facilitates exploratory development of experiments by means of scripts which can be expressed in a number of popular languages, including Ruby, Python and Perl. The framework supplies a repository of gems enabling scripts to interface low-level resources such as PBS queues, EGEE computing elements, LFC directories and other types of Grid resources. Moreover, GridSpace provides a Web 2.0-based Experiment Workbench supporting joint development and execution of virtual experiments by groups of collaborating scientists.

GridSpace is being developed in support of research teams linked to the PL-Grid project. Nonetheless, it's targeted to unrestricted range of communities. First official release took place in August 2010. From then on, new versions are being released each couple of months.

### This is GridSpace installation for DICE Team

This is a development installation of GridSpace



```
require 'json'
baseUri = "#{ENV['GS2_WEBGUI_ENDPOINT']}"
if(baseUri.start_with?('https'))
  require 'net/https'
else
  require 'net/http'
end

startUri = URI.parse("#{baseUri}/start")
http = Net::HTTP.new(startUri.host, startUri.port)
if(baseUri.start_with?('https'))
  http.verify_mode = 0
end
if(baseUri.start_with?('https'))
  http.use_ssl = true
end
startRequest = Net::HTTP::Post.new(startUri.path)

No inputs defined for this snippet. 1 output defined for this snippet.

Generacja losowych liczb Ruby 1.8.7 with zeus.cyfronetpl
Sortowanie Python 2.4.3 with zeus.cyfronetpl
Generacja klatek Bash 3.00 with zeus.cyfronetpl
Generacja filmu Bash 3.00 with zeus.cyfronetpl

Output

response.body
}
}
sleep 1
end

result = JSON.parse(checkData)
size = result['data']['size']
File.open('size.txt', 'w') {|f| f.write(size) }
```

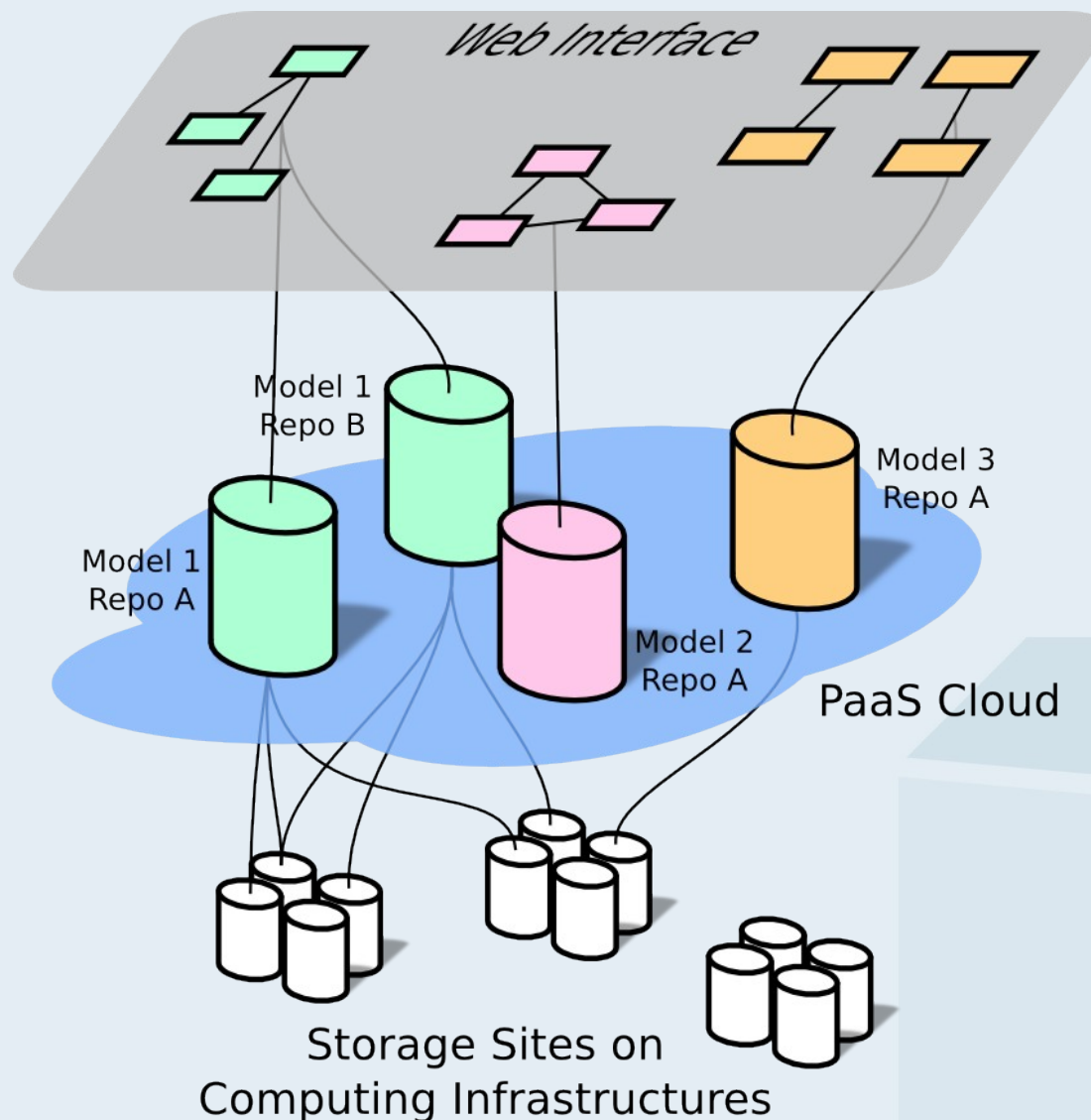
## ■ Rationale

- GridSpace experiments formalized the description of conducting virtual research
- Although experiments can be annotated and provenance is recorded extra meta-information has to be saved and shared
- Each experiment is different and requires dedicated metadata model

## ■ Objectives

- Provide means for **ad-hoc metadata model creation** and deployment of corresponding storage facilities
- Create a research space for **metadata model exchange and discovery** with associated data repositories with access restrictions in place
- Support **different types of storage sites** and **data transfer protocols**
- Continue to support the exploratory paradigm by making the models evolve together with data

- **Web Interface** is used by users to create, extend and discover metadata models
- Model repositories are deployed in the **PaaS Cloud** layer for scalable and reliable access from computing nodes through REST interfaces
- Data items from **Storage Sites** are linked from the model repositories





## ■ DONEs

- Custom CloudFoundry environment was setup as a PaaS platform to ensure quick deployments of required application and storage services
- Preliminary schema for metadata model creation was elaborated and is being evaluated for NoSQL storage service MongoDB
- Prototypes of storage site access libraries were implemented and tested

## ■ TODOs

- Build and deploy a web-based tool to create, discover and manage metadata models
- Integrate storage site access libraries in a web application for convenient data access
- Support various types of metadata storage services to fulfill different application requirements

## Acknowledgements

- This research has been partially supported by the European Regional Development Fund program no. POIG.02.03.00-00-096/10 as part of the PL-Grid PLUS project

## Have a look at our poster and help make DataNet better

## Visit <http://dice.cyfronet.pl> for more information

