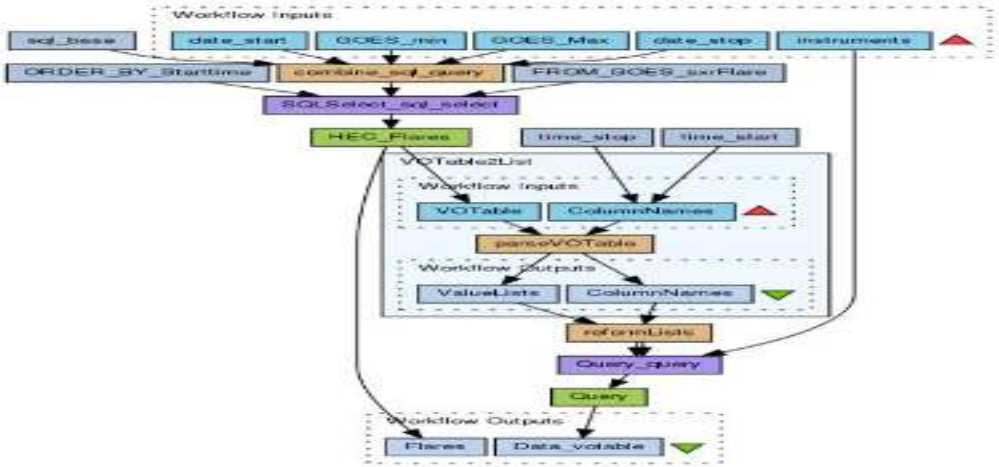


Workflows in Computational Chemistry

Prof. Gabor Terstyanszky
Centre for Parallel Computing
University of Westminster
School of Open Science Cloud
Perugia, Italy
05 – 09 June 2017

Challenges and Motivation

Challenges

- Learn how to create and use workflows
- Many different workflow systems exist that are not interoperable
- Technological choice (data and computer resources) isolates users and user communities

Benefits of workflow

- Share your own workflow, re-use workflows of others
- Create and execute ‘meta-workflows’: built from smaller workflows that use different workflow languages/technologies
- Combine workflows and DCIs

Key Players and Challenges

Domian researchers

Researchers of one particular research field, for example Astrophysicists, Computational Chemists, Heliophysicists, Bio Scientists, etc. with basic computing knowledge

10 or 100 thousands or even millions

Challenges: They are not familiar with the technology to run experiments on computing infrastructures and probably they will never learn it.

Workflow developers

They are familiar with both Computer Science and a particular research field

up to a few thousands

Workflow system developers

Computer Scientists with knowledge about data and compute technologies

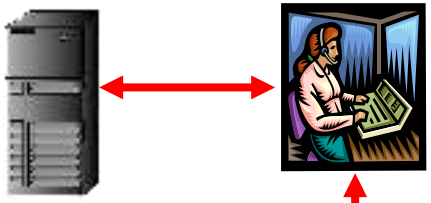
up to a few hundreds

3

Workflow Developer's Scenario

They want
- to develop & publish WFs

They need
- to develop WFs somewhere
- to publish WFs somewhere
- to manage WFs somewhere
- to execute WFs somewhere



Science Gateway

Local clusters

Supercomputers

Clouds

Supercomputer based SGs (DEISA, TeraGrid)

Cluster based service grids (SGs) (EGEE, OSG, etc.)

Desktop grids (DGs) (BOINC, Condor, etc.)

Grid systems

E-science infrastructure

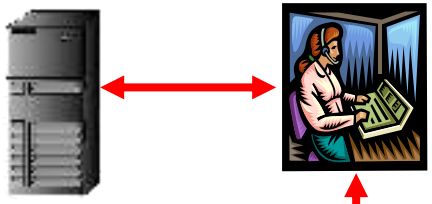
Domain Researcher's Scenario

They want

- to run experiments

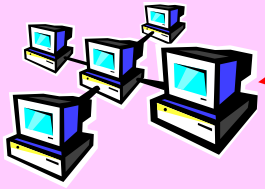
They need

- to run experiments seamlessly i.e. executing workflows which access to data and compute resources hiding all technical details



Workflow Repository

Science Gateway



Local clusters

Supercomputers

Clouds

Supercomputer based SGs (DEISA, TeraGrid)

Cluster based service grids (SGs) (EGEE, OSG, etc.)

Desktop grids (DGs) (BOINC, Condor, etc.)

Grid systems

E-science infrastructure

Co-operation with Research Communities

Phase 1 – introduction to the workflow technology

Target group: communities without any or basic experience in the workflow technology

Phase 2 – creating and running workflows

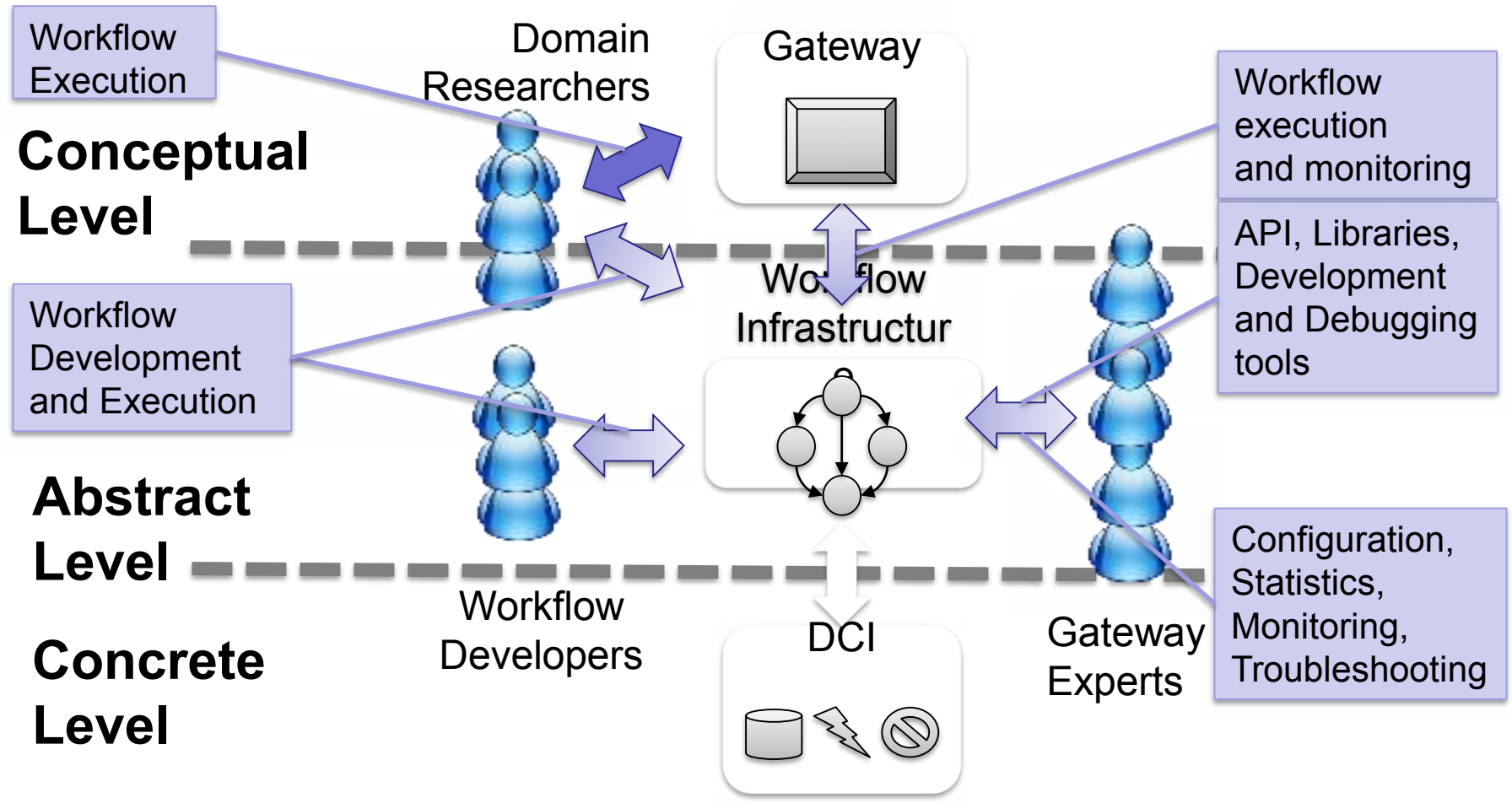
Target group: communities those use workflows to run experiments

Phase 3 – combining workflows of different workflow systems

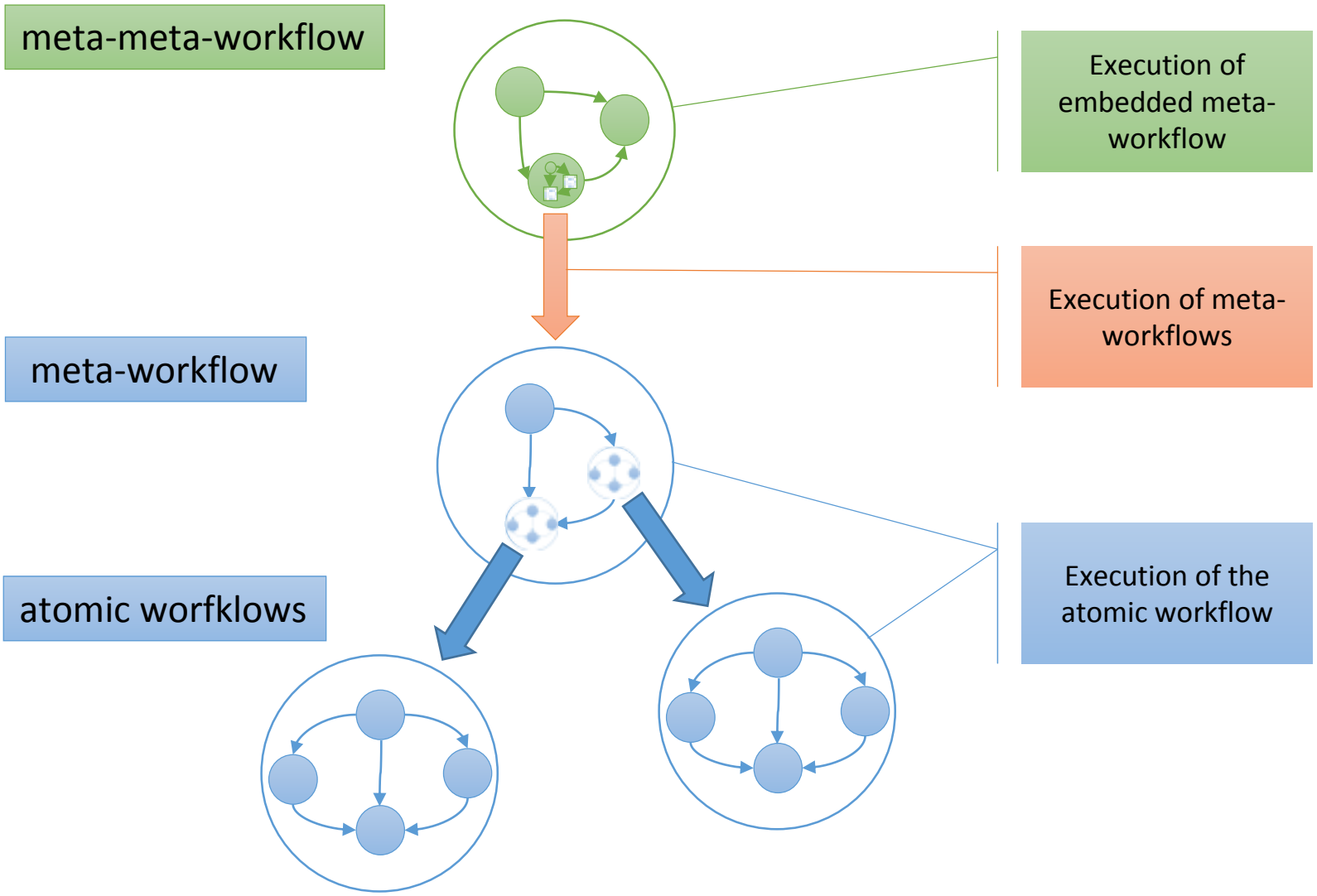
Target group: communities those use workflows to run experiments and are interested in using workflows of other workflow systems

6

Workflow Levels and Users

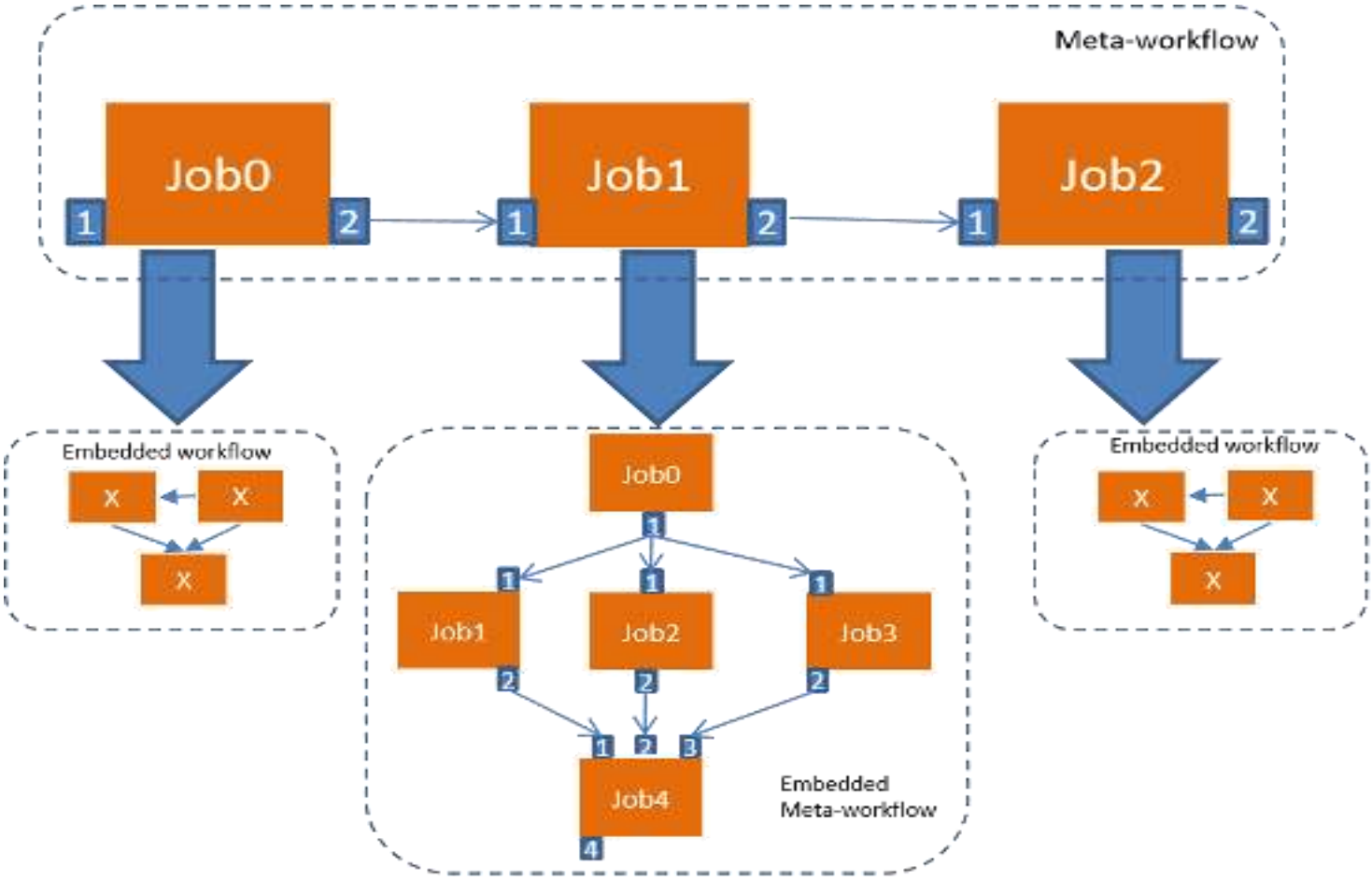


Atomic Workflow Concept

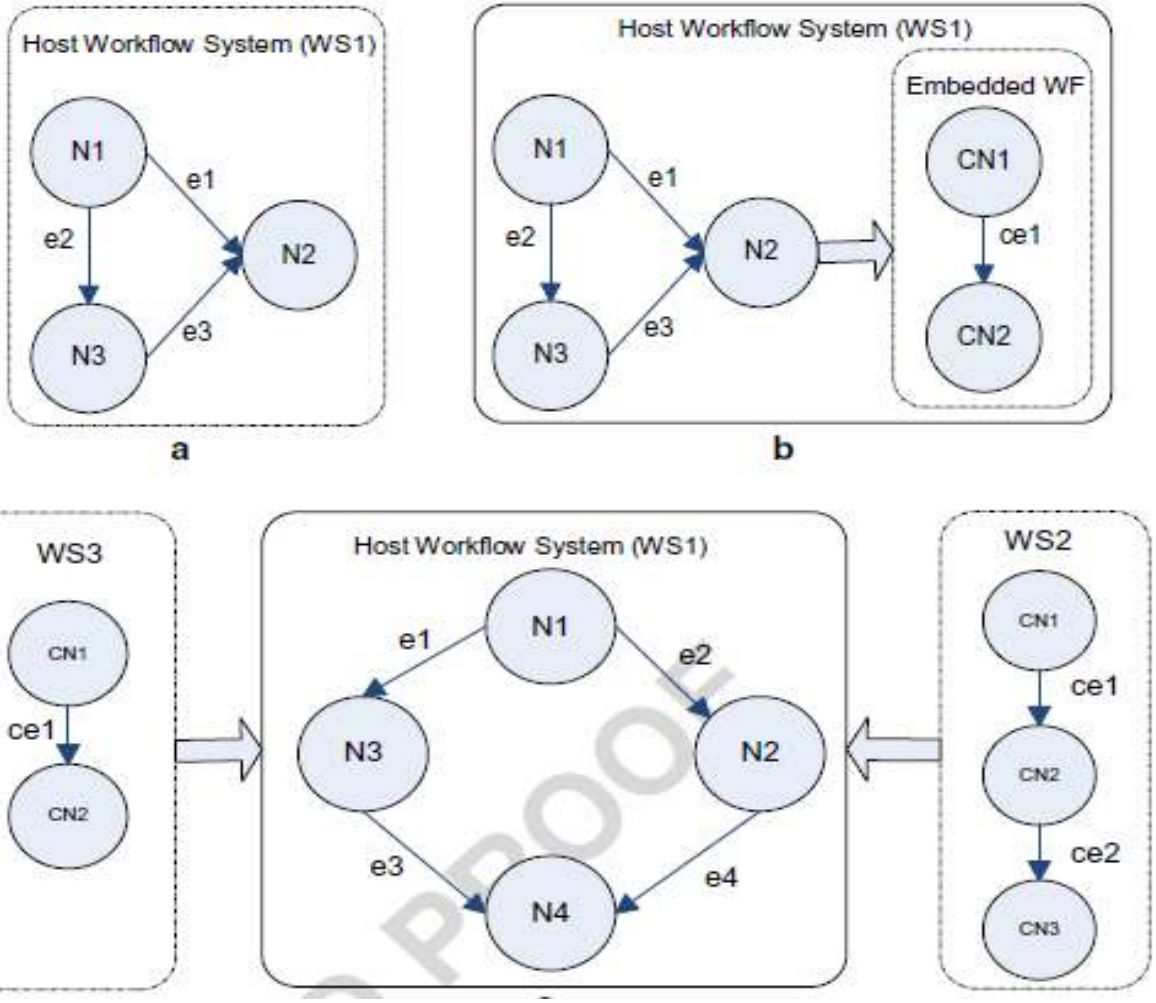


8

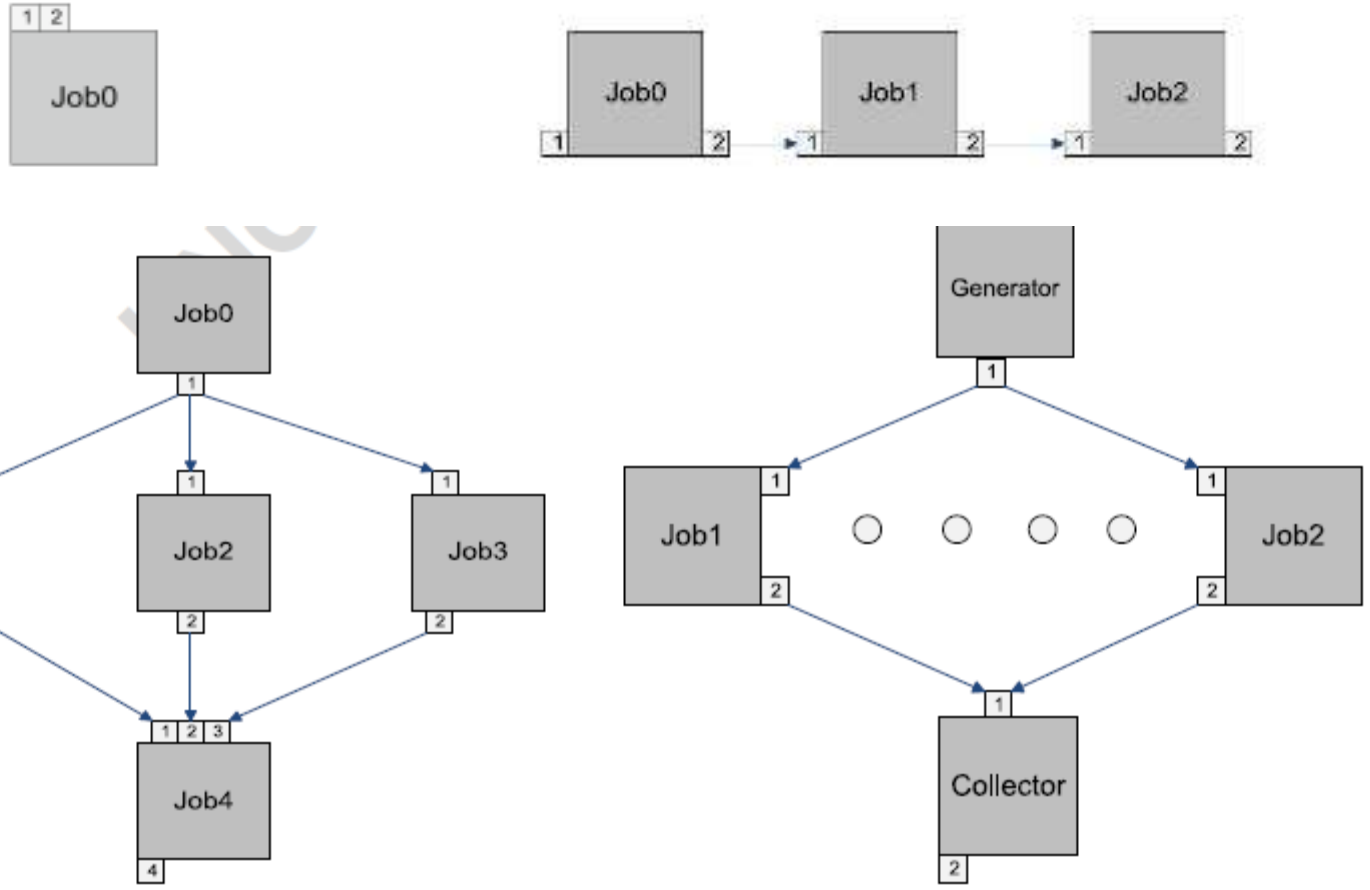
Computational Chemistry: Meta-Workflow Concept (1)



Computational Chemistry: Meta-Workflow Concept (2)



Computational Chemistry: Meta-Workflow Concept (3)



Workflows versus Scientic Experiments

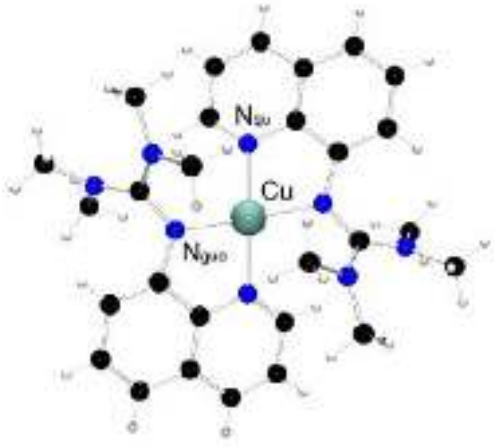
Type of Workflow	Technology	Driven by	Usage
Atomic Workflow	Basic workflows	Use Case	Execution of Atomic Workflows
Meta-workflows	Composite workflows of atomic workflows	Science Case	Orchestration of Atomic Workflows
Iterative meta-workflows	Parameter study of workflows or meta-workflows	Iterative Science Case	Parameter-Sweep Execution of Science Cases

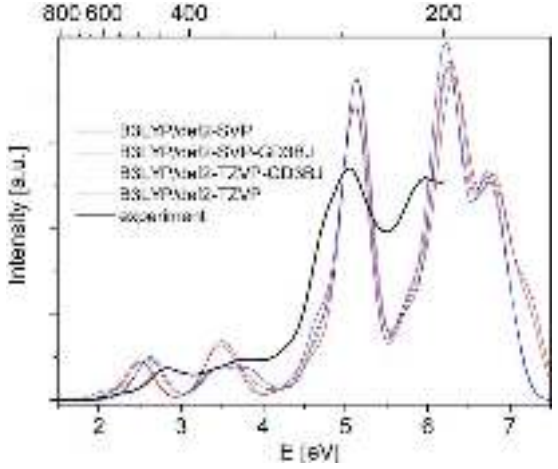
Quantum Chemistry Workflows

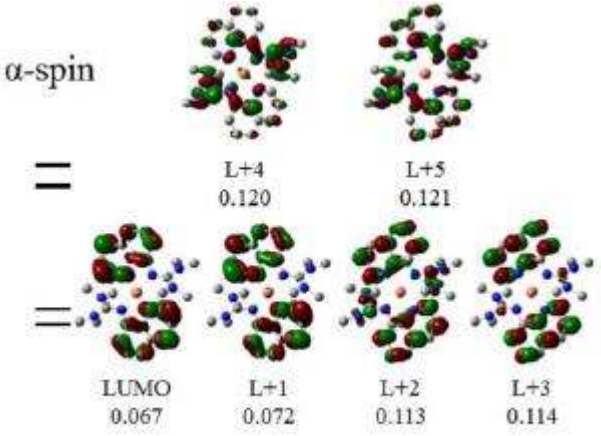
Name	Description	Engine	Middle-ware	Type	ID
Spectroscopic analysis	Explore the spectroscopic characteristics of a molecule	WS-PGRADE	UNICORE	meta-workflow	5739
Spectroscopic benchmarking	Explore the spectroscopic characteristics of a molecule with more functionals/basis sets	WS-PGRADE	UNICORE	meta-meta-workflow	5745
Parameter sweep	Benchmark a molecule using larger arrays of functionals and basis sets	WS-PGRADE	UNICORE	meta-workflow	--
Transition state analysis	Find a reaction transition state and analyse its frequencies together with the spectroscopic properties	WS-PGRADE	UNICORE	meta-workflow	5751
TD-UNI	Calculate the optical response of a molecule	UNICORE	UNICORE	Workflow	
SpecWSUNI	Explore the spectroscopic characteristics of a molecule	UNICORE	UNICORE	Workflow	
PopulationUNI	Apply various population schemes to better electronic understanding of the molecules.	UNICORE	UNICORE	Workflow	
Galaxy QM	Optimise a molecule quantum chemically in Galaxy	Galaxy	UNICORE	Workflow	5754
QM-MD	Optimise the protein scaffold by molecular dynamics and optimise then the metallo-active center by quantum mechanics together with a spectroscopic analysis	WS-PGRADE	UNICORE	meta-meta-workflow	5752

Quantum Chemistry Science Cases

Name	Description	Workflow Links
Spectroscopic Analysis	Explore the spectroscopic characteristics of a molecule	http://www.erflow.eu/spectroscopic-analysis-science-case
Spectroscopic Benchmarking	Calculation of optimized geometries, molecular orbitals, population analyses, frequencies, or optical absorptions.	http://www.erflow.eu/spectroscopic-benchmarking-science-case
Population UNI	Apply various population schemes to better electronic understanding of the molecules.	http://www.erflow.eu/population-uni-science-case

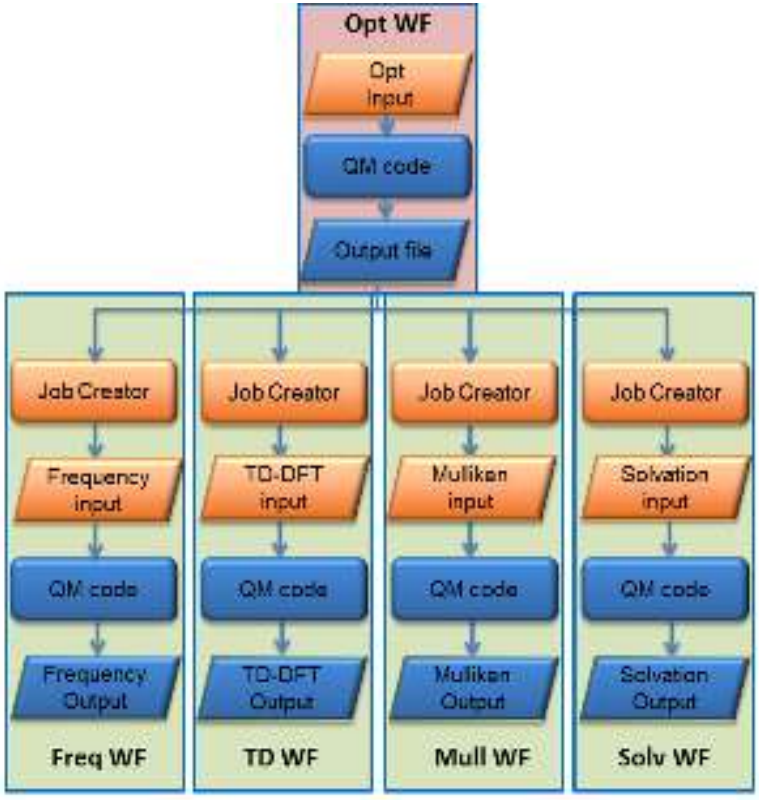


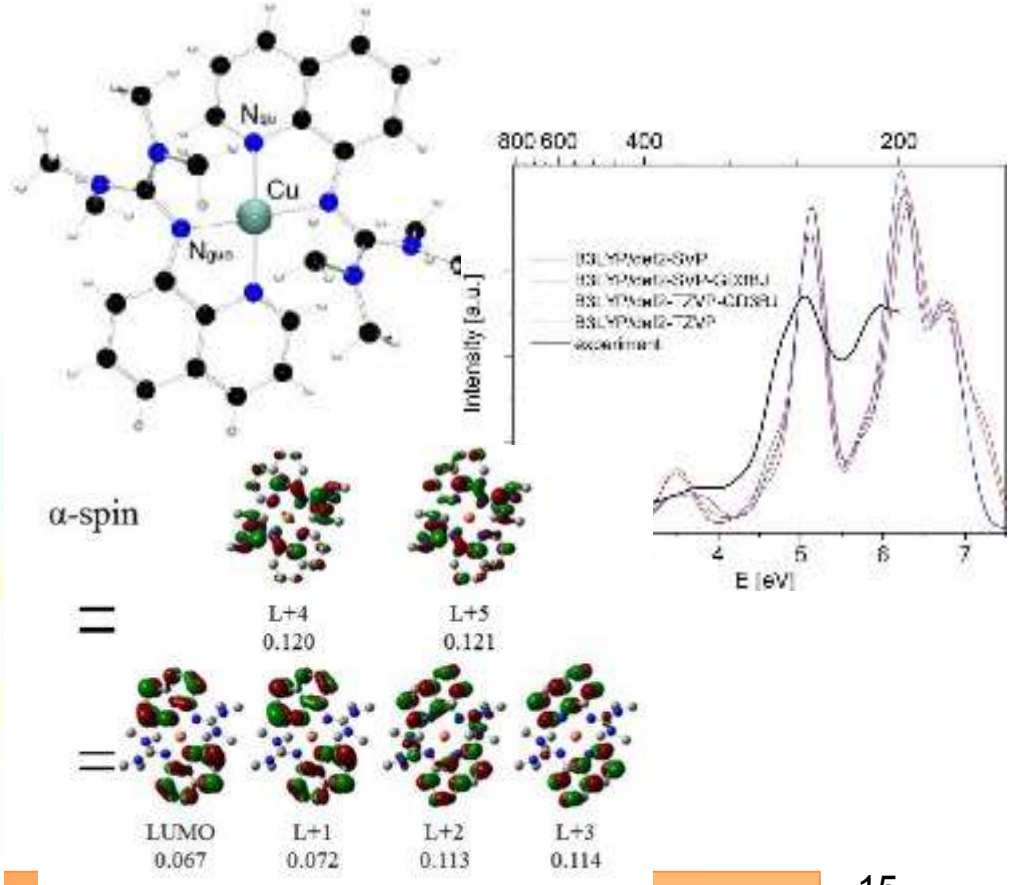




Spectroscopic Analysis Science Cases

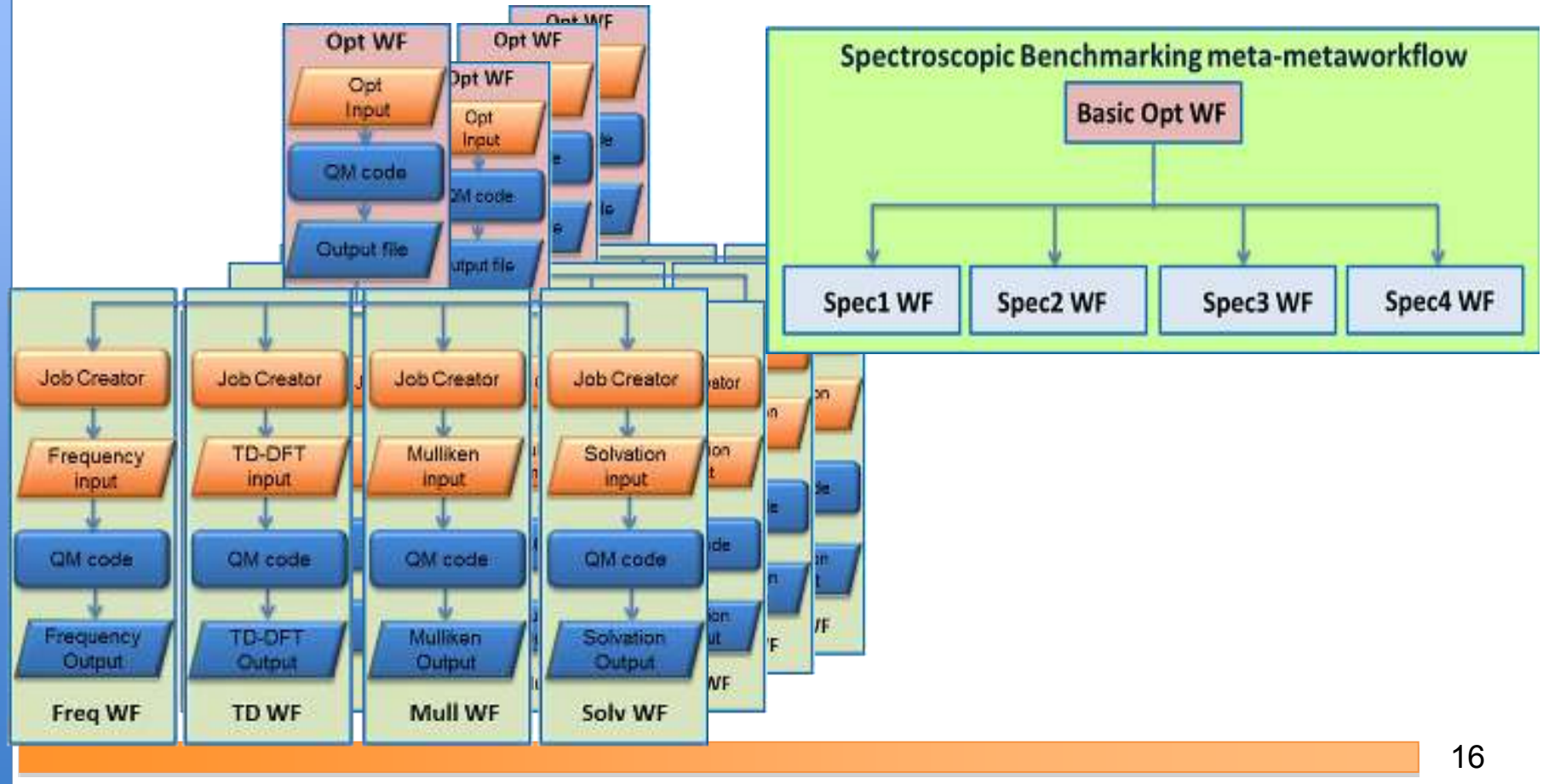
- Highly useful for Quantum Chemists in everyday work
- Full simulation of all spectroscopic features of a molecule
- Combination of 5 atomic workflows



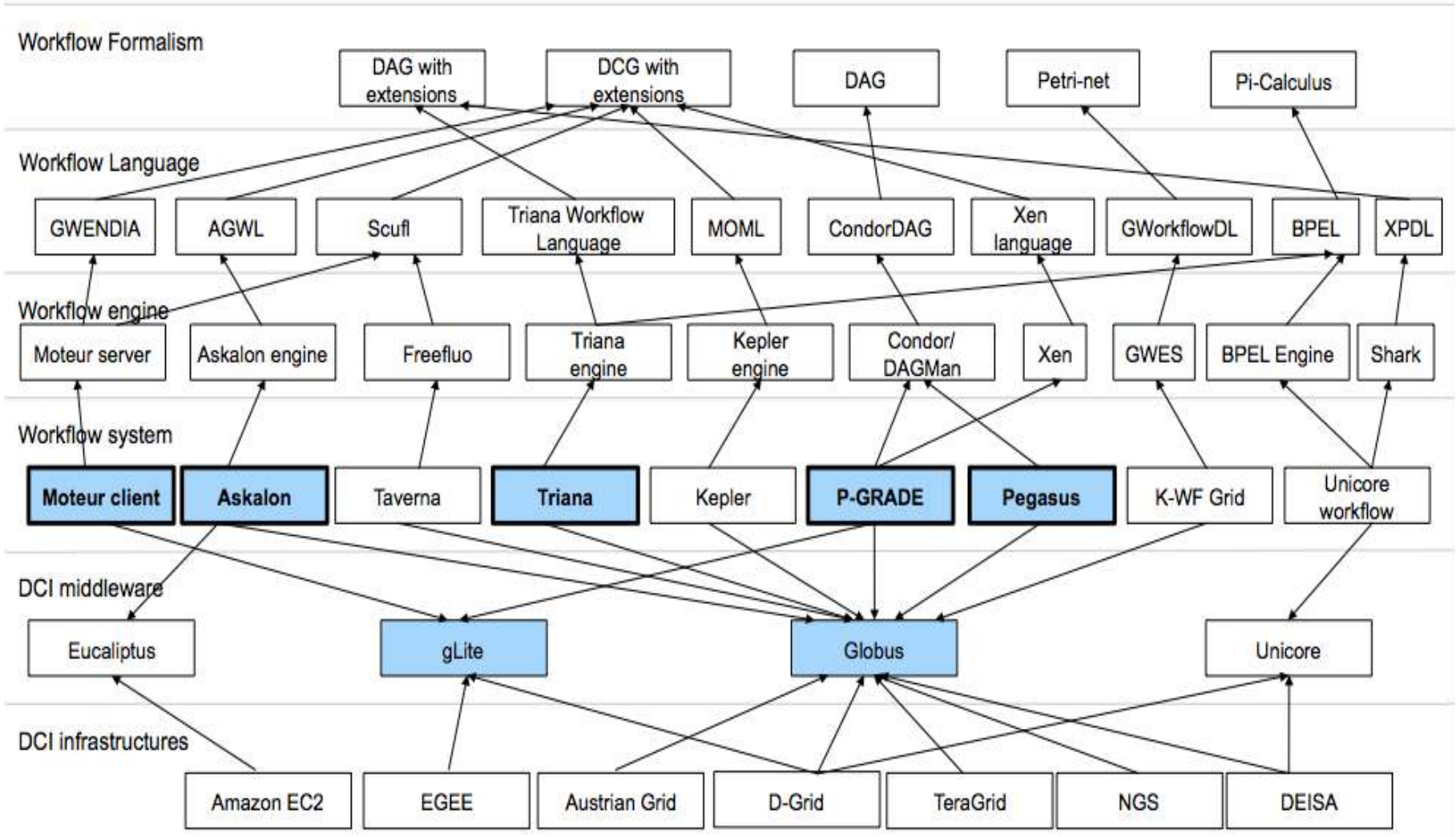


Spectroscopic Benchmarking Science Cases

- Highly useful for Quantum Chemists in everyday work
- Combination of 1 atomic workflow and 4 times the Science case 1



Workflow Interoperability Challenges



Formal Description of WEs and WFs

formal description of workflows

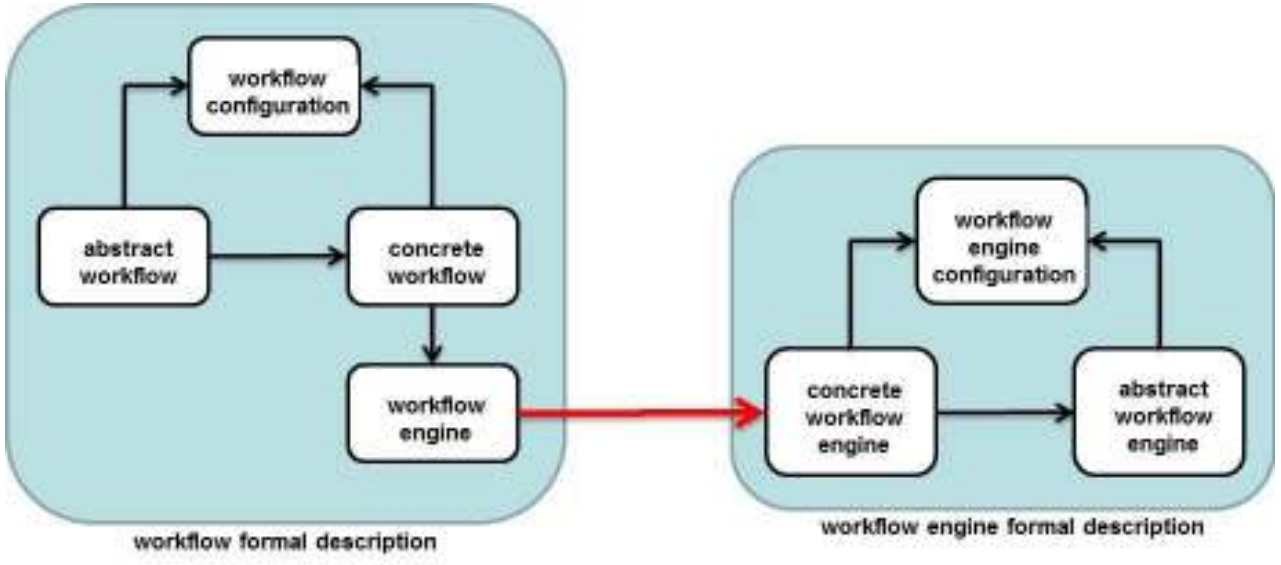
$WF = :\{WF_{abs}, WF_{cnr}, WF_{cnf}, WF_{eng}\}$

where
- WF_{abs} - abstract workflow
- WF_{cnr} - concrete workflow
- WF_{cnf} - workflow configuration
- WF_{eng} - workflow engine

formal description of workflow engines

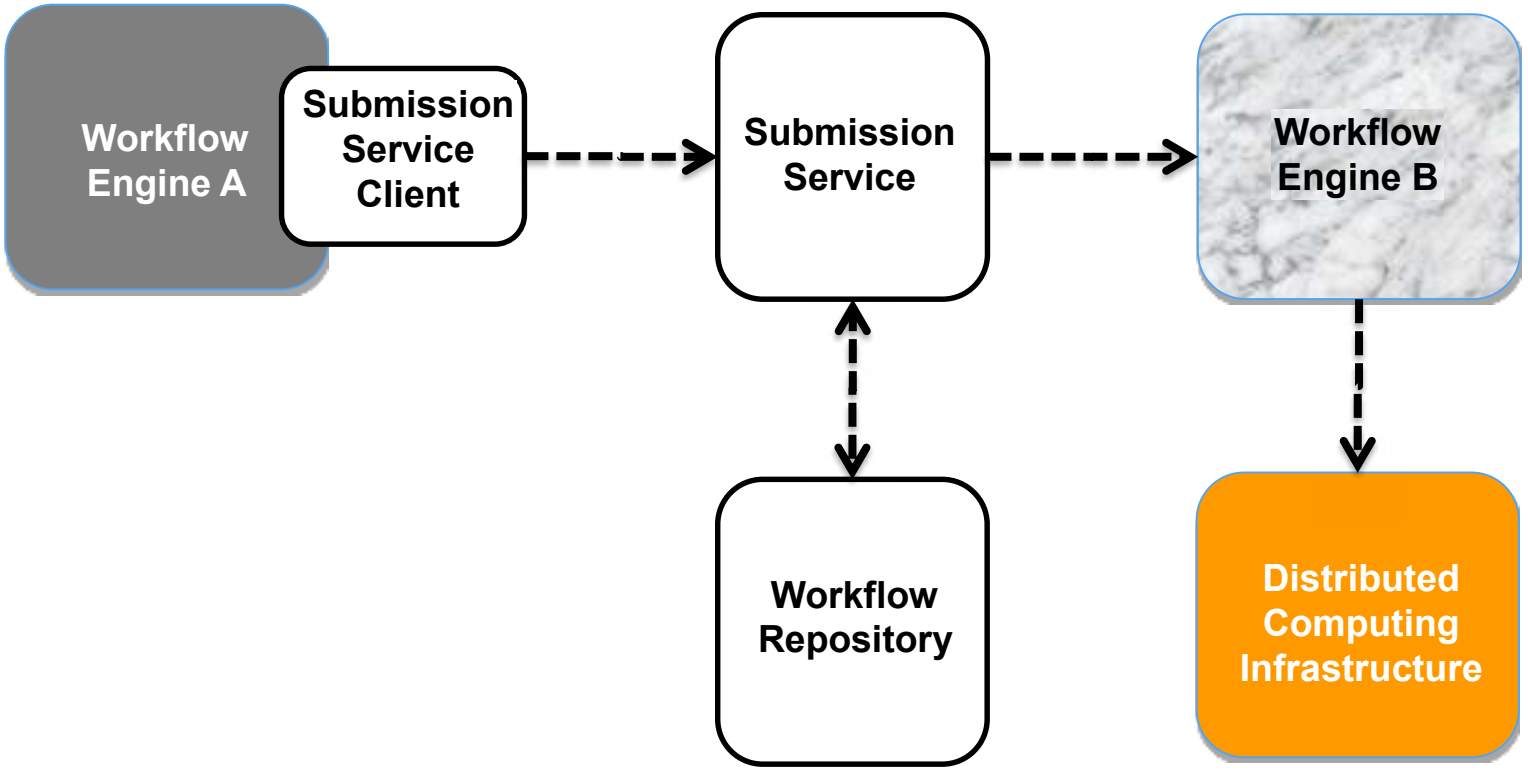
$WE = :\{WE_{abs}, WE_{cnc}, WE_{cnf}\}$

where
- WE_{abs} - abstract workflow engine
- WE_{cnr} - concrete workflow engine
- WE_{cnf} - workflow engine configuration

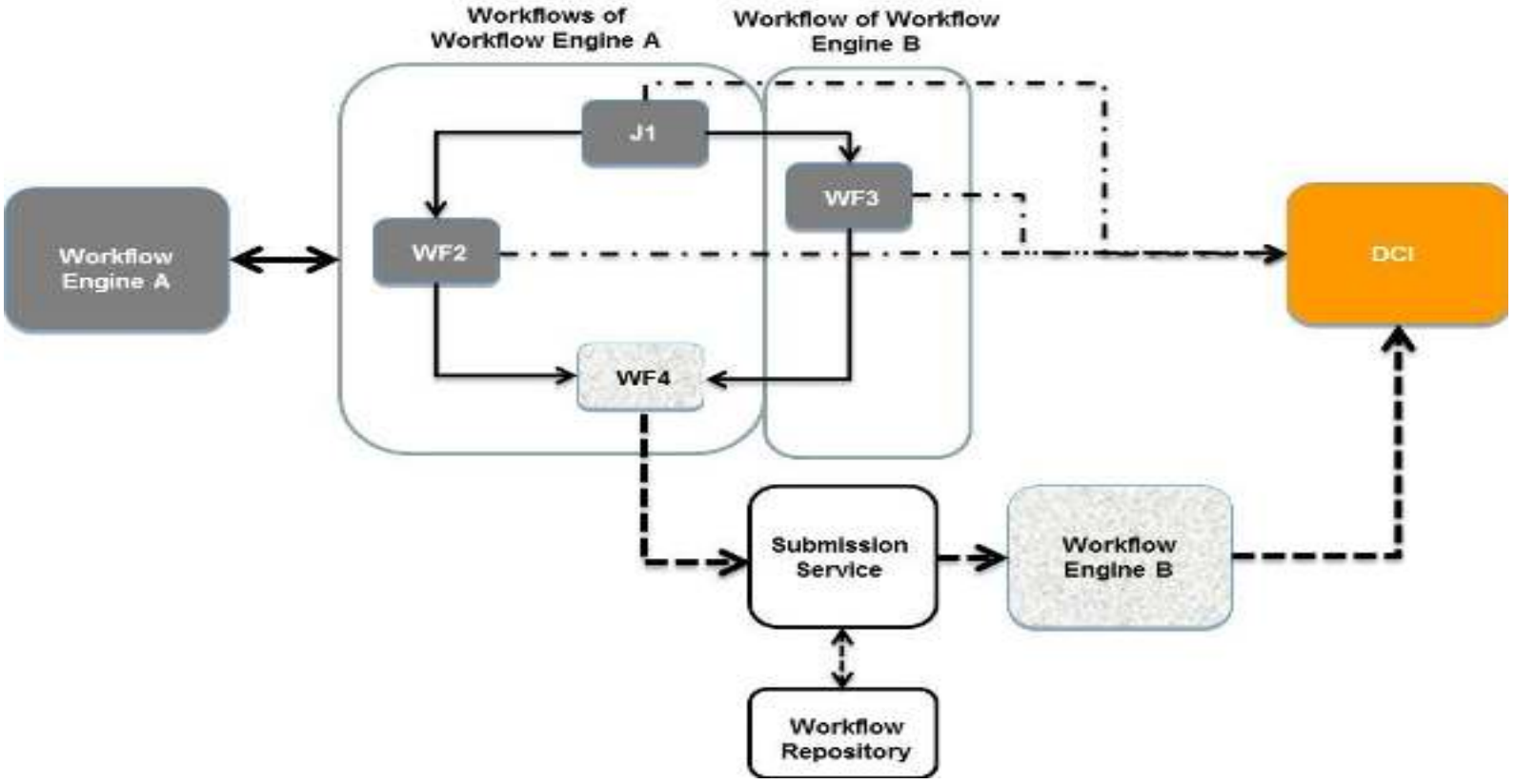


Workflow Interoperability: Coarse-Grained Interoperability (1)

CGI concept = workflow engine integration

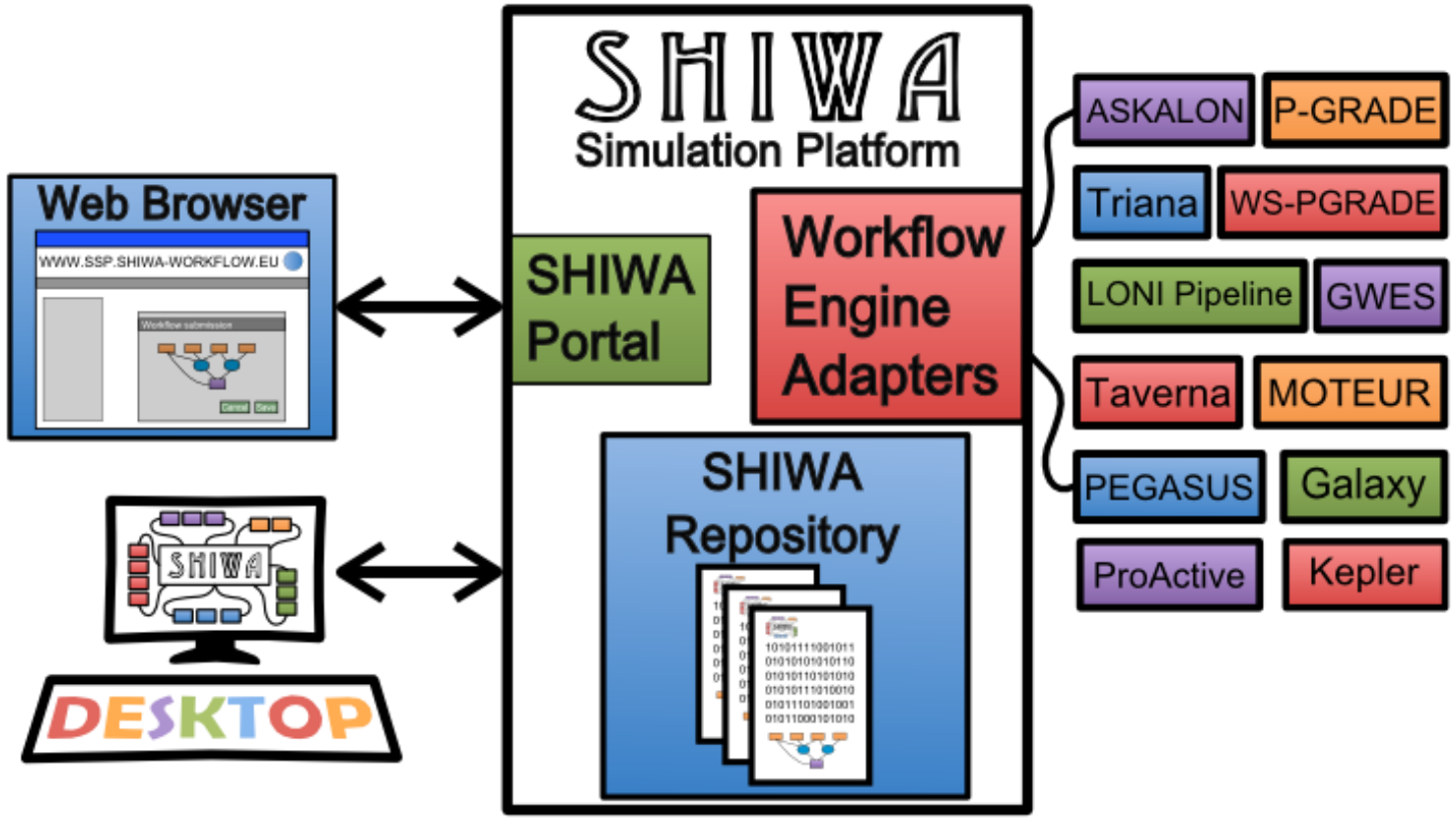


Workflow Interoperability: Coarse-Grained Interoperability (2)

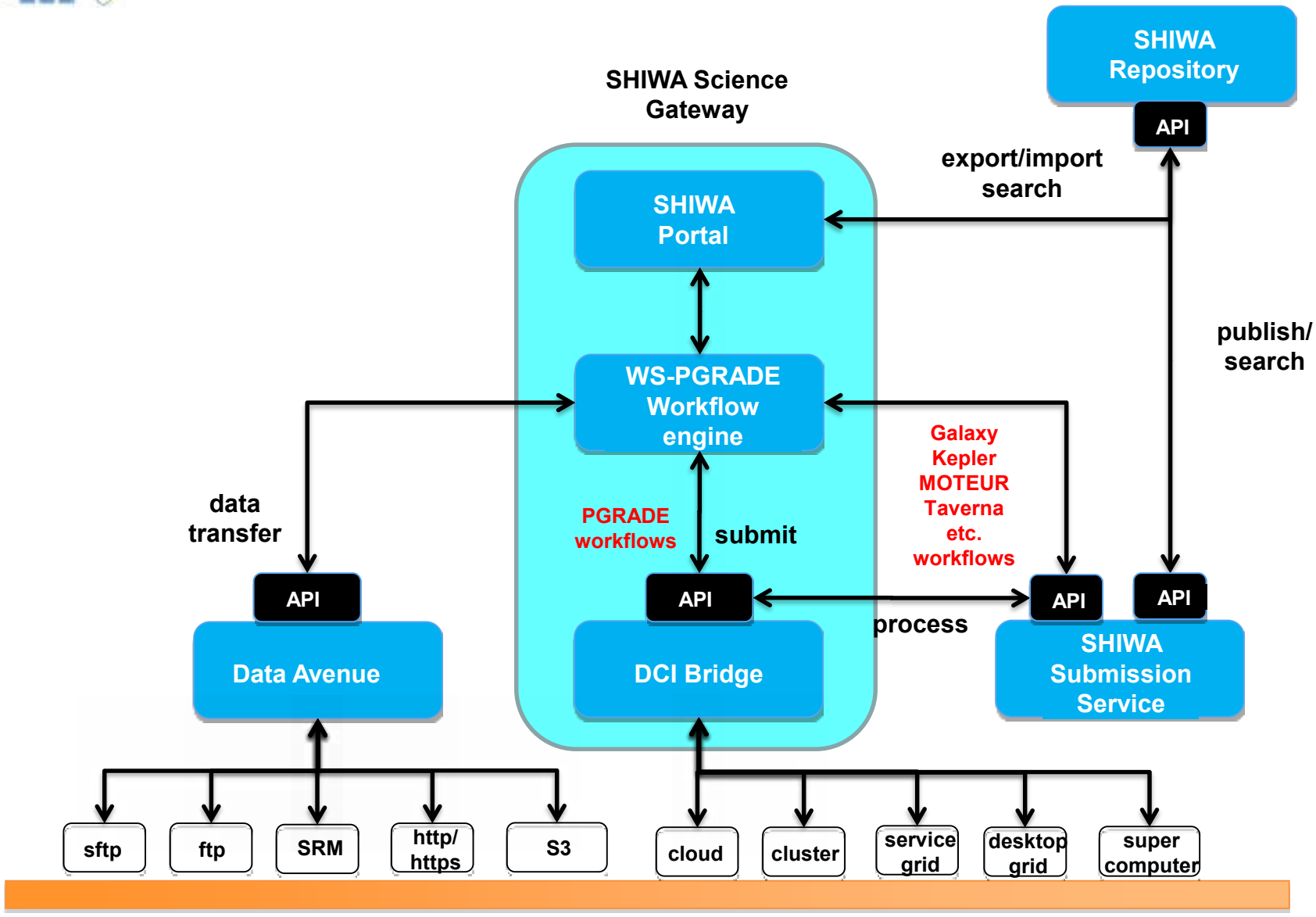


- native workflows: J1, J2, J3
- non-native workflows: WF4
  - black boxes which are managed as legacy code applications

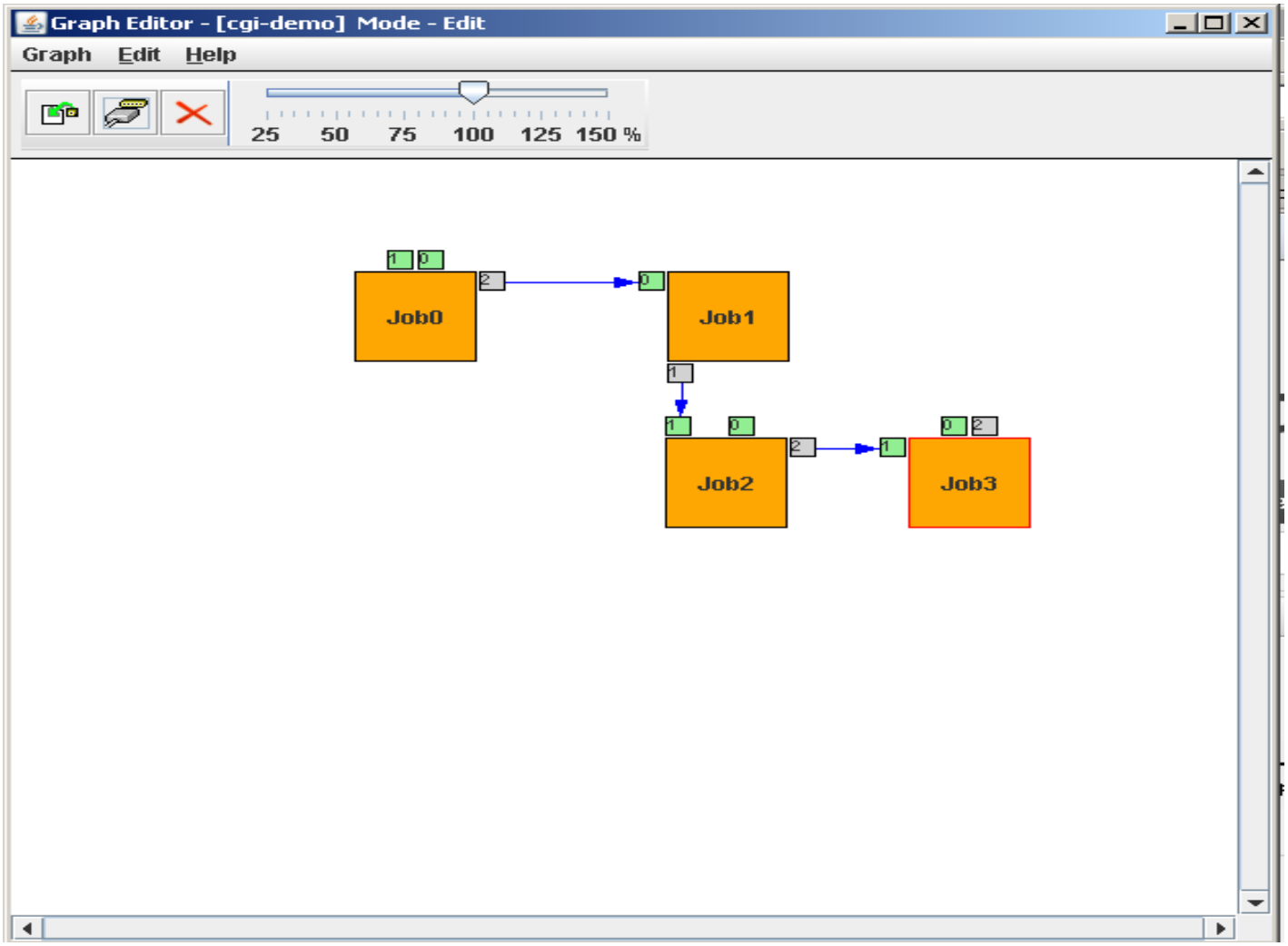
SHIWA Simulation Platform = SSP



SHIWA Simulation Platform

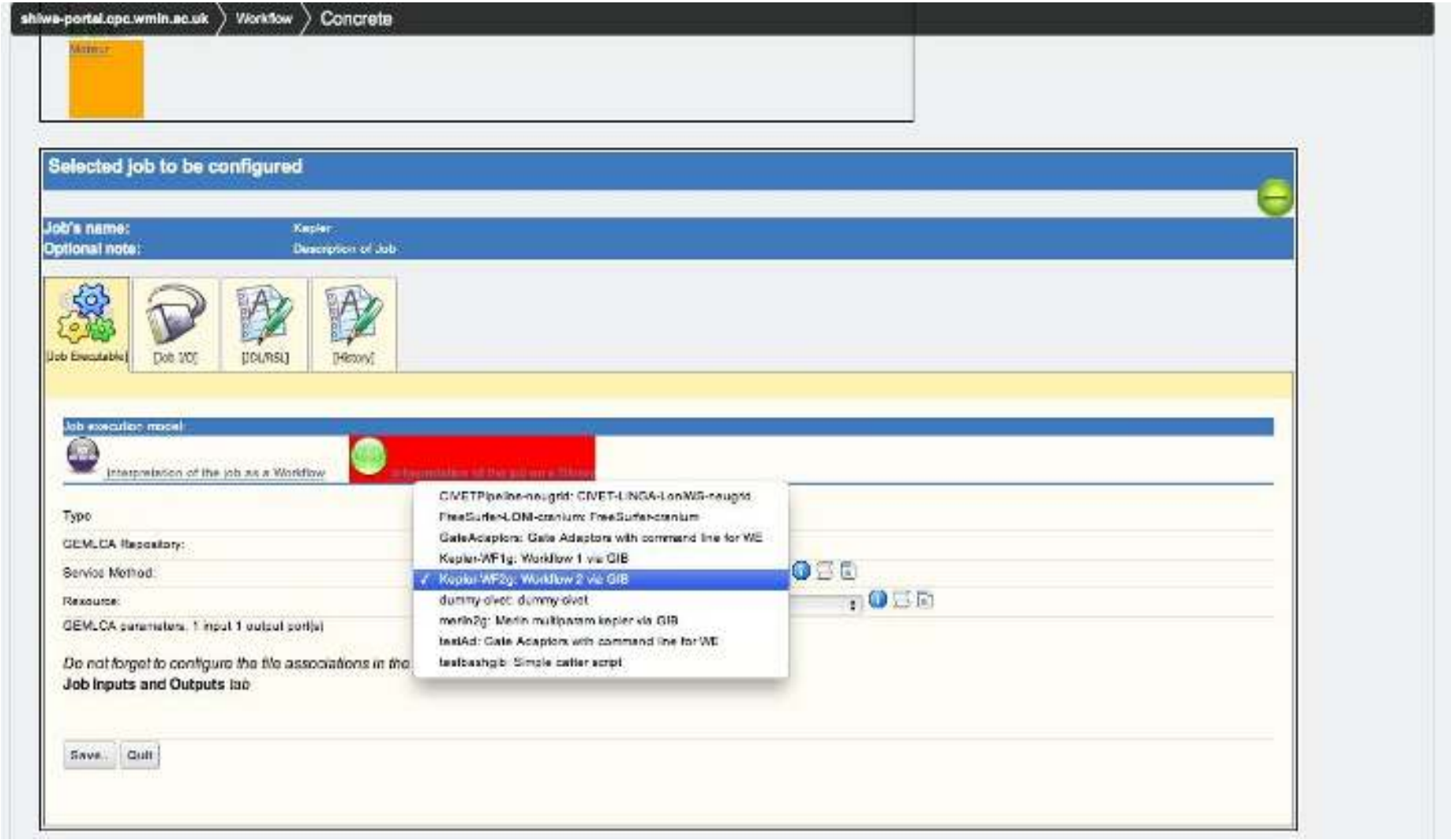


SHIWA Portal: Editing Workflow



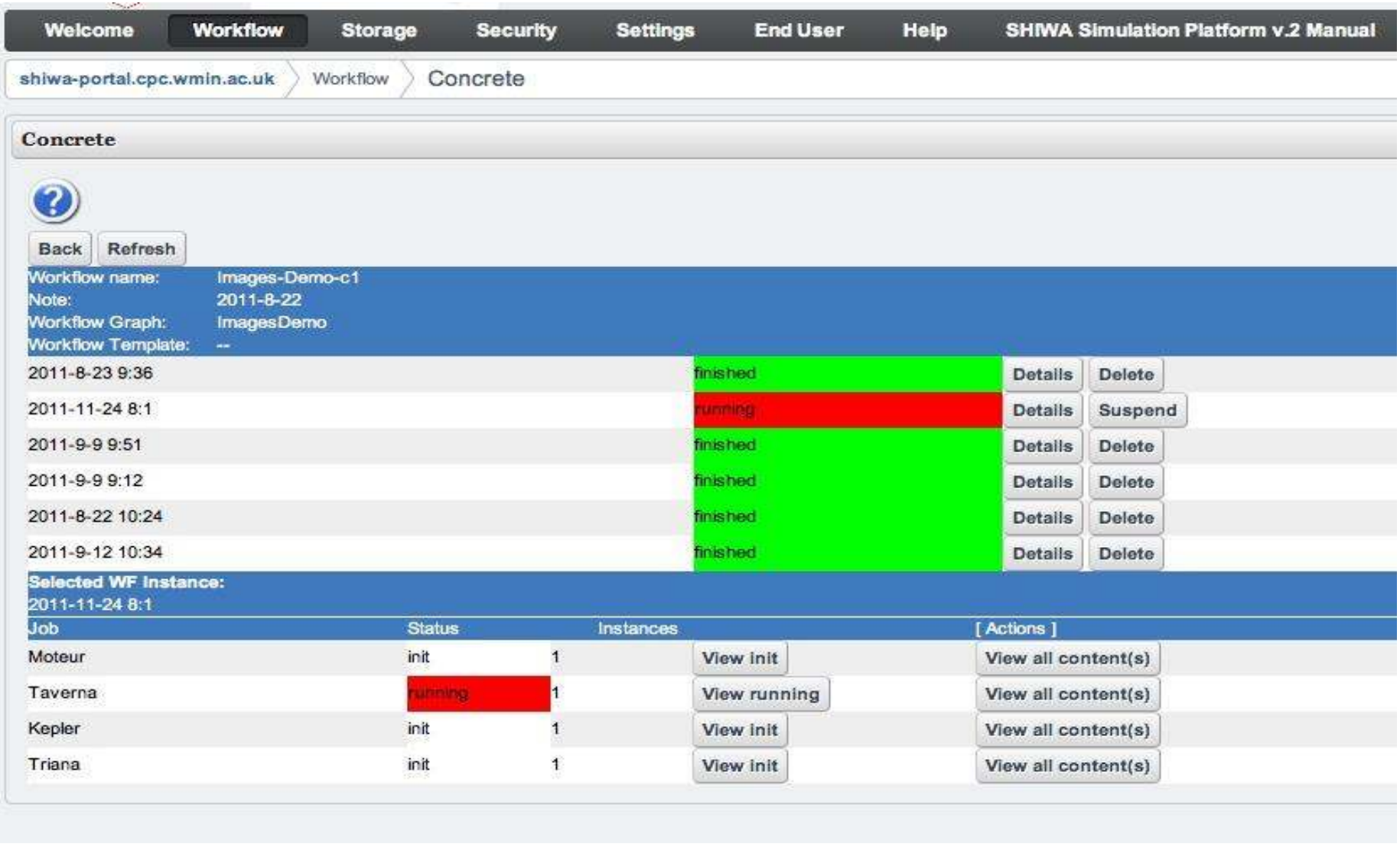
e-infrastructure

SHIWA Portal: Configuring Workflow



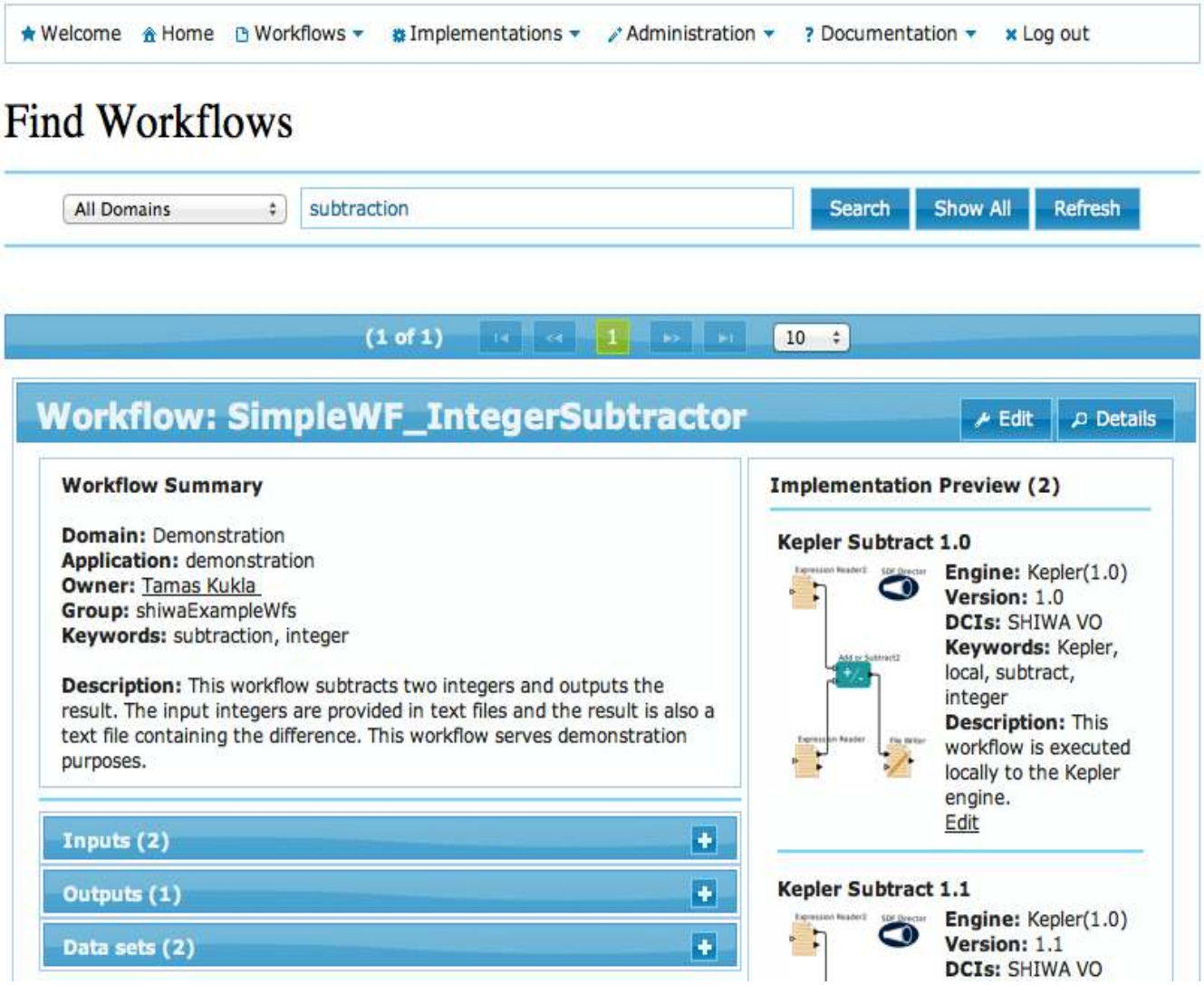
e-infrastructure

SHIWA Portal: Executing Workflow

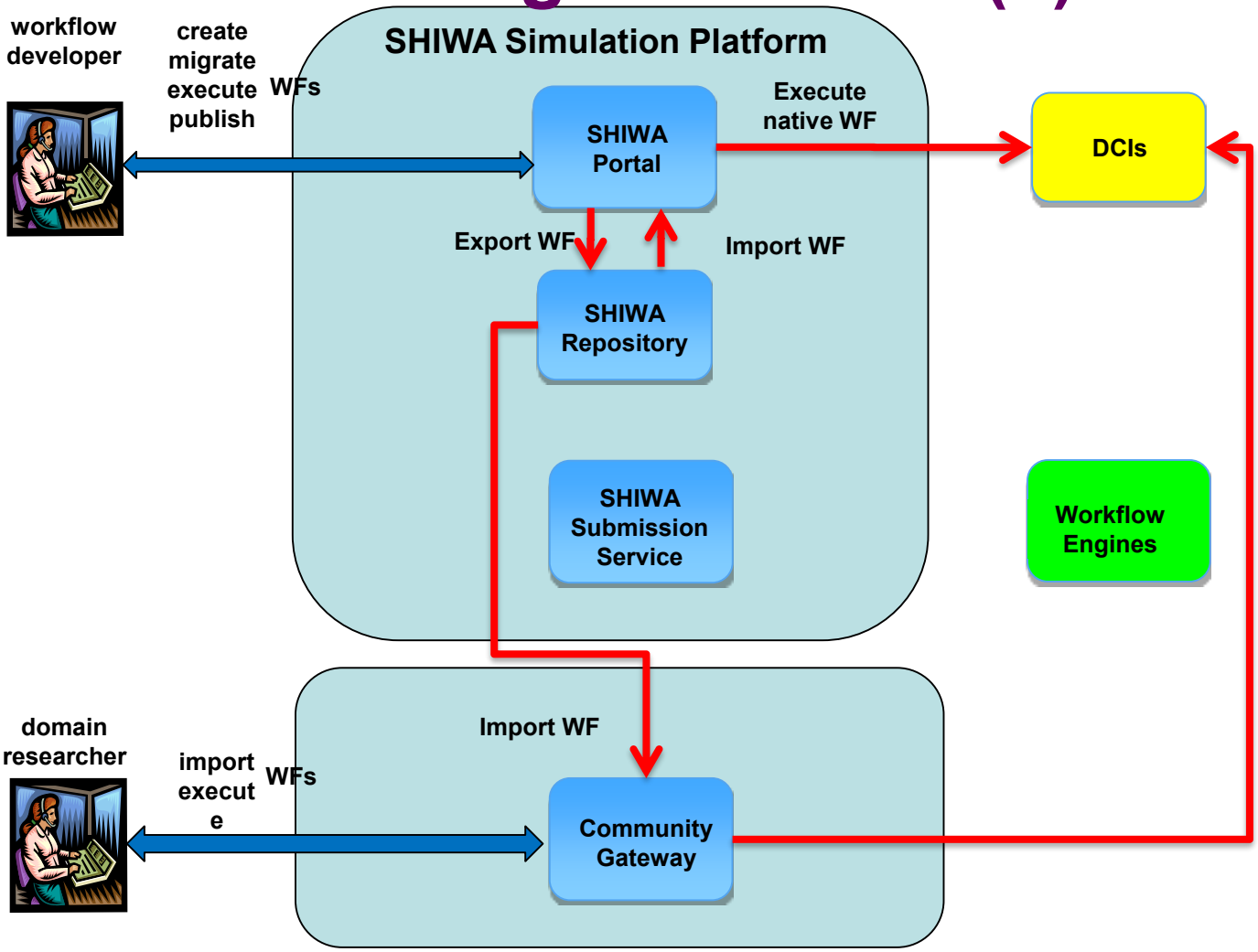


e-infrastructure

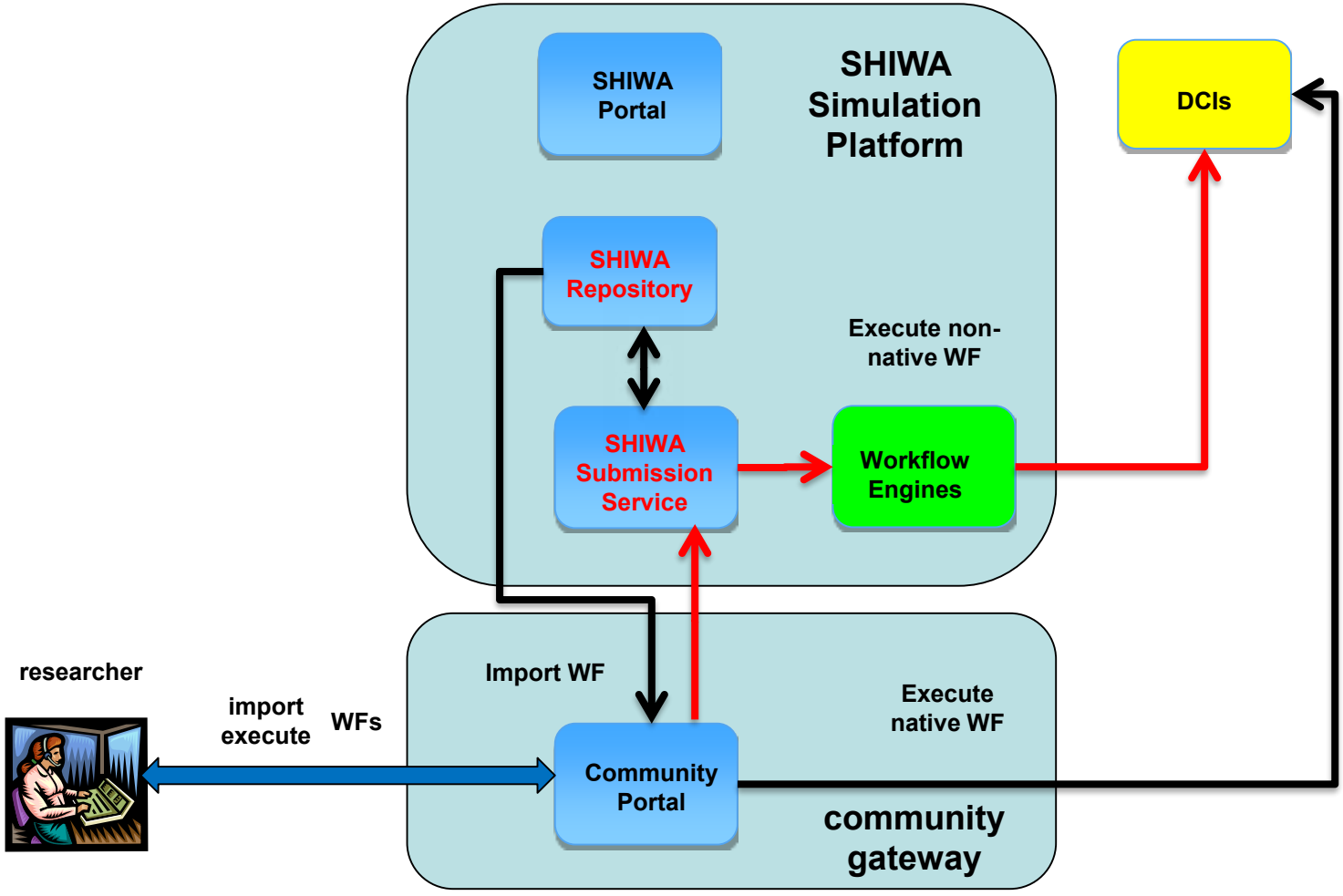
SHIWA Workflow Repository



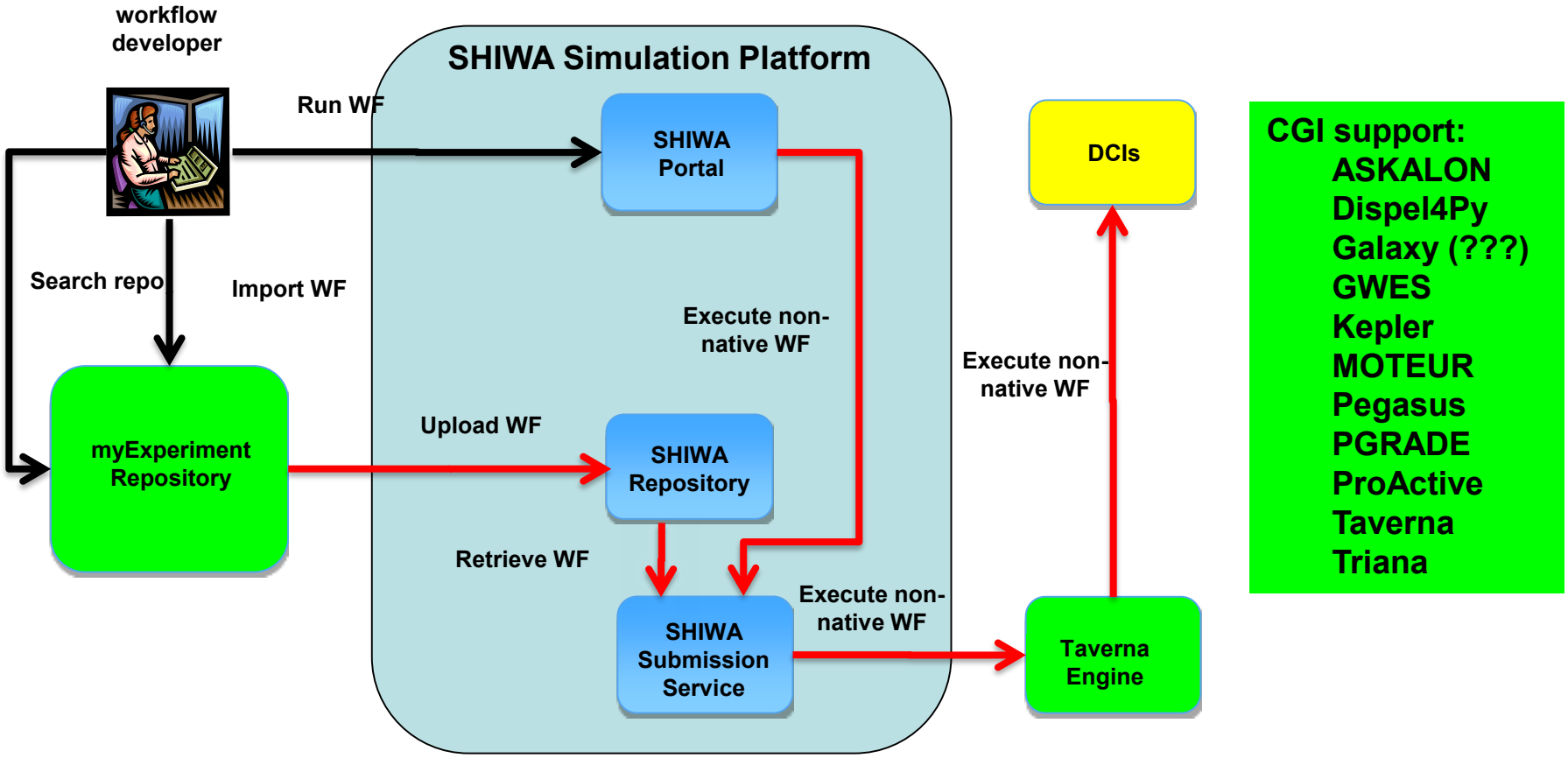
Workflow Interoperability: CGI Usage Scenario (1)



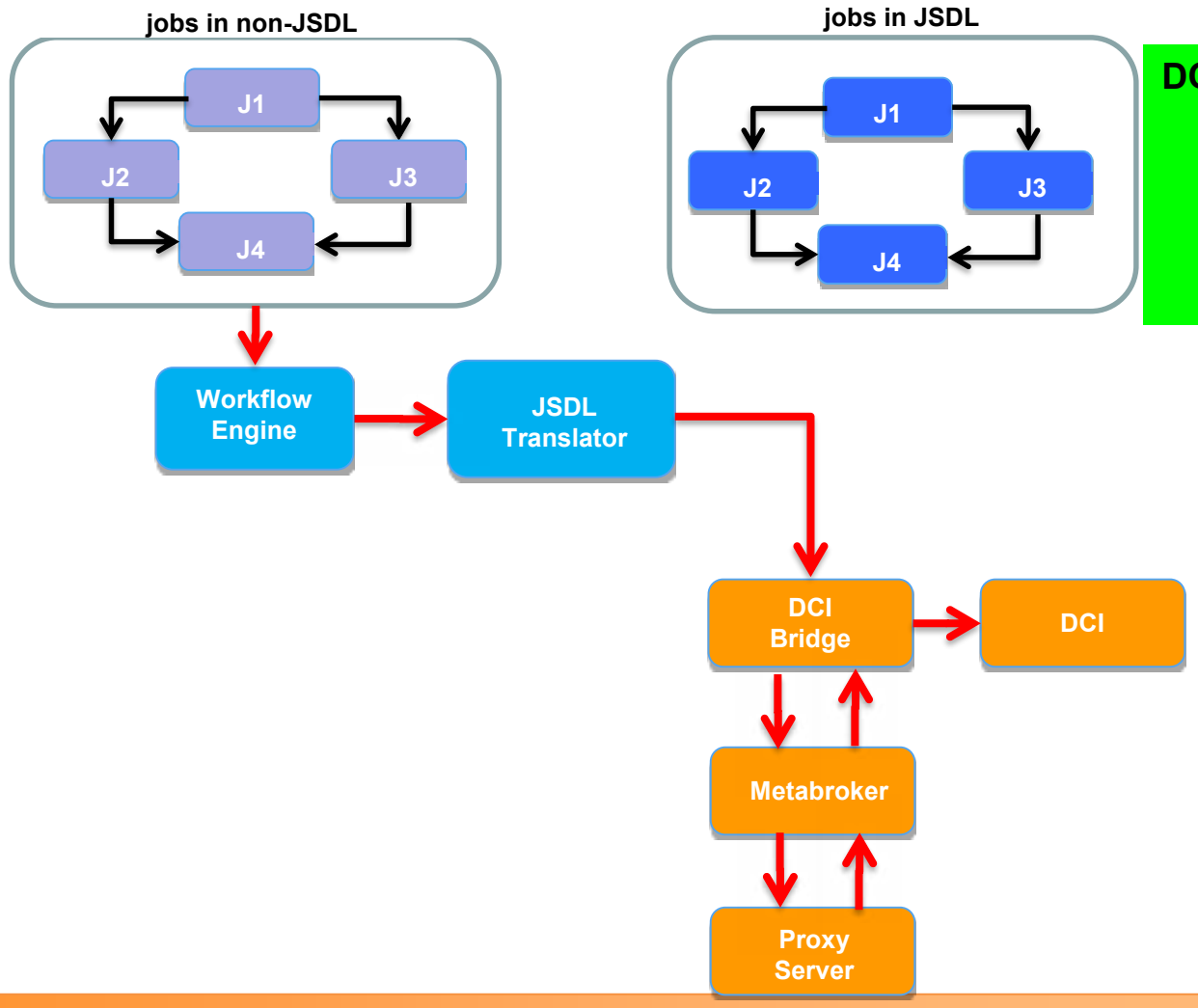
Workflow Interoperability: CGI Usage Scenario (2)



Workflow Interoperability: CGI Usage Scenario + Taverna WF

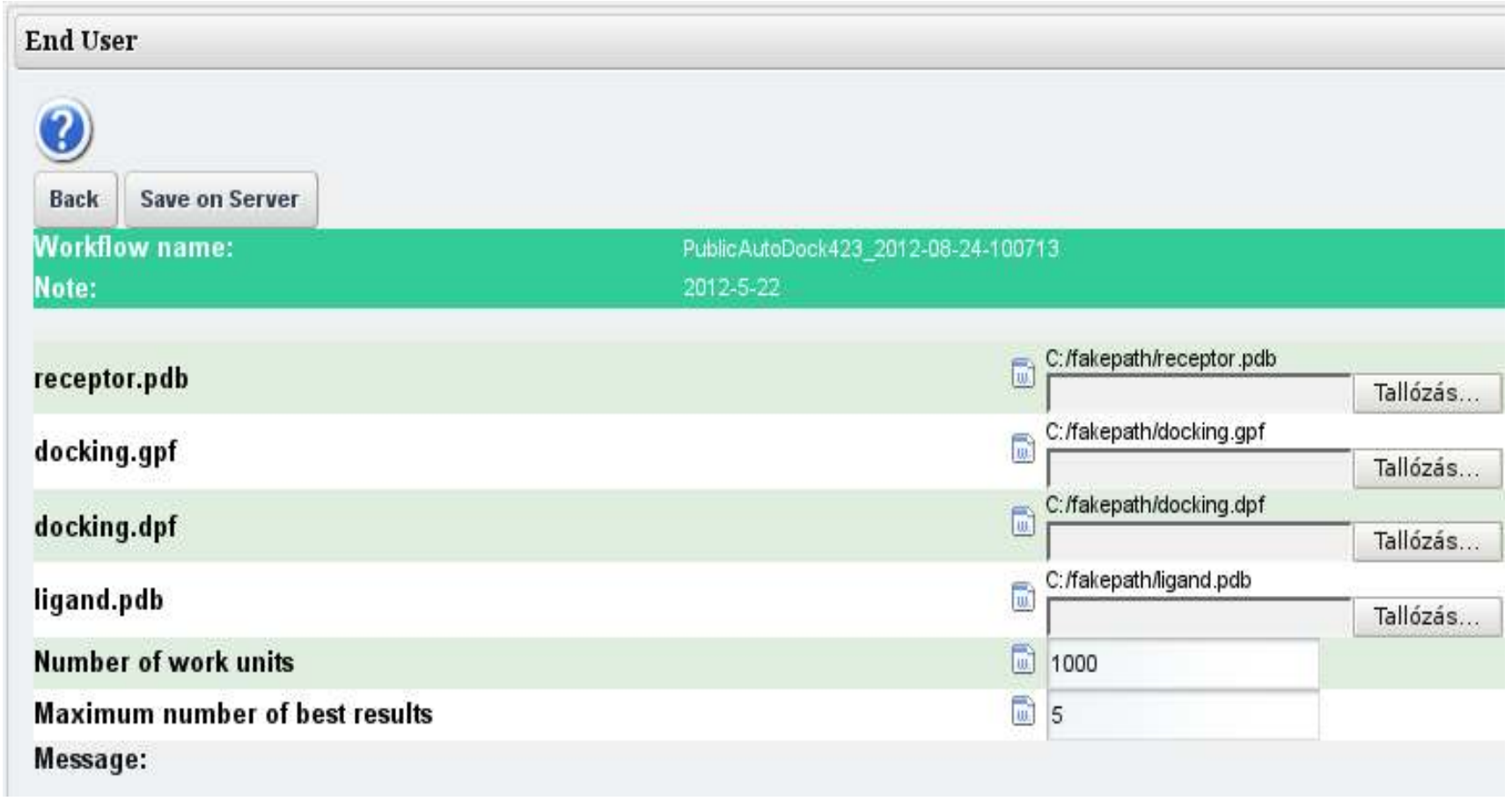


Distributed Computing Infrastructure Interoperability

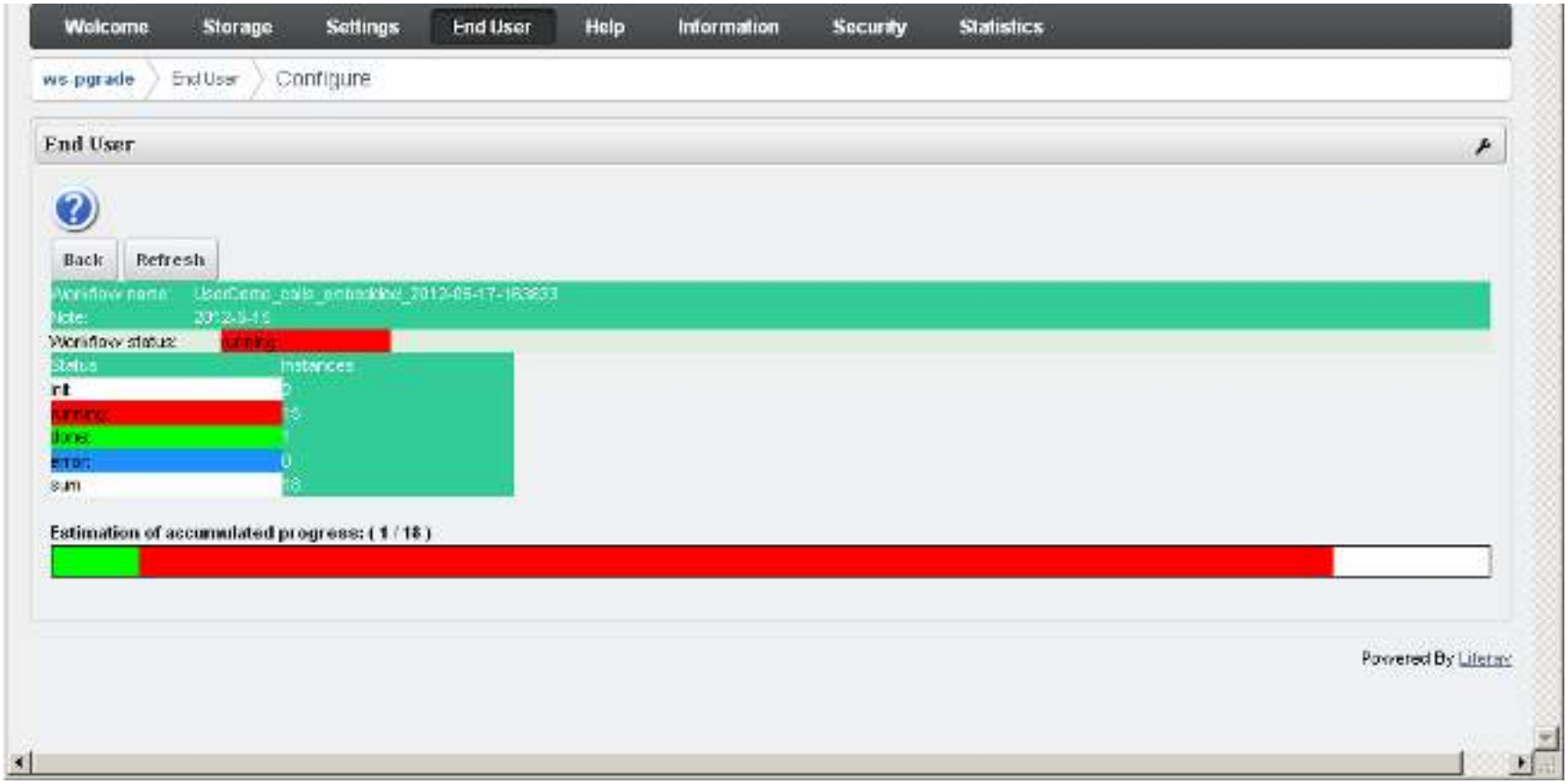


DCI support:
- cloud
- cluster
- desktop grid
- service grid
- supercomputer

Usage Scenario: Domain Researcher View



Usage Scenario: Domain Researcher View

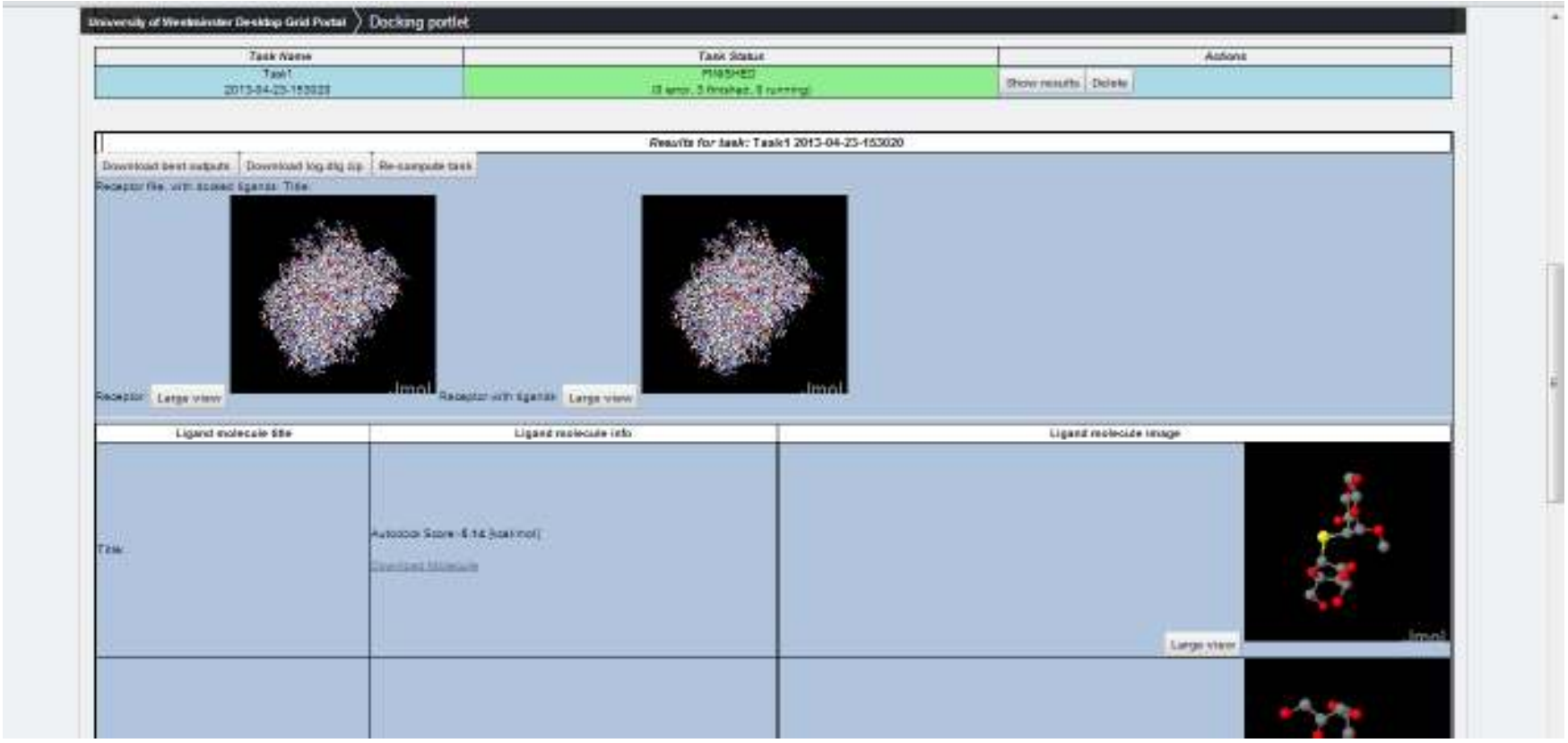


e-infrastructure

Usage Scenario: Customised View

e-infrastructure

Usage Scenario: Customised View



e-infrastructure

Knowledge Transfer and Research Community Support

Academic communities

Astrophysics	PGRADE + Taverna
Computational Chemistry	Galaxy + PGRADE + UNICORE
Heliophysics	PGRADE + Taverna
Hydrometeorology	PGRADE
Life Sciences	Galaxy + Moteur + PGRADE + Taverna
Meteorology	PGRADE
Material Sciences	PGRADE
Particle Physics	PGRADE
Seizmology	Dispel4Py + PGRADE

Non-academic communities

Engineering and manufacturing SMEs

- Business Process Simulation
- Discrete Event Simulation
- Fluid Dynamics Simulation

PGRADE

Creating and Executing Workflows

workflows in the repository in 2013

abstract	- 123
concrete	- 119
total	- 242

workflows in the repository in 2016

abstract	- 213
concrete	- 385
total	- 598

workflow execution number

SHIWA Simulation Platform

- 512 (dev) / 331 (test) / 181 (training)

Community gateways

Astro workflows	- 182 (dev) / 73 (test) 203 (prod)
Compchem workflows	- 550 (dyn) / 400 (dock) / 300 (quan)
Helio workflows	- 41 (prod)
Life Sciences workflows	- 79 (dev) / 325 (prod)

Research Community Support

Phase 1 – introduction to the workflow technology

Platform: SHIWA Simulation Platform

Support: workflow creation and execution

Training: platform and workflow training

Phase 2 – creating and running workflows

Platform: SHIWA Simulation Platform or community portal

Support: portal deployment + workflow porting

Training: gateway deployment and management

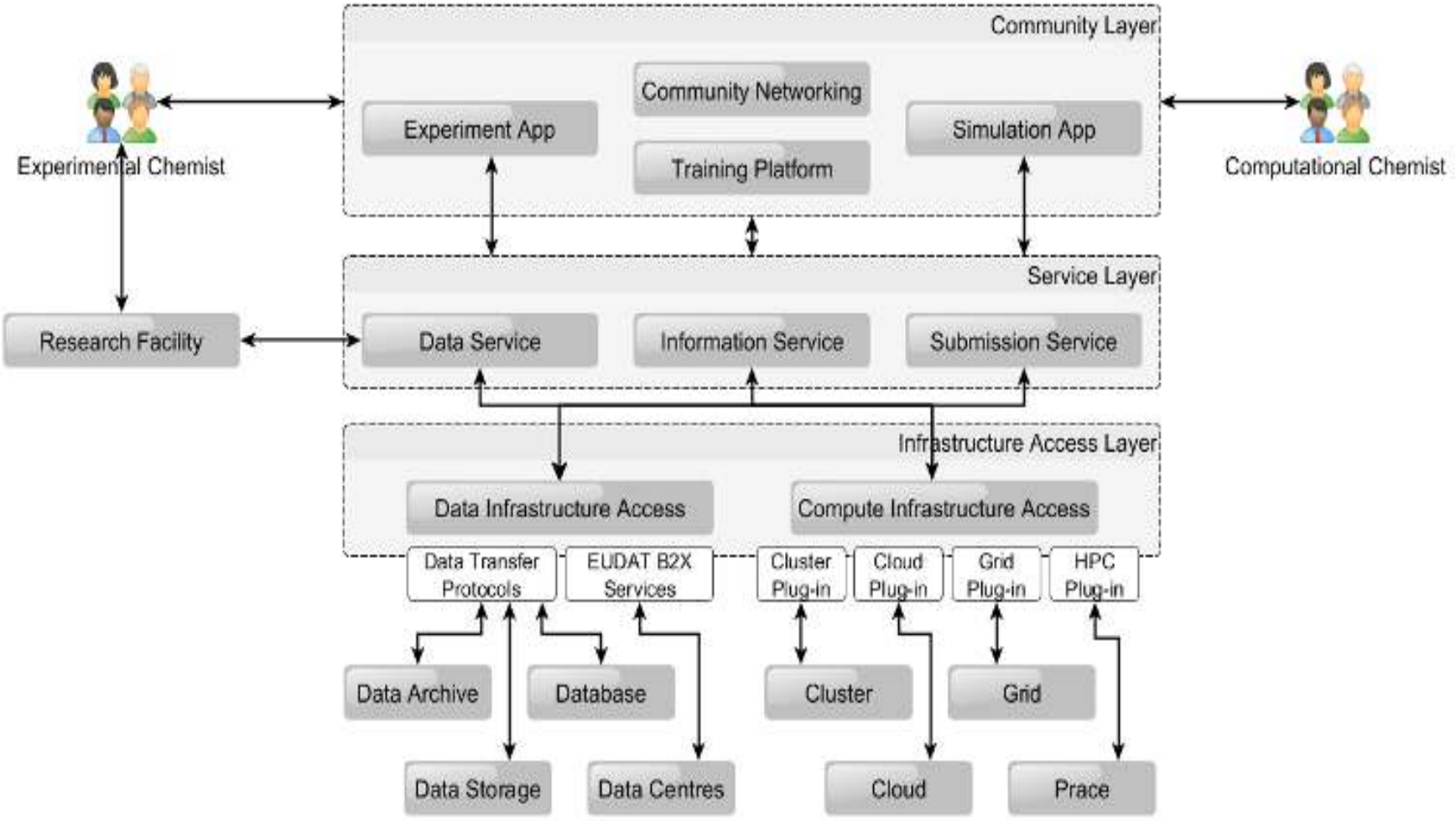
Phase 3 – combining workflows of different workflow systems

Platform: SHIWA Simulation Platform + community portal

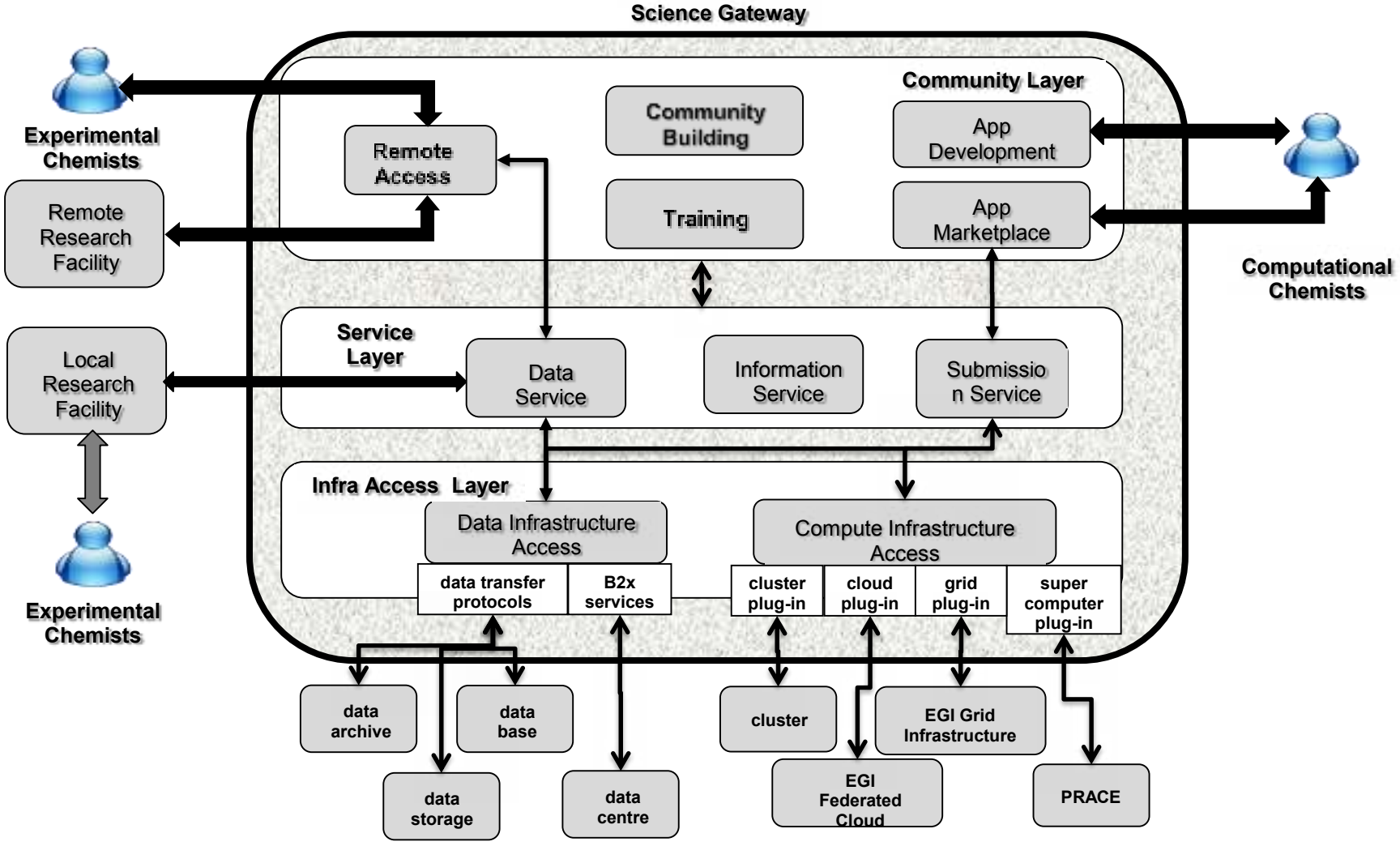
Support: access to repository + submission service

Training: CGI training

Architecture of the Research Infrastructure



Access to the Research Infrastructure



Access to the Research Infrastructure

community layer

- will offer social media type services allowing Experimental Chemists to run experiments on remotely available research facilities.
- will provide to access to simulation applications to run simulations for Computational Chemists.
- will also support training activities and community building.

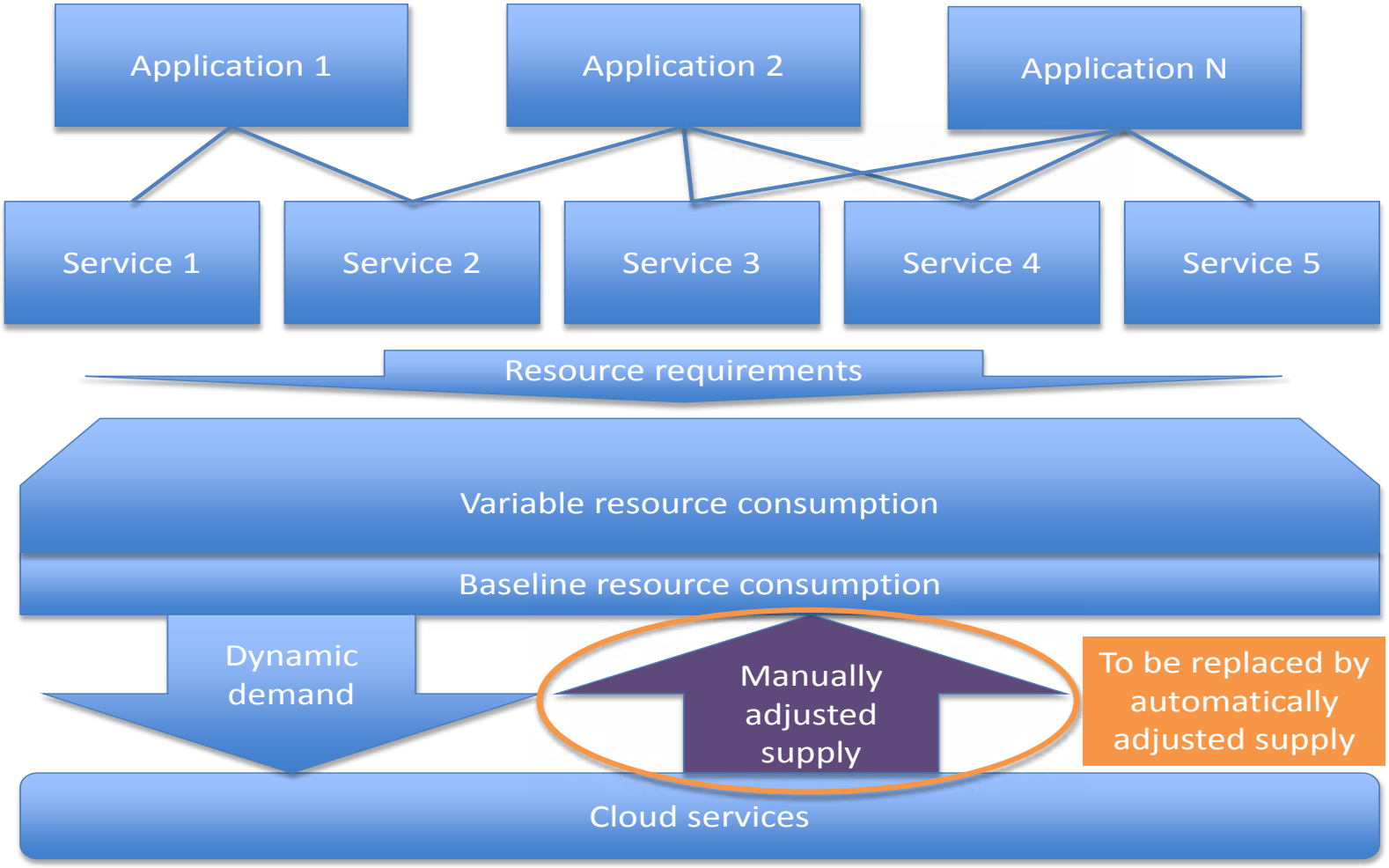
service layer

- will connect researchers to the research facilities and e-infrastructure resources using microservices managed by a service orchestrator.
- will provide data service that will connect Experimental and Computational Chemists through scientific data
- Experimental Chemists will use the data service to manage experimental data while Computational Chemists will run simulations through the submission service using the data service.

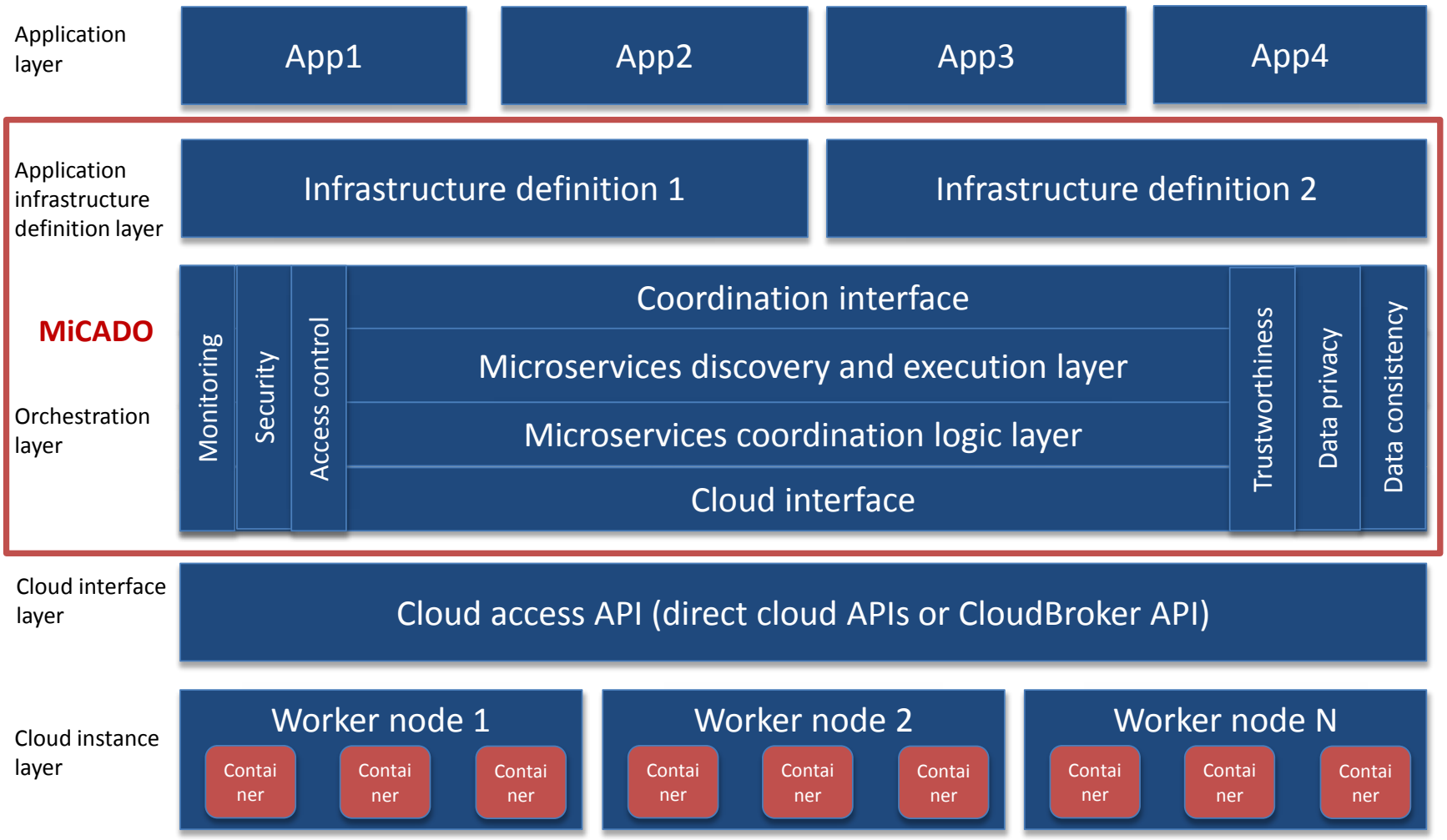
infrastructure access layer

- computing infrastructure access service - will manage access to major computing resources such as cloud, cluster, grid and supercomputer.
- data infrastructure access service - will manage data using different types of data resources, such as data archives, databases, data collections, data storages using EUDAT B2xx and MASi services, and major data transfer protocols

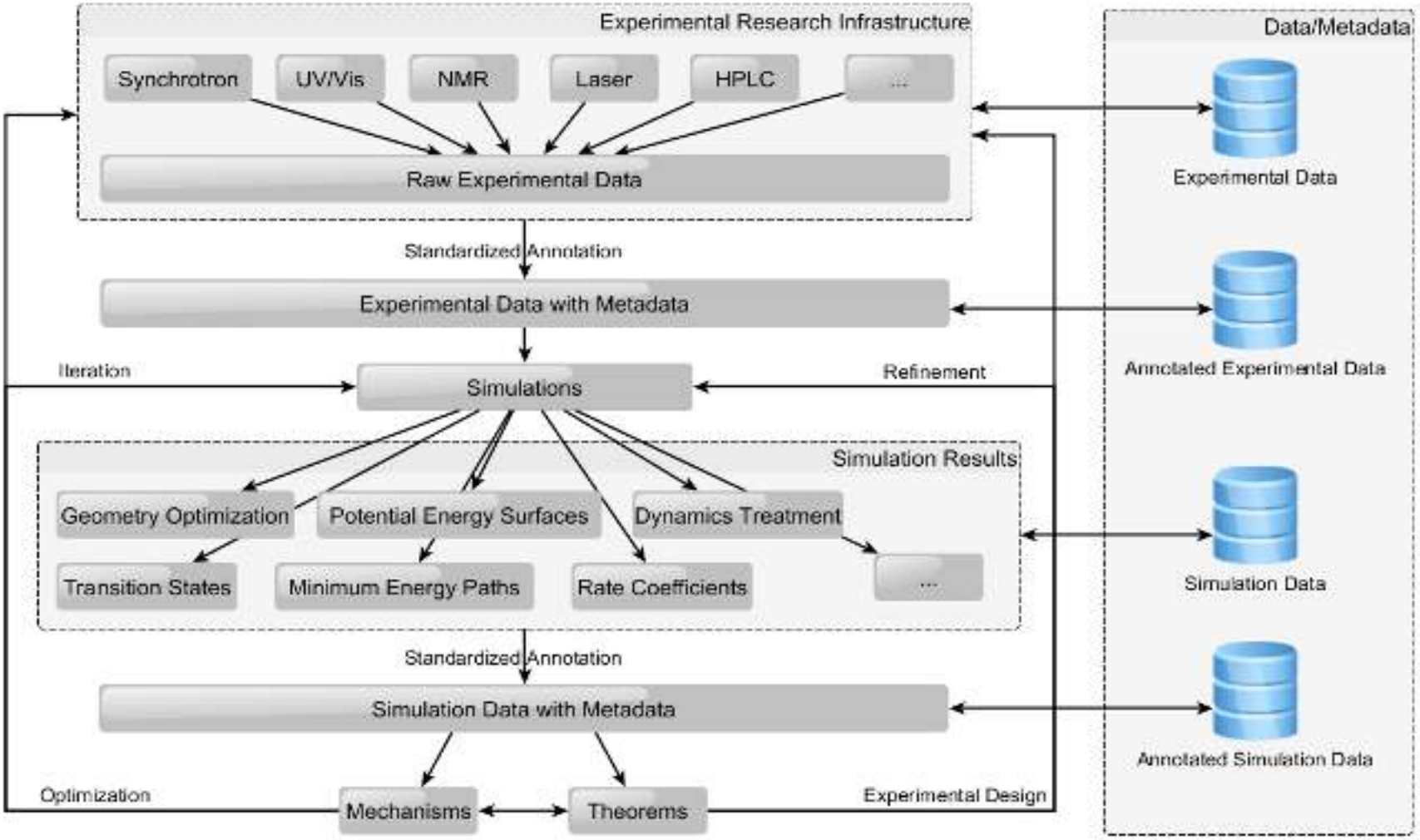
Access to Cloud Resources in the Research Infrastructure (1)



Access to Cloud Resources in the Research Infrastructure (2)



Data Flows in the Research Infrastructure



e-infrastructure

Acknowledgements

Prof. Sonja Herres-Pawlis

Dr. Jens Kruger

Questions?