Virtual Earthquake and seismology Research Community e-science environment in Europe
Project 283543 – FP7-INFRASTRUCTURES-2011-2 – www.verce.eu – info@verce.eu

The VERCE Science Gateway: Webinar on data-intensive processing with dispel4py

(dispel4py training)

15-16 October 2014

What is a workflow?

Why use a workflow?

There are many workflow languages - why invent dispel?

What is dispel4py good for?

Malcolm Atkinson

What is a workflow?

A composition of steps to make a data-handling + data-analysis+simulation journey

Many ways of forming steps

- Require good libraries of ready made steps
- Learn to add your own

Many ways of combining steps

Running in many computing environments

Recursive — a journey can be a step in another journey

Why use a workflow?

Rapid prototyping and experiment
Saving you labour and repeated drudgery
Reducing error rates
Saving you from doing your own housekeeping
  Returning resources such as file space
  Gathering all your results
Acceleration due to workflow optimisation, e.g. parallelisation
Sharing & getting credit for methods
Incrementally improving methods
Combining methods developed by different experts

There are many workflow languages - why invent dispel?

Raising the level of discourse
- Removing much technology specific information - technology changes
- Relieving users from concerns about optimisation

Improving the logical description
- Streams of data with auto-iteration over data units
- Multiple streams in & multiple streams out
- Behaviour, data interpretation & data representation

Covering existing models
- Distributed query
  - Optimisation based on avoiding IO & characterising operators
- Real-time processing
- Task-based batch processing

Achieving scalability

What is dispel4py good for?

That is what you will learn today and tomorrow in the context of seismology

Everything ….

but investment in libraries is needed for each new topic

plus common libraries for shared activities, such as data handling

Everything ….

but the dispel4py engineering team need to

make it perform at the scales you need

make it excel on the DCIs you use

- laptop to cloud via supercomputers & clusters

make it reliable

So I will hand you over to their tender mercies

Wednesday 15/10

Please note that all times are in BST

Session 1		
9:00-9.15	Tea & Microphones	Tom Garth
9:15-9:30	Welcome and Introduction	Malcolm Atkinson
9:30-11:00	Introduction to dispel4py	Rosa Filgueira
Session 2		
13:00-13:10	Review of progress	Rosa Filgueira
13:10-13:20	Experience of dispel4py in Seismology	Federica Magnoni
13:20-14:20	dispel4py functions: chaining and adding PEs	Amy Krause
14:20-14:50	Writing PEs	Amy Krause

Thursday 16/10

Session 3		
9:00-9.10	Tea & Microphones	Tom Garth
9:10-9:20	Review of progress	Amy Krause
9:20-9:50	The dispel4py registry: selecting and using PEs	Iraklis Klampanos
9:50-10:10	X-correlation	Amy Krause and Rosa Filgueira
Session 4		
10:10-10:25	Introduction to the provenance data model	Alessandro Spinuso
10:25-10:35	Provenance-powered PEs	Alessandro Spinuso
10:35-10:45	Accessing provenance data	Alessandro Spinuso
Closing		
10:45-11:00	Plans for developing dispel4py further	Amy Krause et. al.