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

DataNet – GridSpace Data Management Framework

Daniel Harężlak[1], Eryk Ciepiela[1], Marek Kasztelnik[1], Bartosz Wilk[1], Marian Bubak[1,2]

[1]ACC Cyfronet AGH,
[2]AGH University of Science and Technology, Institute of Computer Science AGH

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

Presentation Plan

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

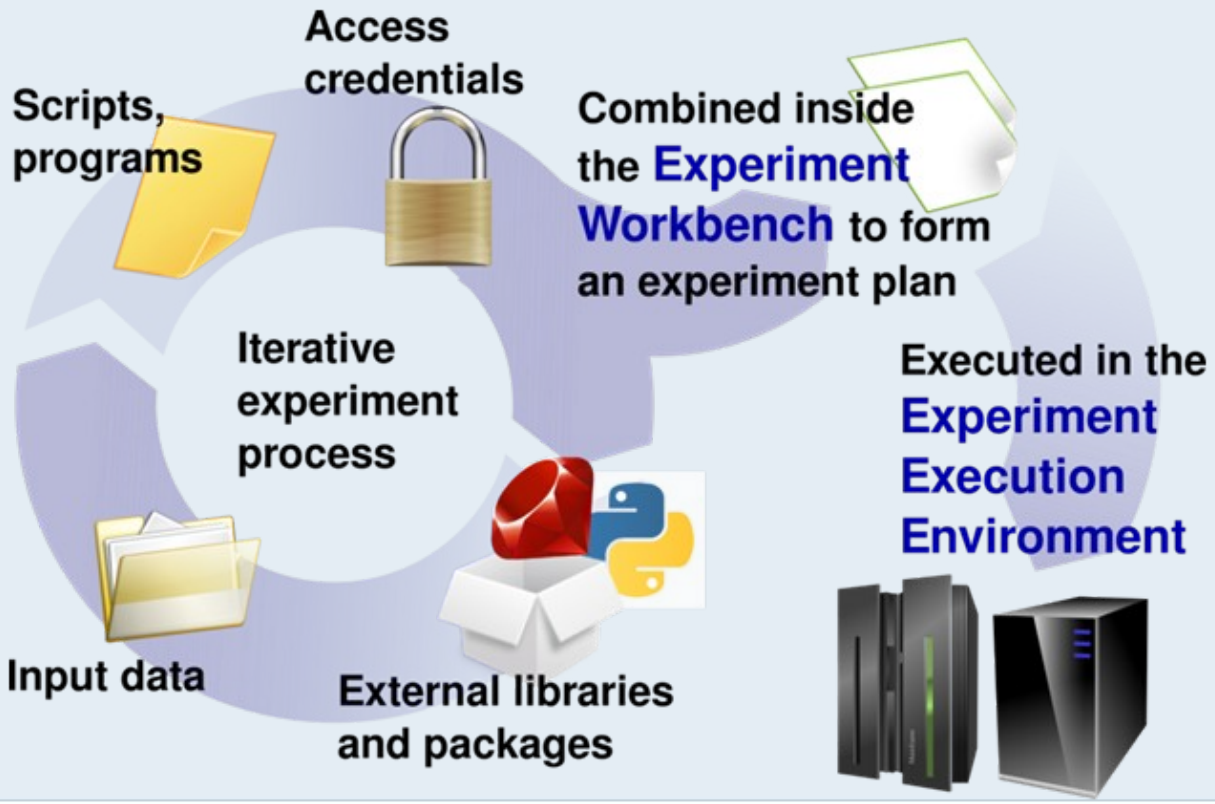
GridSpace Introduction - About

- 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 colaborative 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 Introduction - Experiment

- **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

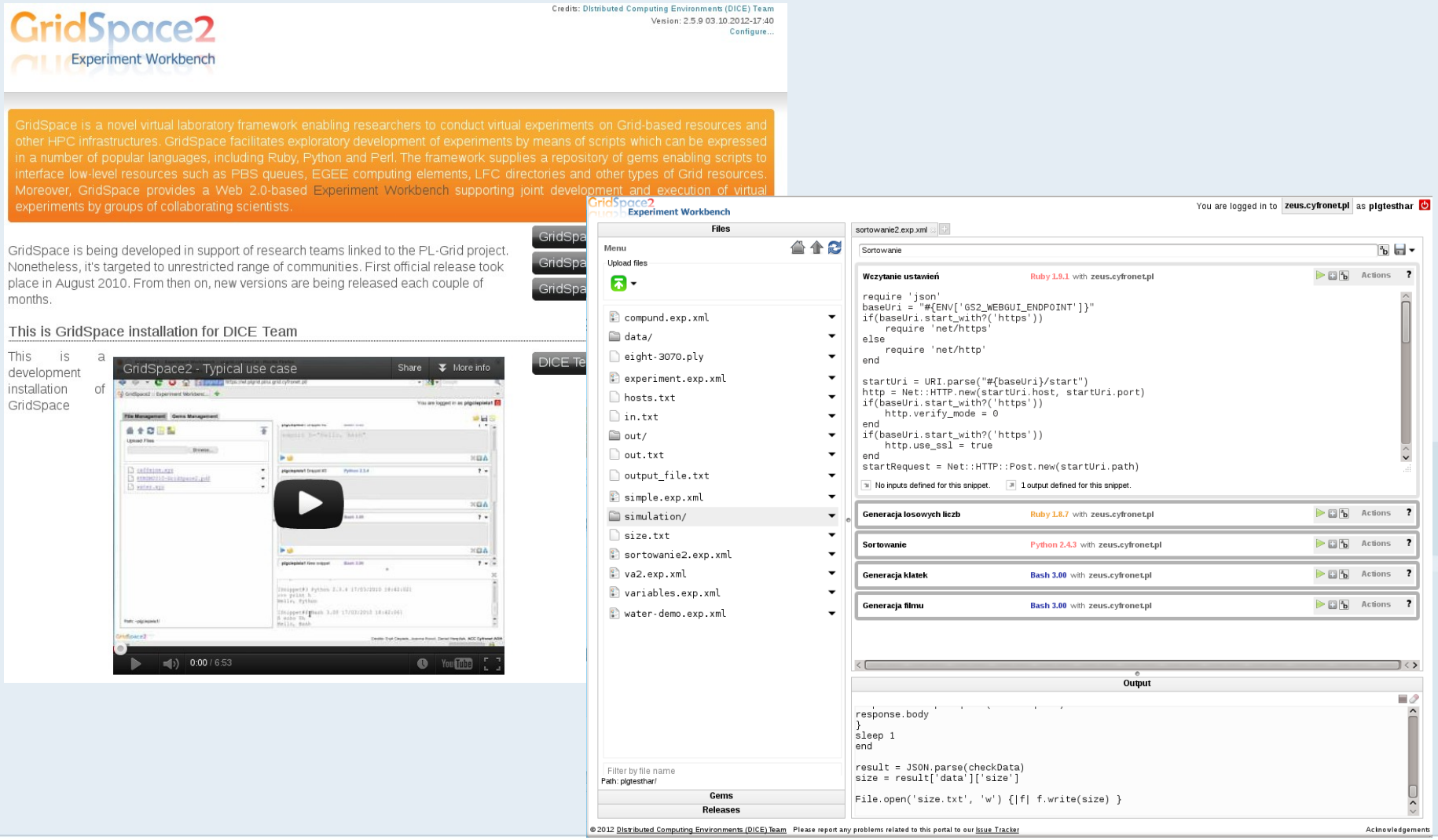


GridSpace Introduction – Deployments

- 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

Demo

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

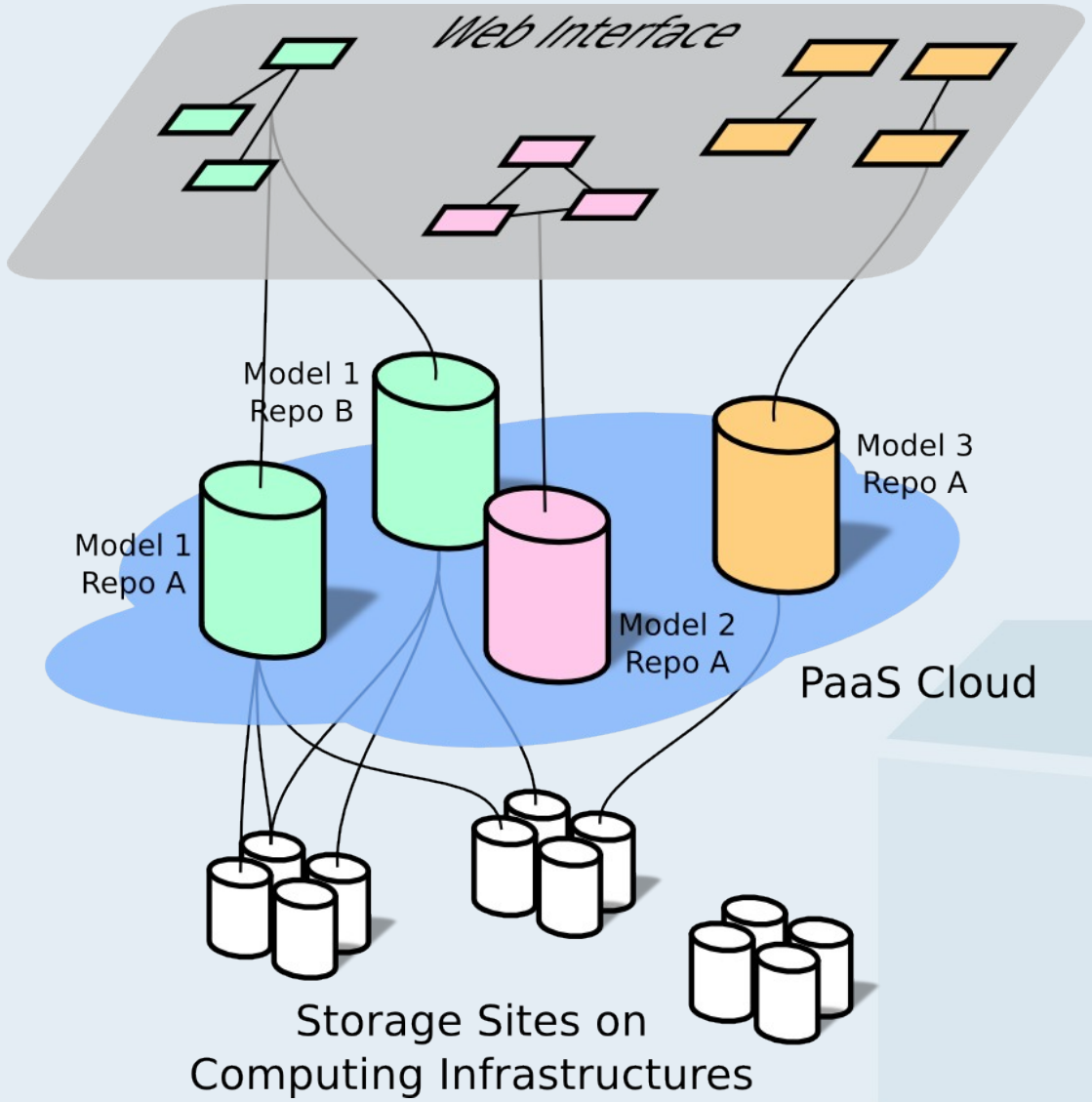


DataNet – Rationale and Objectives

- 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

DataNet – Architecture

- **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



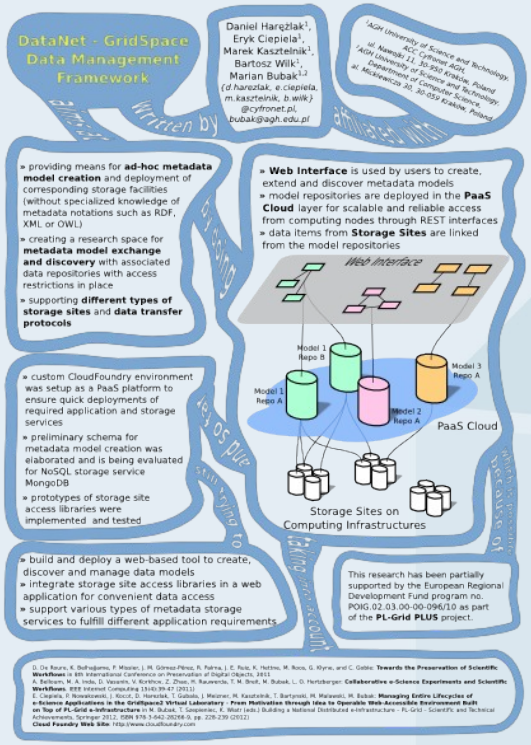
DataNet – DONEs and TODOs

- 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

Thank You

- 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