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

PE Registration in VERCE

Iraklis A. Klampanos
Data-Intensive Research Group
University of Edinburgh

VERCE @ University of Liverpool, 3 September 2012

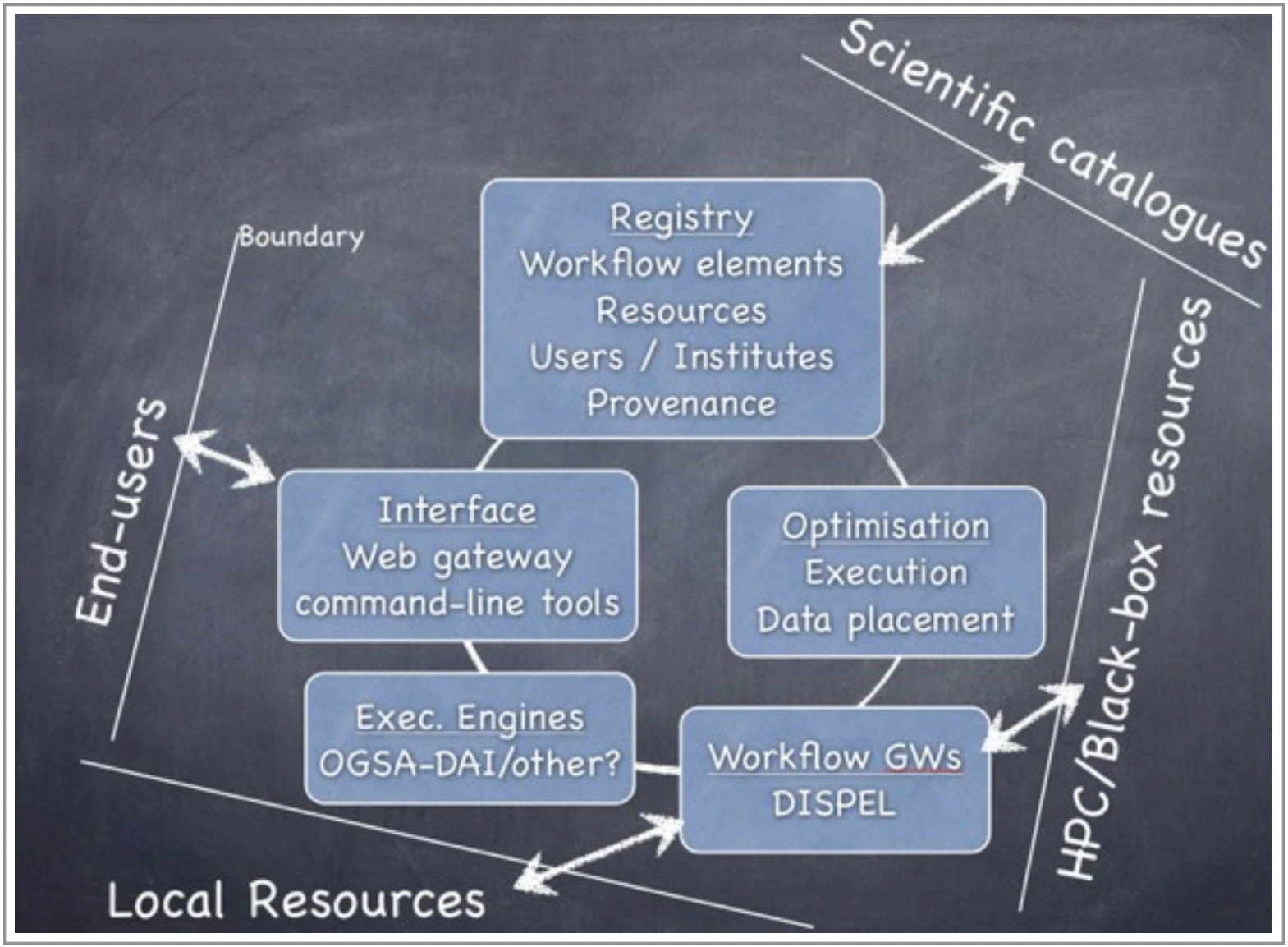
Overview

- Rationale
- Bird's eye view
  - Maintaining PEs, catalogues, provenance, users
- Technologies
  - Relational DBs
  - Linked data stores
  - Triple stores / RDF
  - *Hybrid*
- Current version and how to use for the exercises
- Steps to be completed within the 2nd year

Rationale

- Versioning, provenance and attribution
- Coordination of remote components
  - Consistent view of the world
- Hints to execution engines
  - Store, execute, backup and deliver optimally
- VERCE scientific gateway
  - Interfacing with scientists and other users

Components


Scientific catalogues
Boundary
Registry
Workflow elements
Resources
Users / Institutes
Provenance
End-users
Interface
Web gateway
command-line tools
Optimisation
Execution
Data placement
HPC/Black-box resources
Exec. Engines
OGSA-DAI/other?
Workflow GWs
DISPEL
Local Resources

Technologies

- Relational DBs - *e.g. MySQL*
  - Very mature and efficient
  - Widely supported
  - Strongly typed, strict schemas
- Linked data / Column stores - *e.g. Cassandra*
  - Networks of entities
  - Weakly typed, flexible
- Triple stores - *e.g. Apache Jena*
  - RDF-friendly
  - Weakly typed, sort-of flexible

Technologies [2]

- Hybrid approach
- Scientific catalogues
  - primarily in RDF formats
- RDF schemas for resources, users, access policies, etc.
- Distributed, “eventually consistent” column stores may be appropriate for user-related data
- Relational consistency may be appropriate for driving data movement and computation

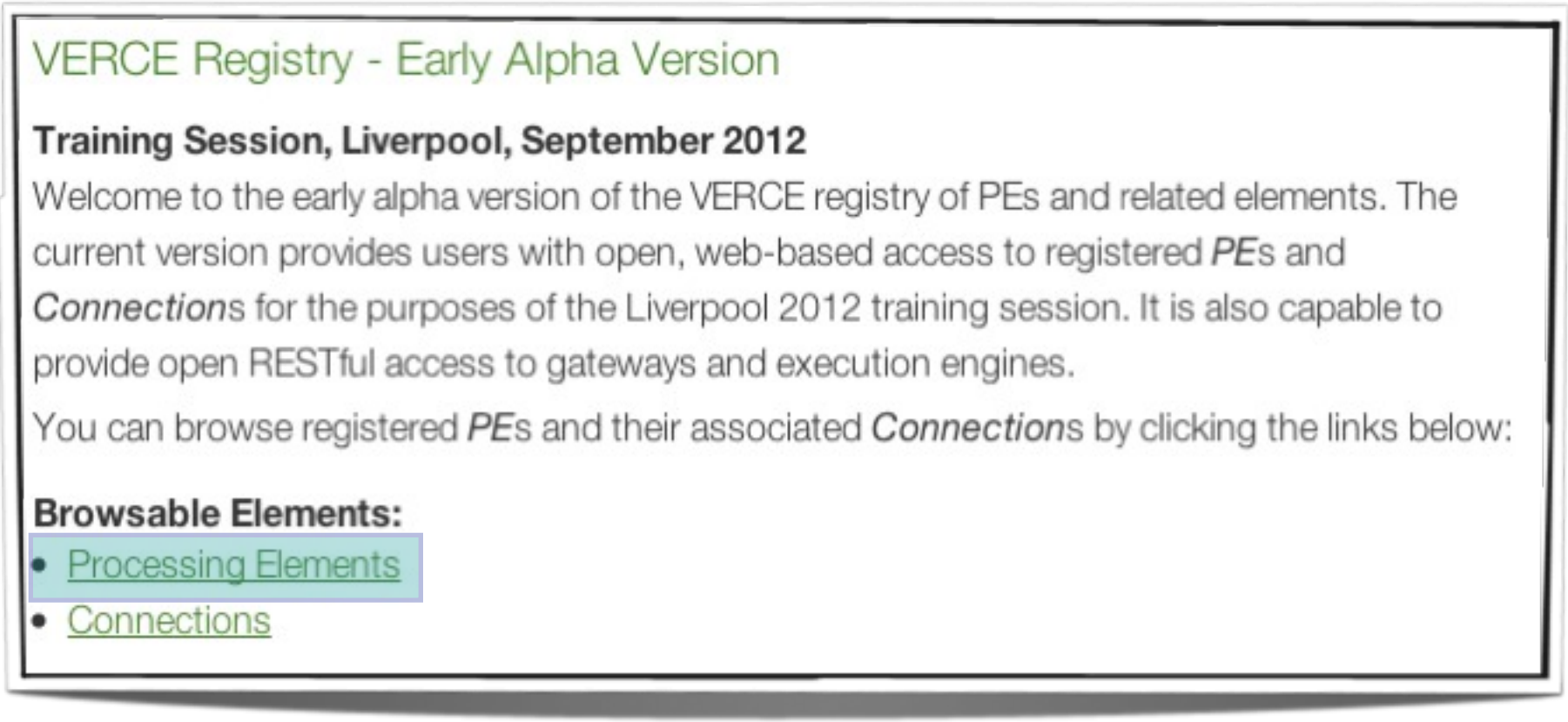
VERCE PEs Registry

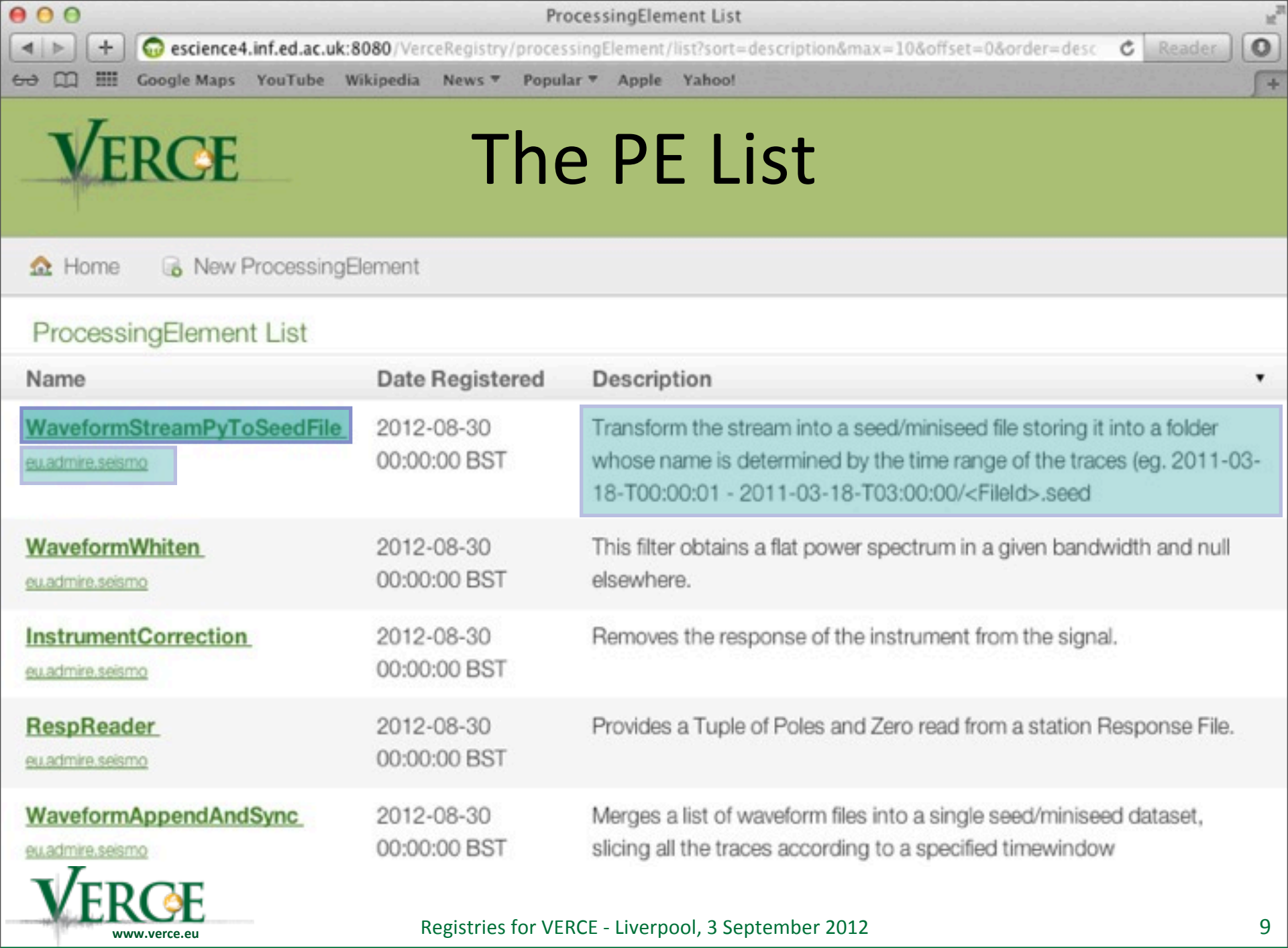
- Relational
- Restful
  - VERCE gateways (ADMIRE)
  - Execution engines (OGSA-DAI)
- Browsable
- Rough but usable
- Accessible at

  http://escience4.inf.ed.ac.uk:8080/VerceRegistry/

http://escience4.inf.ed.ac.uk:8080/VerceRegistry/

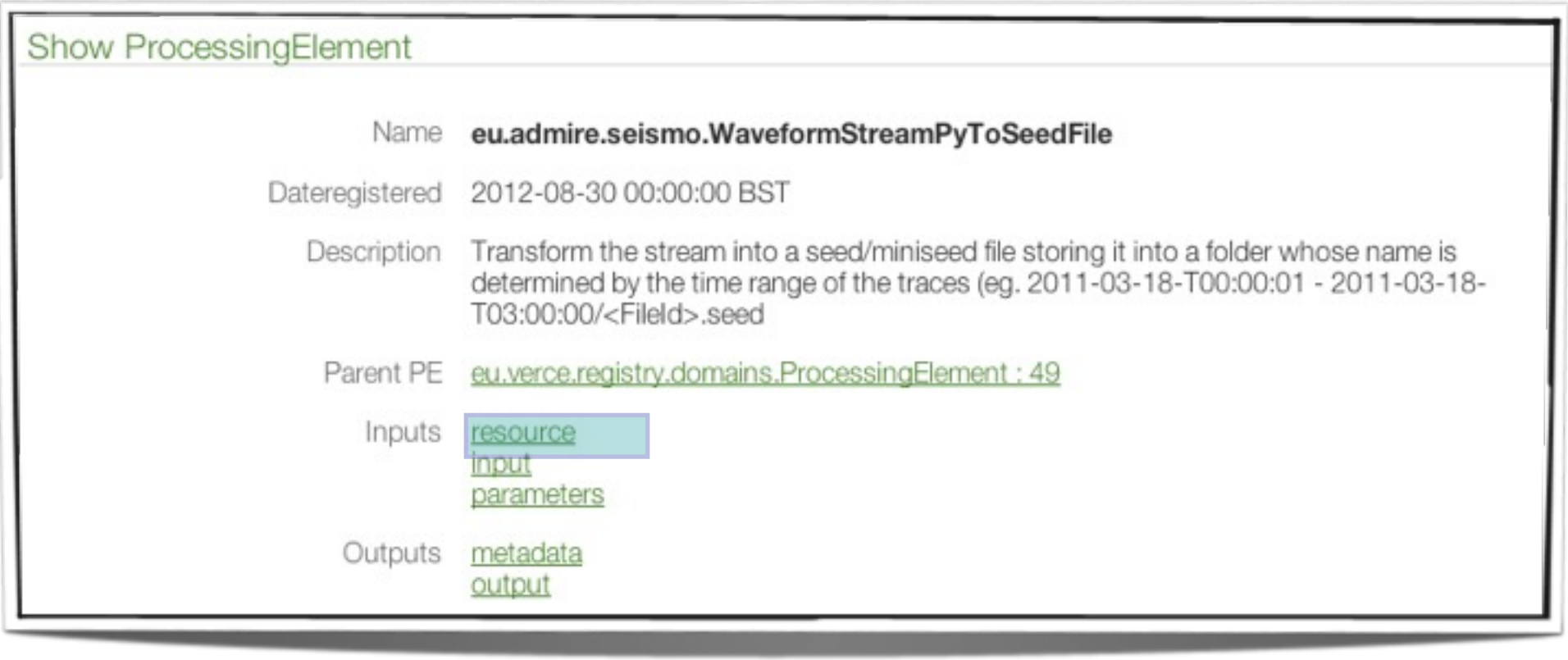
Using the Registry for the Exercises




ProcessingElement List
escience4.inf.ed.ac.uk:8080/VerceRegistry/processingElement/list?sort=description&max=10&offset=0&order=desc
Reader
Google Maps
YouTube
Wikipedia
News
Popular
Apple
Yahoo!
VERCE
Home
New ProcessingElement
ProcessingElement List
Name
Date Registered
Description
WaveformStreamPyToSeedFile
eu.admire.seismo
2012-08-30 00:00:00 BST
Transform the stream into a seed/miniseed file storing it into a folder whose name is determined by the time range of the traces (eg. 2011-03-18-T00:00:01 - 2011-03-18-T03:00:00/<FileId>.seed
WaveformWhiten
eu.admire.seismo
2012-08-30 00:00:00 BST
This filter obtains a flat power spectrum in a given bandwidth and null elsewhere.
InstrumentCorrection
eu.admire.seismo
2012-08-30 00:00:00 BST
Removes the response of the instrument from the signal.
RespReader
eu.admire.seismo
2012-08-30 00:00:00 BST
Provides a Tuple of Poles and Zero read from a station Response File.
WaveformAppendAndSync
eu.admire.seismo
2012-08-30 00:00:00 BST
Merges a list of waveform files into a single seed/miniseed dataset, slicing all the traces according to a specified timewindow

The PE List

PE Details


Show ProcessingElement
Name
eu.admire.seismo.WaveformStreamPyToSeedFile
Dateregistered
2012-08-30 00:00:00 BST
Description
Transform the stream into a seed/miniseed file storing it into a folder whose name is determined by the time range of the traces (eg. 2011-03-18-T00:00:01 - 2011-03-18-T03:00:00/<FileId>.seed
Parent PE
eu.verce.registry.domains.ProcessingElement : 49
Inputs
resource
input
parameters
Outputs
metadata
output

Connection Details

Next Steps

- Fix and secure location of service
  - likely on EDIM1
- Apply backup policies
- Identification of PEs
  - User and session-specific
- Validation and typing semantics
- Integrate with the VERCE Web gateway
- Add support for DISPEL functions