GridSpace2
Towards Science-as-a-Service Model

Eryk Ciepiela, Bartosz Wilk, Daniel Harężlak, Marek Kasztelnik, Maciej Pawlik, Jan Meizner, Marian Bubak
Academic Computer Center CYFRONET AGH

KUKDM 2013, Zakopane, 28.02.2013
Outline

- Science-as-a-Service
- e-Science-as-a-Service
- From computational experiment to service
- From computational experiment to executable scientific publications
- GridSpace2 – web-oriented distributed computing platform
- Collage Authoring Environment – executable publication framework
- Science made possible by GridSpace2/Collage Authoring Environment
- Conclusions
- Future work
Science-as-a-Service
Applying software-as-a-service principles and IT best practices to science

“as-a-service” model
- Proved successful in multiple areas in IT
- Infrastructure-as-a-Service
- Platform-as-a-Service
- Software-as-a-Service

Software-as-a-service model applied to science
- Core service providers
- Outsourcing
- Service market, competitiveness, ecosystem
- Cost-effectiveness, resource sharing, low entry costs, pay-per-use
- Scientific experiments as a service
- Science as an enterprise
- Crowdsourced science
- Open science
e-Science-as-a-Service
Applying software-as-a-service principles and IT best practices to e-science

...and especially e-science

- Data centers as computing power and data storage providers
- Software providers
- IT staff to hire
- Base software platform, tools, problem solving environments
- Computational experiment as a service
- Shared e-infrastructure, reusable and reproducible experiments
- Executable scientific publications
From computational experiments to services
Common aspects ensured or streamlined by GridSpace2

From computational experiment to service
(e.g. PL-Grid Plus domain service)

- Integration with common e-infrastructure
- User access management
- Respected intellectual property rights
- Cataloging, indexing
- Accessibility
- Reusability
- Experiment availability
- Documentation availability
- Examples availability
- Monitoring
- Accounting
- Maintenance
- User support
- Quality assurance
- Security assurance

Provisioning of services at little cost with GridSpace2 in platform-as-a-service model
From computational experiments to e-publications
Common aspects **ensured** or **streamlined** by GridSpace2

**From computational experiment to e-publication**
(e.g. Computers and Graphics Journal Special Issue)

- Scientific relevance
- Originality
- Reproducibility
- Verifiability
- Transparency
- Primary data and results availability
- Support for review process
- Publication factors
- Publishing medium

**Provisioning of e-publications at little cost with GridSpace2 in platform-as-a-service model**
GridSpace2 and Collage Authoring Environment

Concept
GridSpace2 and Collage Authoring Environment
Managing computational experiments

Collage Authoring Workbench

Menu
Upload files

- arguments.txt
- collatz.exp.xml
- hwm.png
- iters.png
- last.png
- results_raw.txt

Filter by file name
Path: eciepiela/collatz/

Files

Computing sequences for Ruby 1.8.7 with collage-exphost.elsevier.com

Generating plots GnuPlot 4.2.6 with collage-exphost.elsevier.com

set output 'collatz/hwm.png'
plot 'collatz/results_raw.txt' using 1:3 title "The biggest number (high water mark) reached when iterating"

set output 'collatz/last.png'
plot 'collatz/results_raw.txt' using 1:4 title "Value in last iteration"

Output

gnuplot> set output 'collatz/hwm.png'
gnuplot> plot 'collatz/results_raw.txt' using 1:3 title "The biggest number (high water mark) reached when iterating"
GridSpace2
and Collage Authoring Environment
Accessing services and interacting with executable publications

The Collatz Conjecture
ceiepia, DOI: 10.0000/1358511059290

The experiment was released by eciepia on Fri Jan 18 13:10:59 CET 2013 in the private scope. No, below is not an article. It’s only generated text with injected labels that navigate to particular experiment items to show you how in-text links work.

Data 4: The biggest number (high water mark) reached when iterating

Data 4 (your copy): The biggest number (high water mark) reached when iterating.

Data 2 (original): Raw results to be visualized afterwards.

Data 3: Number of iterations

Data 4: The biggest number (high water mark) reached when iterating

Data 5: Value in last iteration

Save experiment
Reset
Science made possible by GridSpace2/Collage platform


- Your publication maybe?
Publications describing GridSpace2/Collage and its applications


- P. Pierzchała: *Multiscale Applications in the GridSpace Virtual Laboratory*, Master of Science Thesis supervised by Katarzyna Rycz; AGH University of Science and Technology (2012)

Conclusions

- Inefficient traditional model pushes scientists and research organizations to seek new business models for science.
- Software-as-a-service model proved effective in IT market, science-as-a-service can too, especially in e-science.
- E-science services are not only software and data but complete and consumable products.
- Scientific findings and methods need new ways for communicating and disseminating to embrace reusability, verifiability, reproducibility, transparency, executability.
- GridSpace2 is a platform-as-a-service that facilitates provisioning of experiment-as-a-service at little cost.
- Collage Authoring Environment overlays GridSpace2 and enables executable scientific publications.
- Collage Authoring Environment is integrated with Elsevier ScienceDirect portal.
Future Work

- In the context of PL-Grid Plus project GridSpace2 will be used as a platform for streamlined provisioning of domain services.
- Multiscale applications as services will be powered by GridSpace2 platform in the scope of the Mapper project.
- Collage Authoring Environment will empower special issue of “Computers and Graphics” journal featuring executable publications.

- We’re eager to support research teams in provisioning of services and e-publications.