vl-e

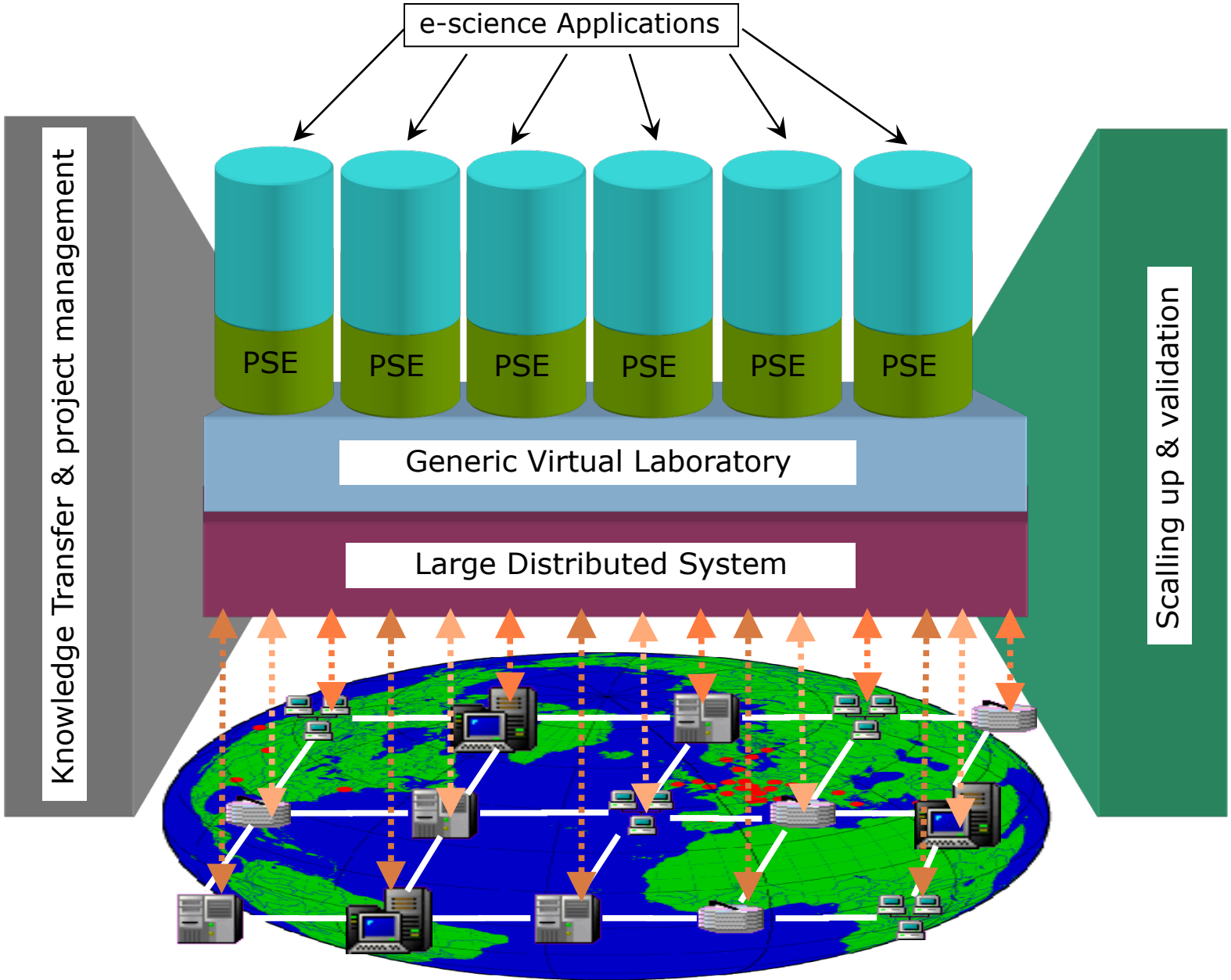
virtual laboratory for e-science

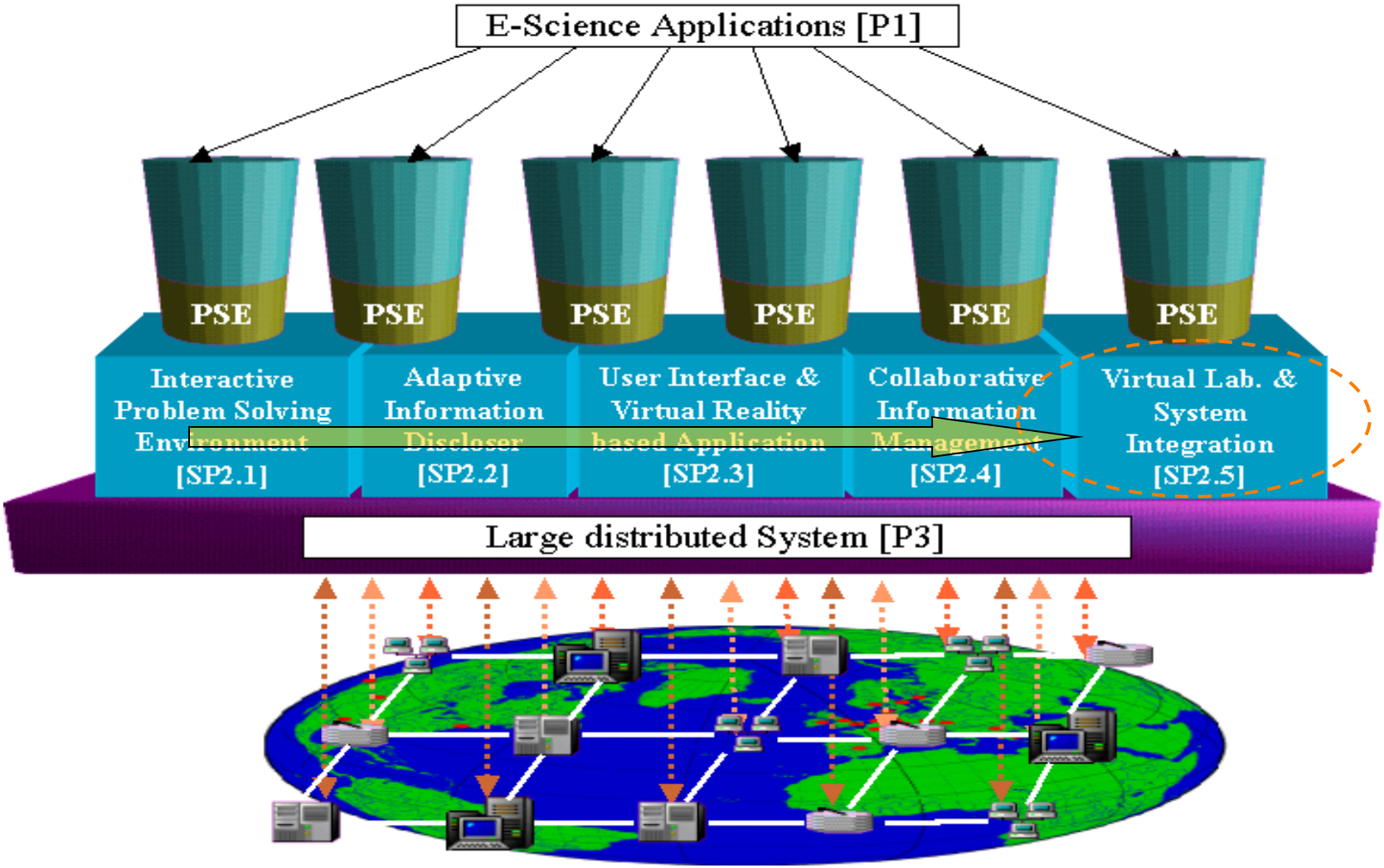
VL-e eScience Framework: from design to sharing

Adam Belloum
Institute of Informatics
University of Amsterdam

UvA

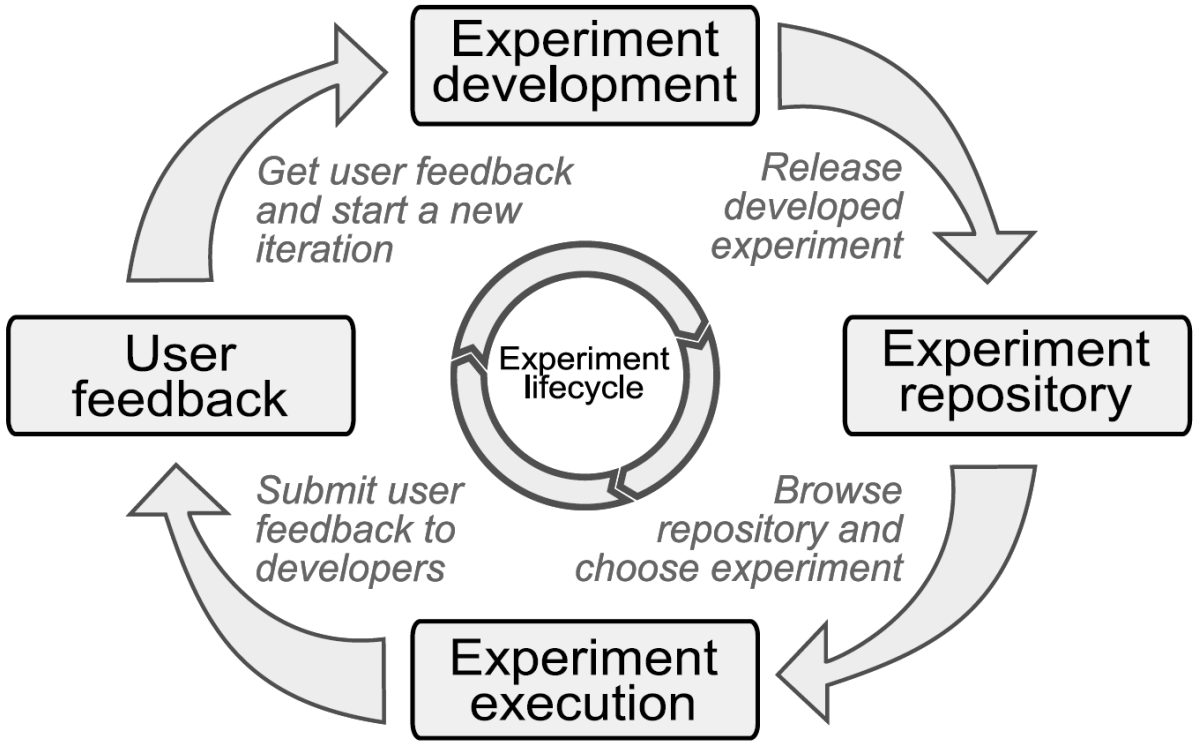
UNIVERSITEIT VAN AMSTERDAM


e-science Applications
Knowledge Transfer & project management
PSE
PSE
PSE
PSE
PSE
PSE
Generic Virtual Laboratory
Large Distributed System
Scalling up & validation

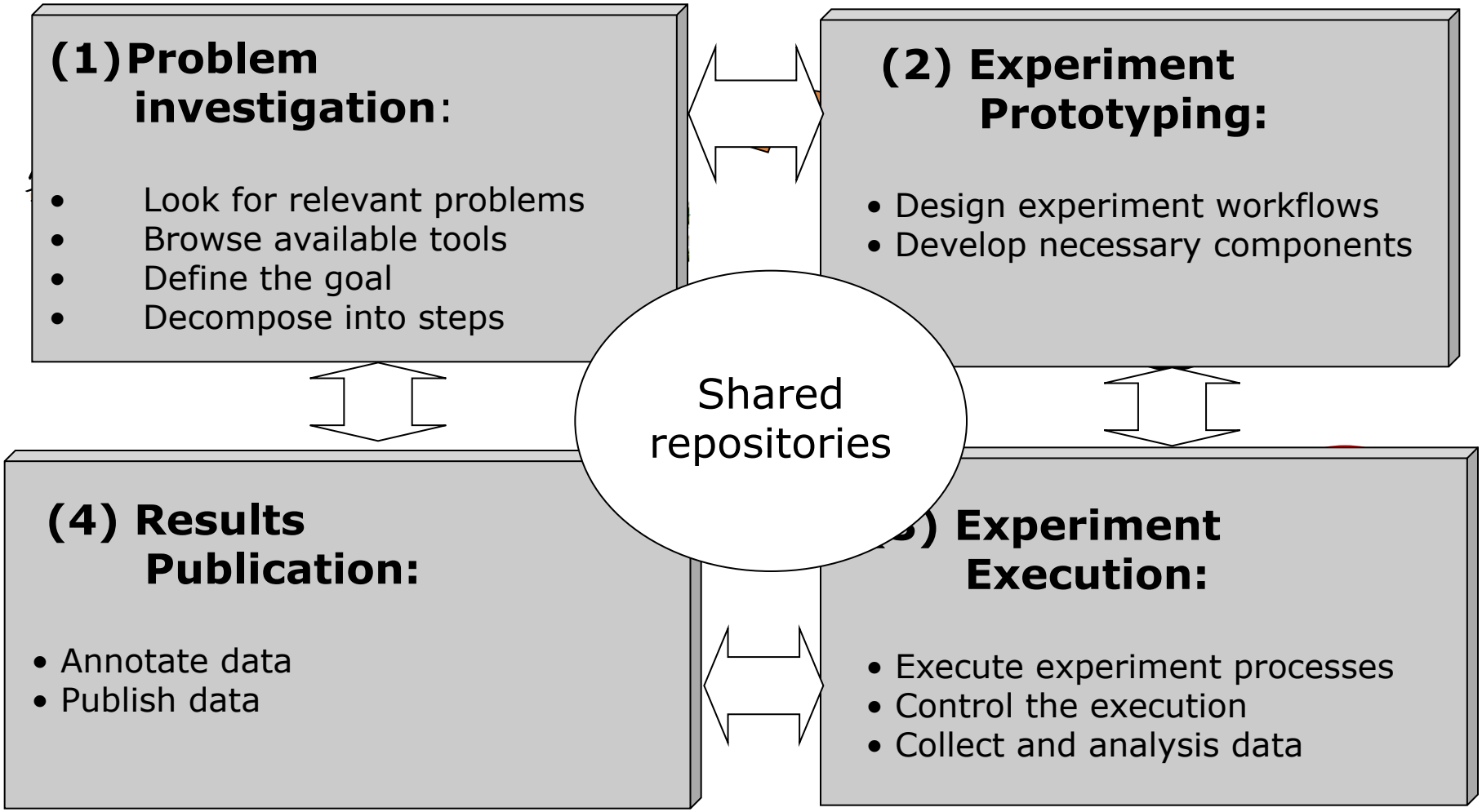

E-Science Applications [P1]
PSE
PSE
PSE
PSE
PSE
PSE
Interactive Problem Solving Environment [SP2.1]
Adaptive Information Discloser [SP2.2]
User Interface & Virtual Reality based Application [SP2.3]
Collaborative Information Management [SP2.4]
Virtual Lab. & System Integration [SP2.5]
Large distributed System [P3]

Experiment Lifecycle

- *Experiment Pipeline* - a collaborative planning and execution process that may create a new experiment



Complex Scientific experiments model



Objective

- The Objective of this work is to address the research issues related to building an e-Science Framework which will support scientists to share their knowledge and to perform specific experiments on grid resources

Keywords:

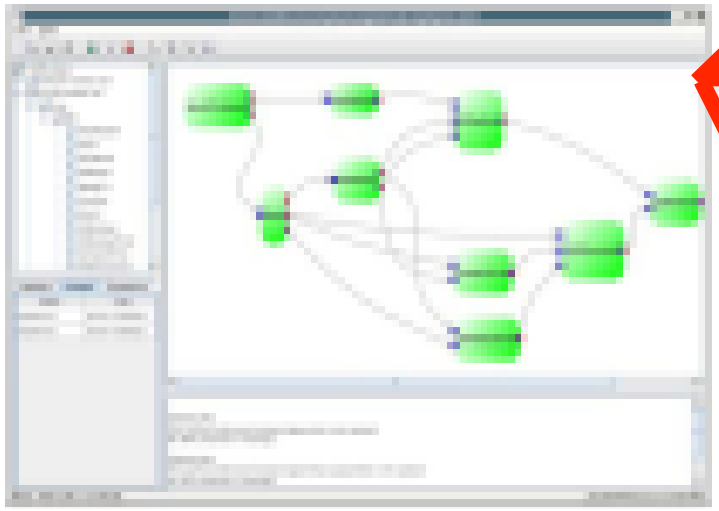
- Grid, Scientific workflow, SOA, provenance, interoperability

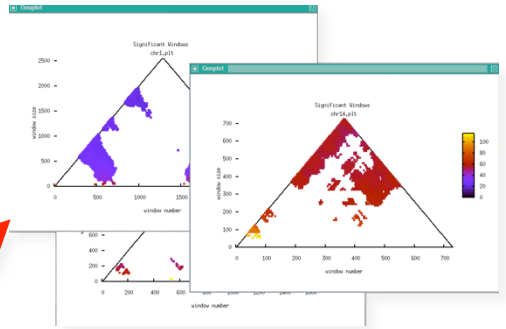
WS-VLAM e-Studio

- So far it is **simple program**, used to start the different **tools developed** in the VL-e project by the *Virtual Laboratory and system integration* group

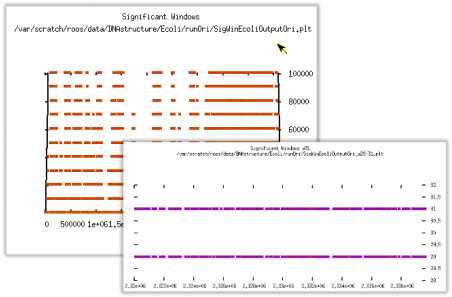
The Application Workflow …

Developed in the context of the Ridge-O-grammer Application use case by Micro-array department UvA

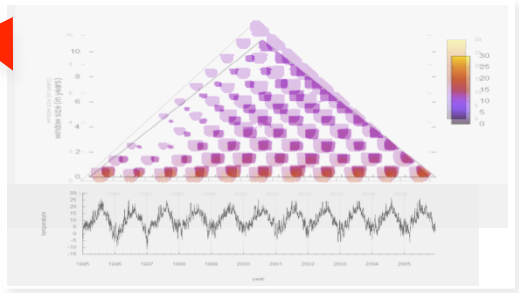




Human transcriptome map



DNA curvature of the *Escherichia coli* chromosome

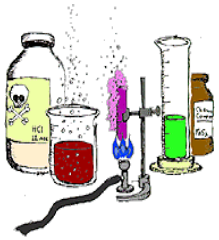


Temperature in Amsterdam

(1) Problem investigation

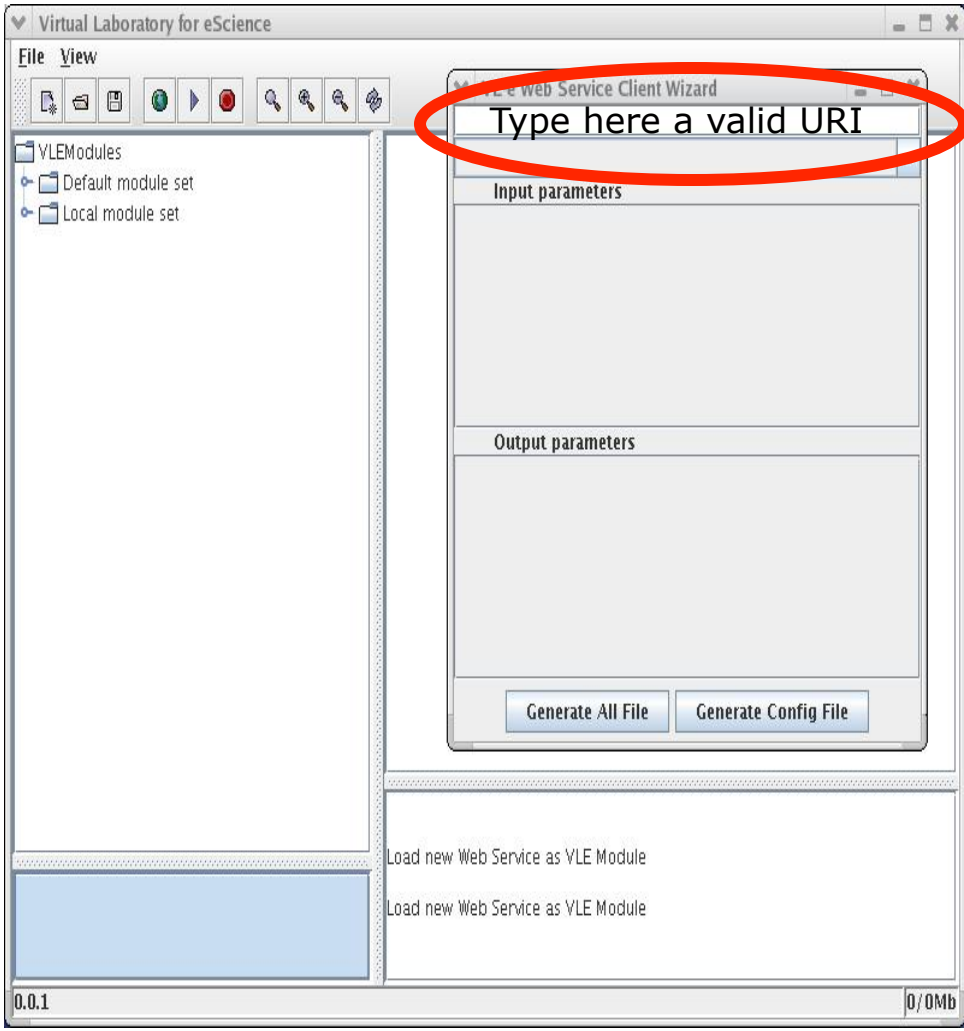
- Tools to improve **re-usability**, and **share of expertise**. These tools should allow:
  - Advanced **search capabilities**
    - HAMMER: **H**ybrid-b**A**sed **M**atch-**M**aker for **E**-Science **R**esources
    - Web Service Harvester (RPC Style WSDL)

designing the experiment



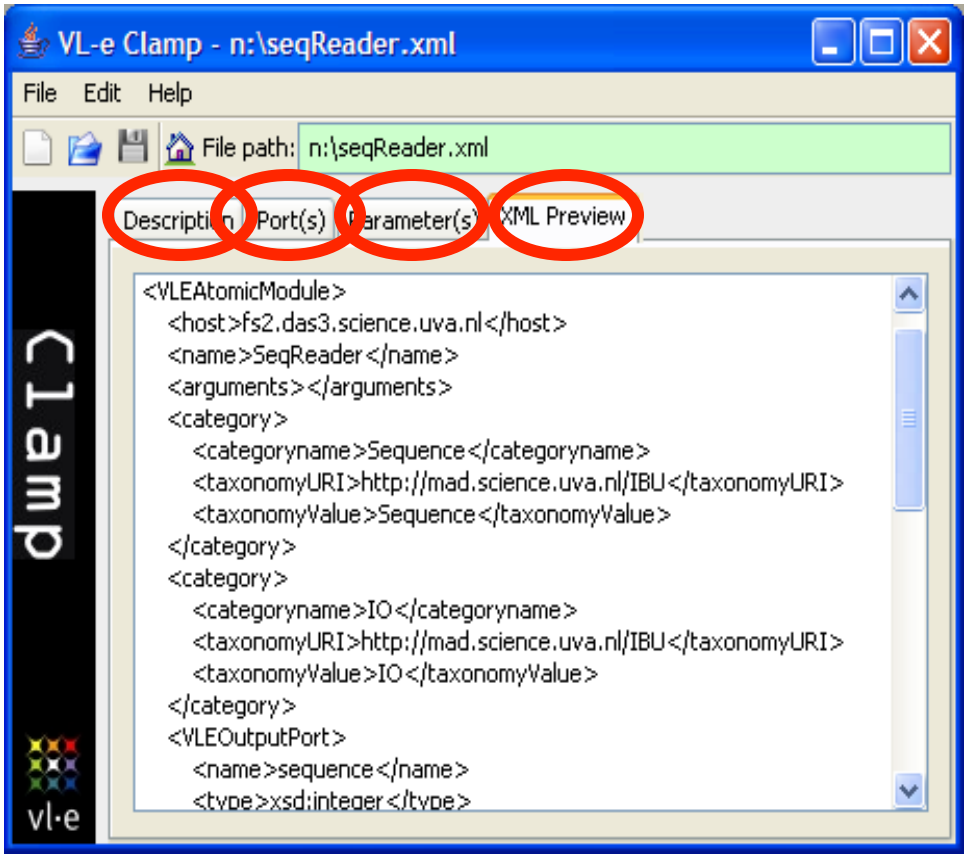
(1) Problem investigation

- Tools to improve **re-usability**, and **share of expertise**. These tools should allow:
  - Advanced **search capabilities**
    - HAMMER: **H**ybrid-b**A**sed **M**atch-**M**aker for **E**-Science **R**esources
    - Web Service Harvester (RPC Style WSDL)



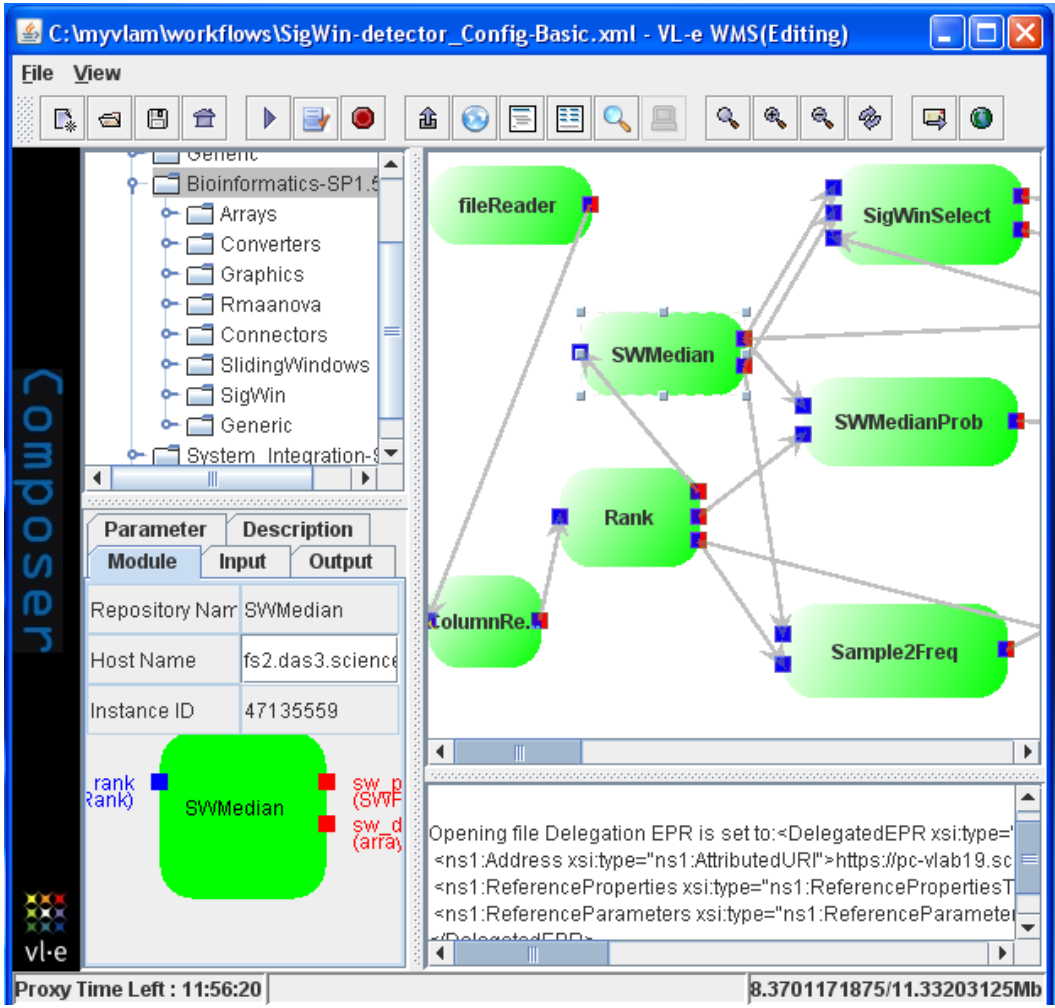
(1) Problem investigation

- Tools to design **new components**. These tools should allow:
  - **Speedup** and **simplify** the design of the new
    - CLAMP: **C**onnecting **LA**nguage for **M**odules and **P**rograms



(2) Experiment Prototyping
Grid Enabled workflow management system

- Tools to **speedup** and **simplify** building prototypes. These tools should allow:
  - **Combine** existing software components
  - **Seamless access** to computational and storage resources.

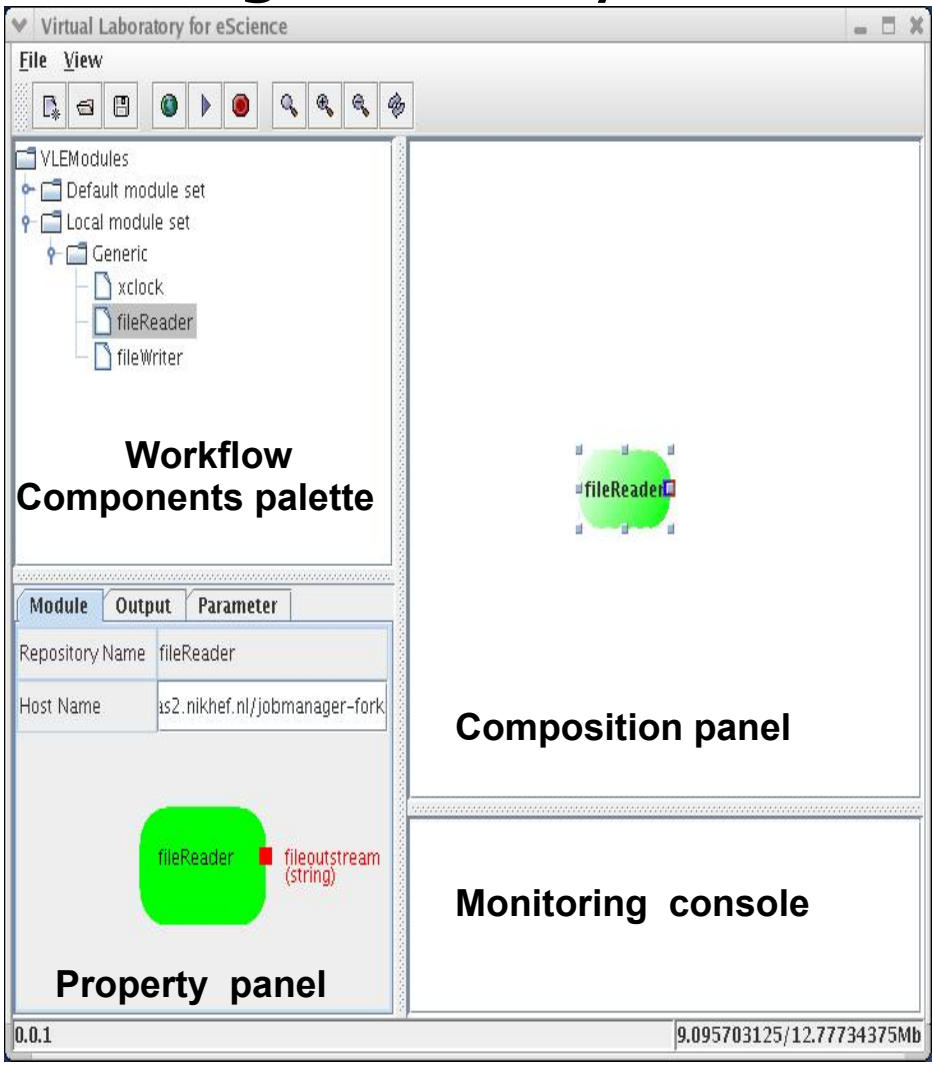


http://staff.science.uva.nl/~gvlam/wsvlam

(2) Experiment Prototyping
Grid Enabled workflow management system

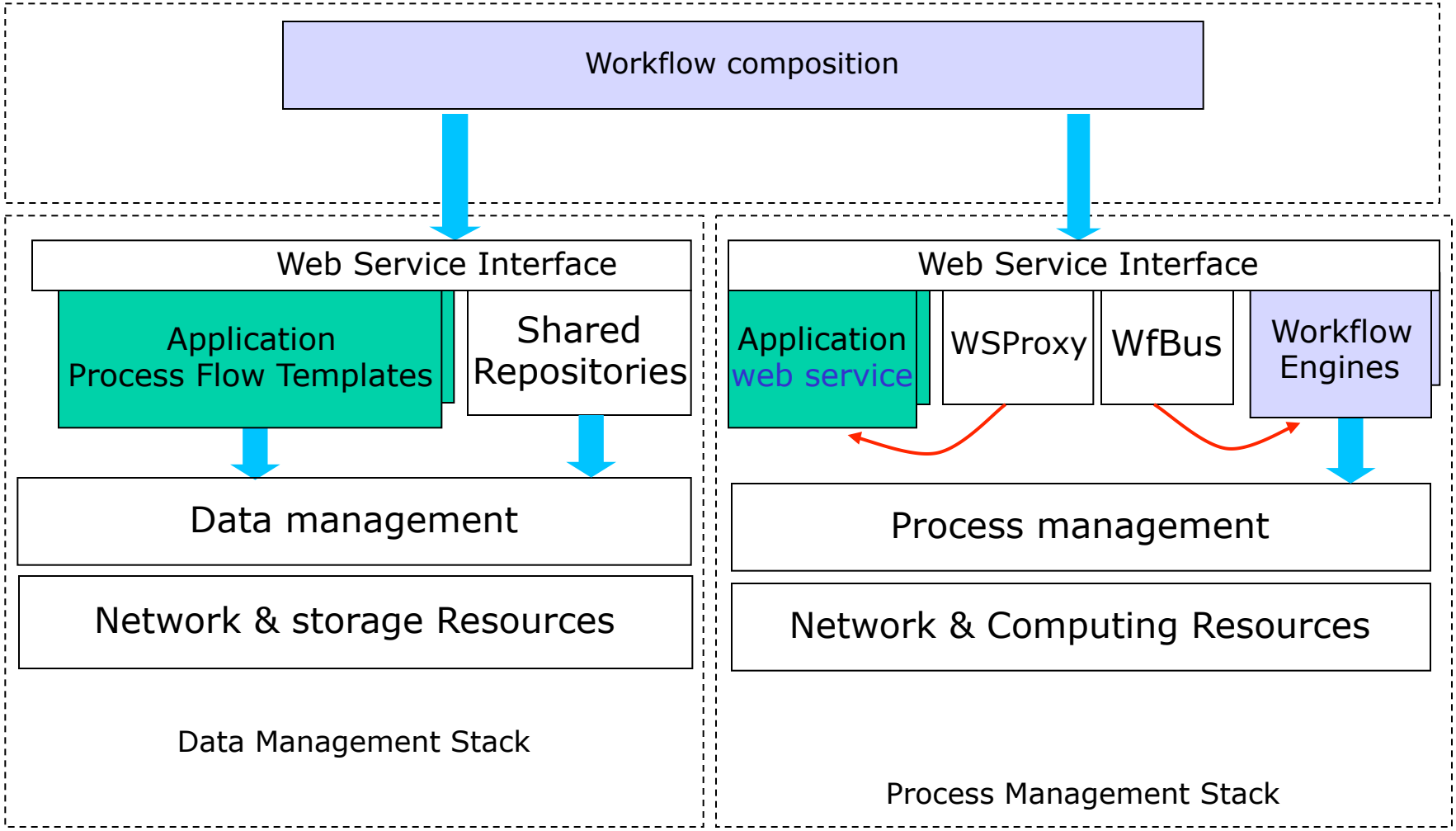
- For Application developers
  - API for JAVA, C/C++, PYTHON,
  - Data Streaming capabilities

- End-users (**scientists**)
  - Workflow monitoring
  - Drag-and-drop composition
  - Hierarchical workflow creation
  - User in the loop
  - Detach/re-attach capability



Tutorials for both developers and End-users are available at
http://www.science.uva.nl/~gvlam/wsvlam/Documentation

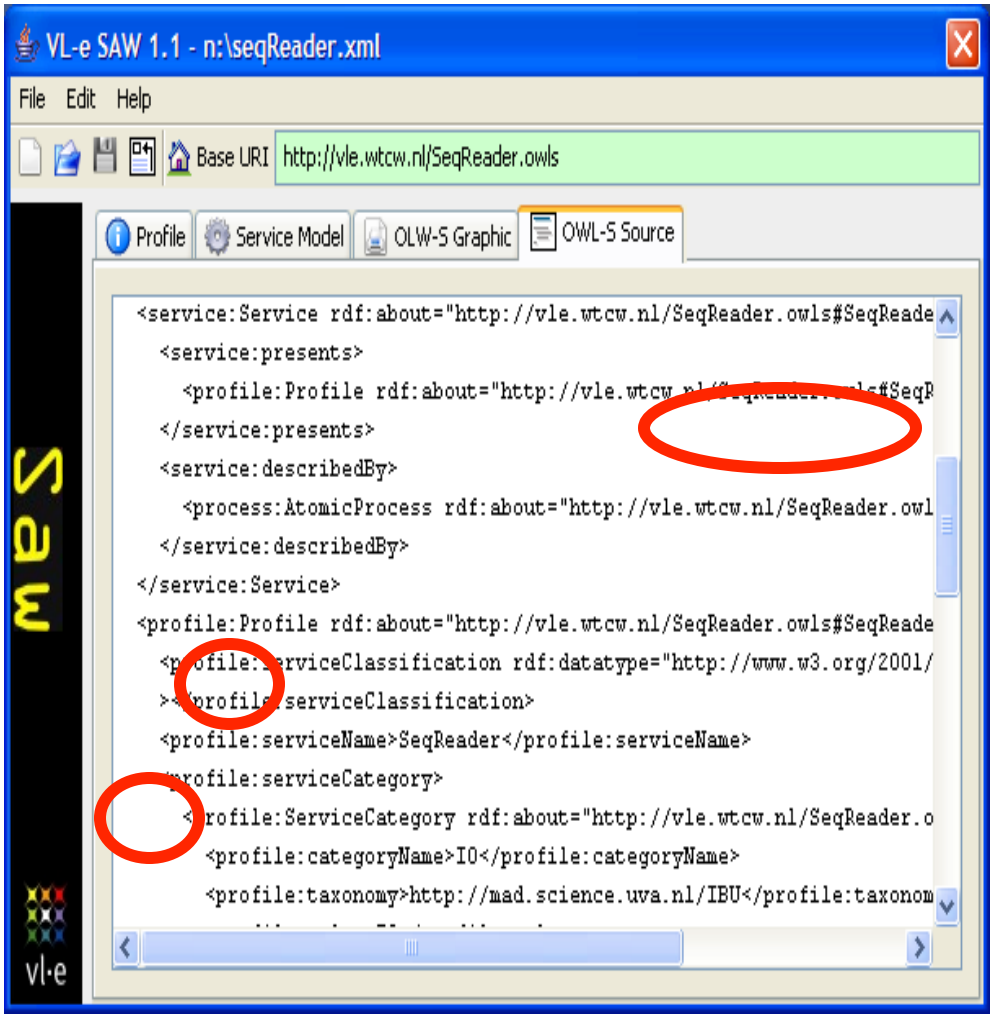
A WSRF enabled workflow engine



Bob Hertberger keynote talk at *2nd IEEE Conf on eScience & grid computing , Amsterdam 2006*

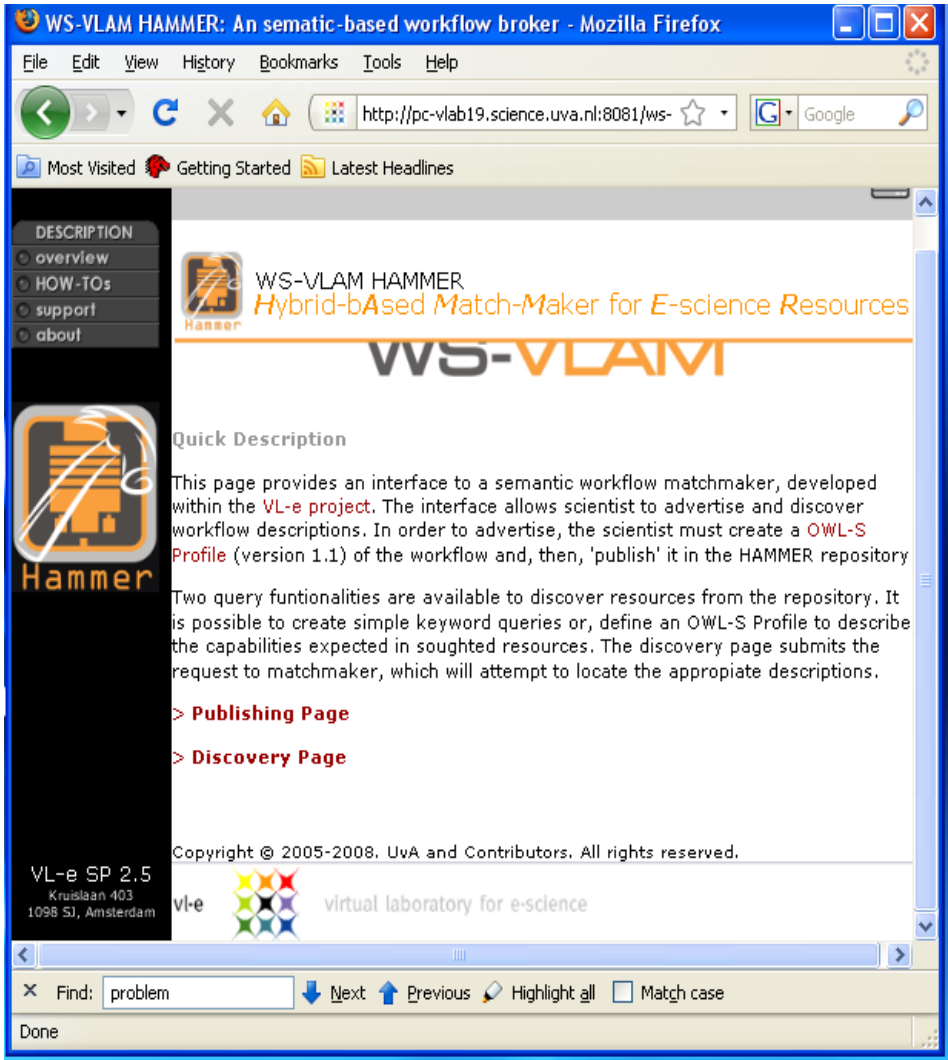
(2) Experiment Prototyping Semantic tools

- Tools to improve **sharing** and **re-usability**. These tools should allow:
  - **Annotate** and **publish** the designed experiment prototype
    - SAW: Semantic Annotation for Workflow and workflow components
    - HAMMER: **H**ybrid-b**A**sed **M**atch-**M**aker for **E**-Science Resources



(2) Experiment Prototyping Publishing tools

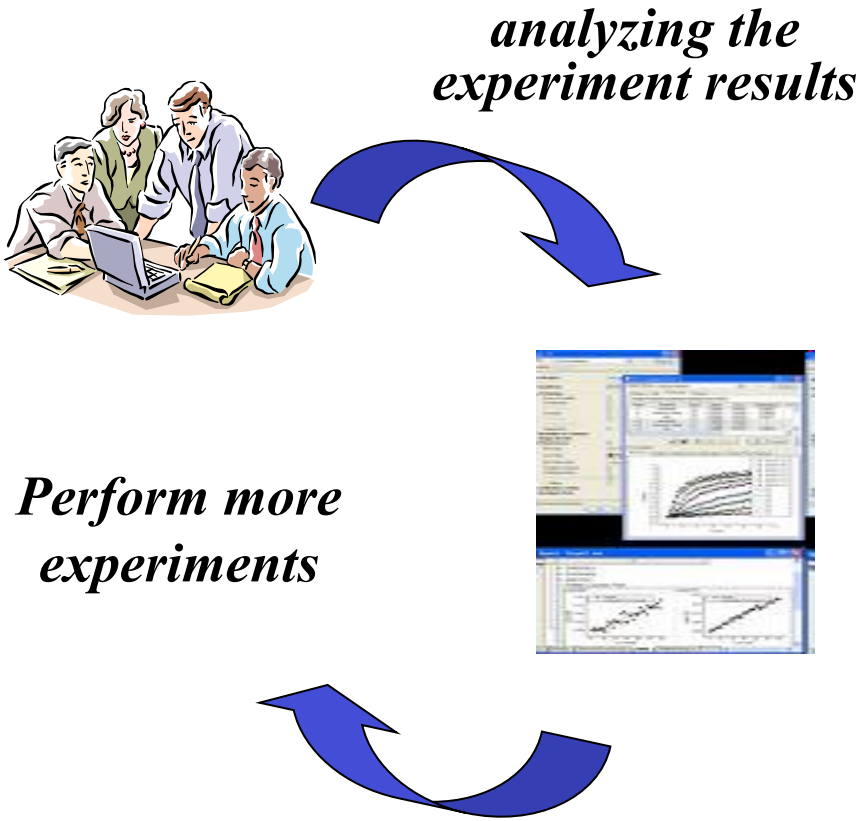
- Tools to improve **sharing** and **re-usability**. These tools should allow:
  - **Annotate** and **publish** the designed experiment prototype
    - SAW: Semantic Annotation for Workflow and workflow components
    - HAMMER: **H**ybrid-b**A**sed **M**atch-**M**aker for **E**-Science Resources



(3) Experiment Execution

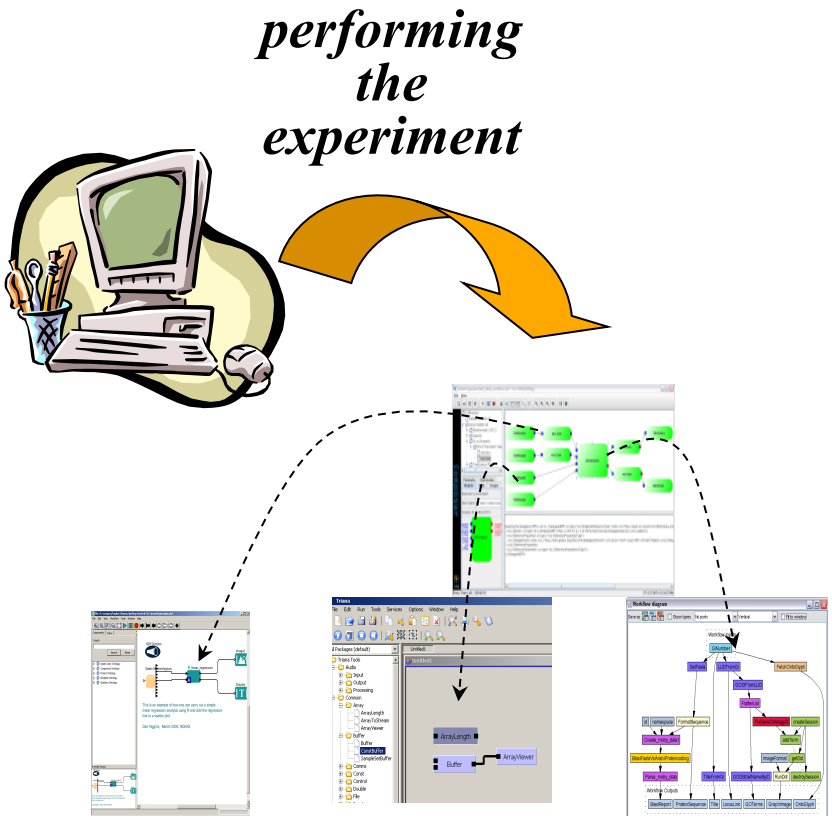
Virtual resource Browser

- Tools with **intuitive interface** allowing end users (Scientists):
  - **Find** results: access resources to manipulate data
  - **Re-start** experiment and monitor its execution



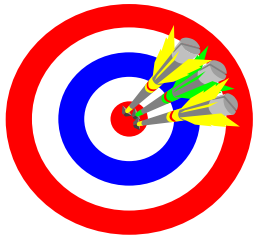
(3) Experiment Execution VLE-WFBus

- Tools to **improve interoperability** among various execution platforms.
  - Recognize different workflow descriptions.
  - Coordinate the execution of workflows, and monitor their execution.



(4) Results Publication

- Tools to improve **share of expertise**. These tools should allow:
  - **Annotate** and **publish** data, workflows, etc
    - HAMMER: **H**ybrid-b**A**sed **M**atch-**M**aker for **E**-Science Resources (at the project level)
    - MyExperiment: (at the community level)



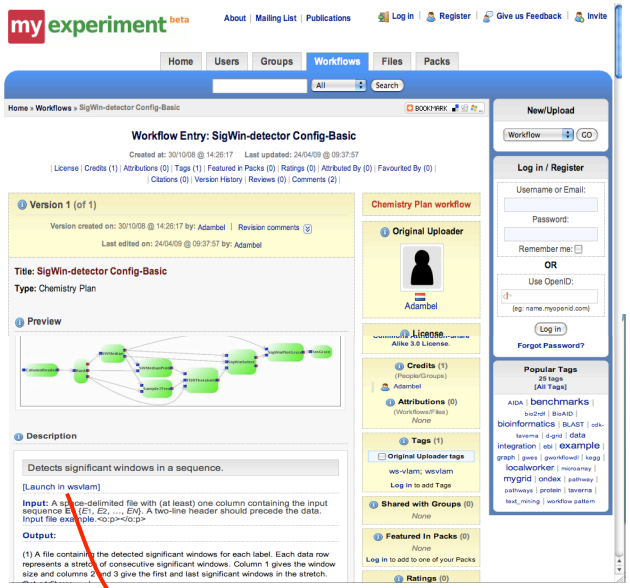
success

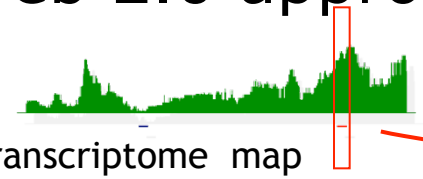
publish

(4) Results Publication

- Workflow can be invoked form other systems
- Workflow can be made available to entire community (using Web 2.0 approach)

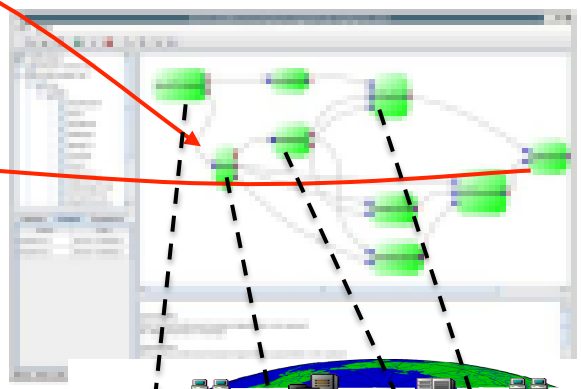
•myexperiment web site

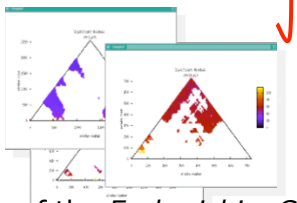




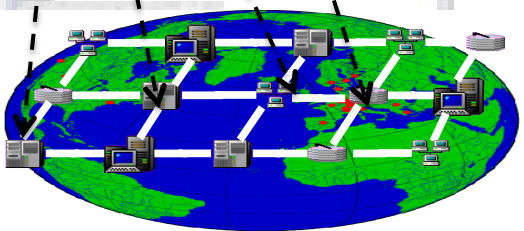
•Human transcriptome map

•WS-VLAM composer





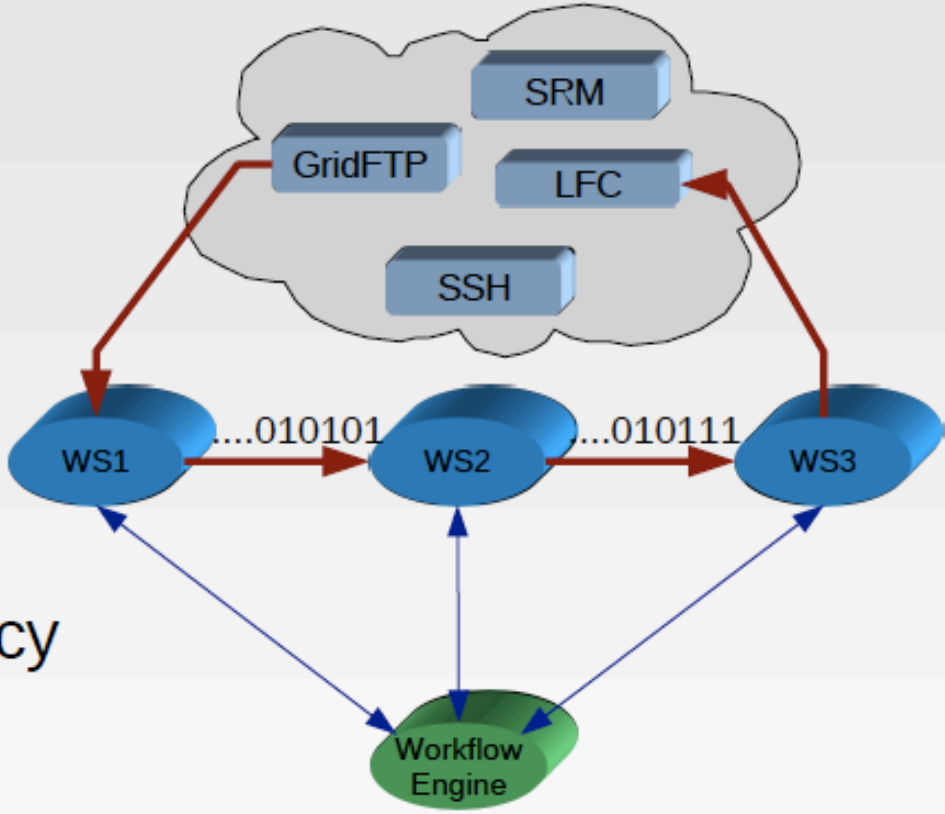
•DNA curvature of the *Escherichia Coli* chromosome



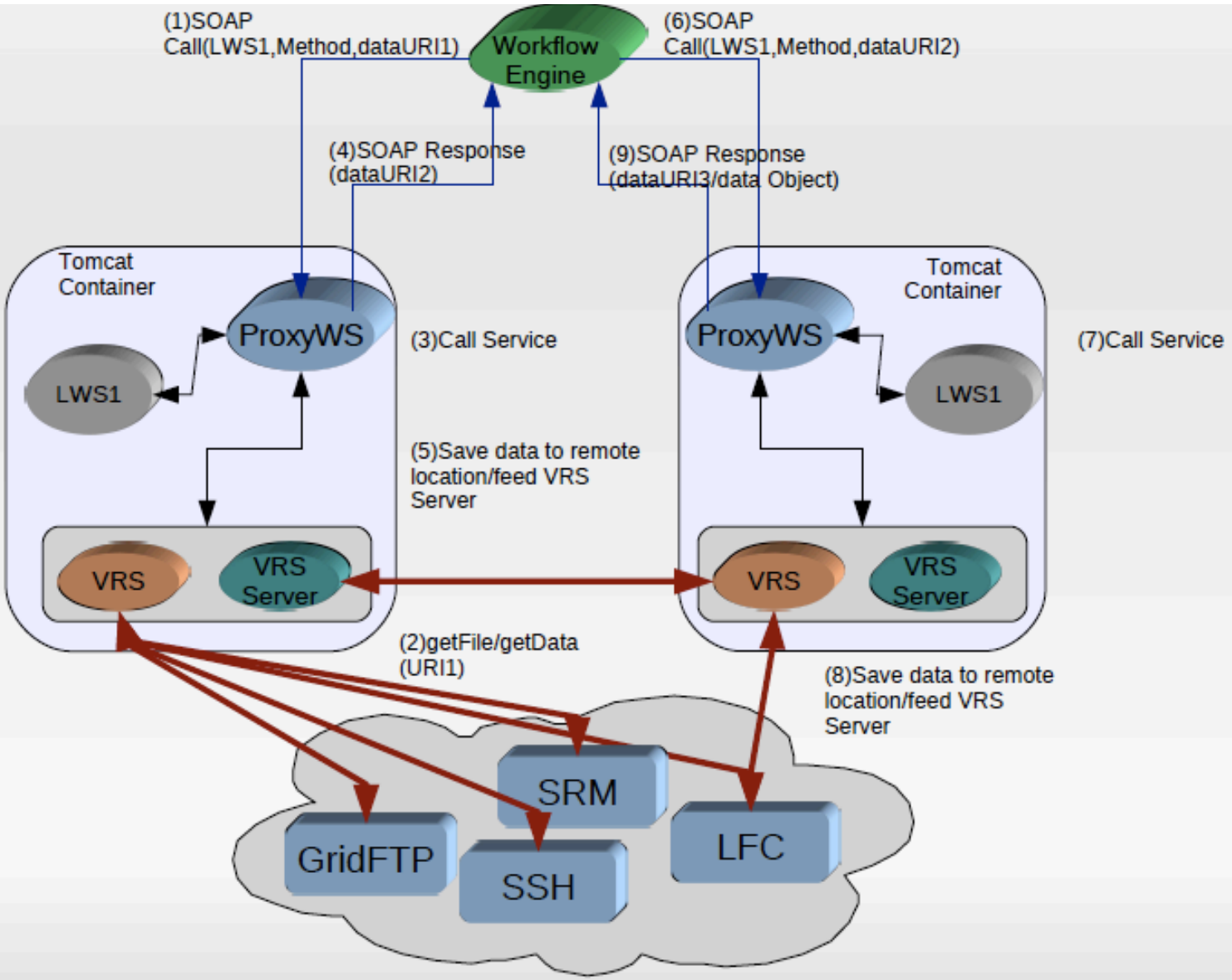
Marcia A Inda, Marinus F van Batenburg, Marco Roos, Adam SZ Belloum, Dmitry Vasunin, Adianto Wibisono, Antoine HC van Kampen and Timo M Breit SigWin-detector: *a Grid-enabled workflow for discovering enriched windows of genomic features related to DNA sequences,* BMC Research Notes 2008,

(*) Web Service and Data centric workflows

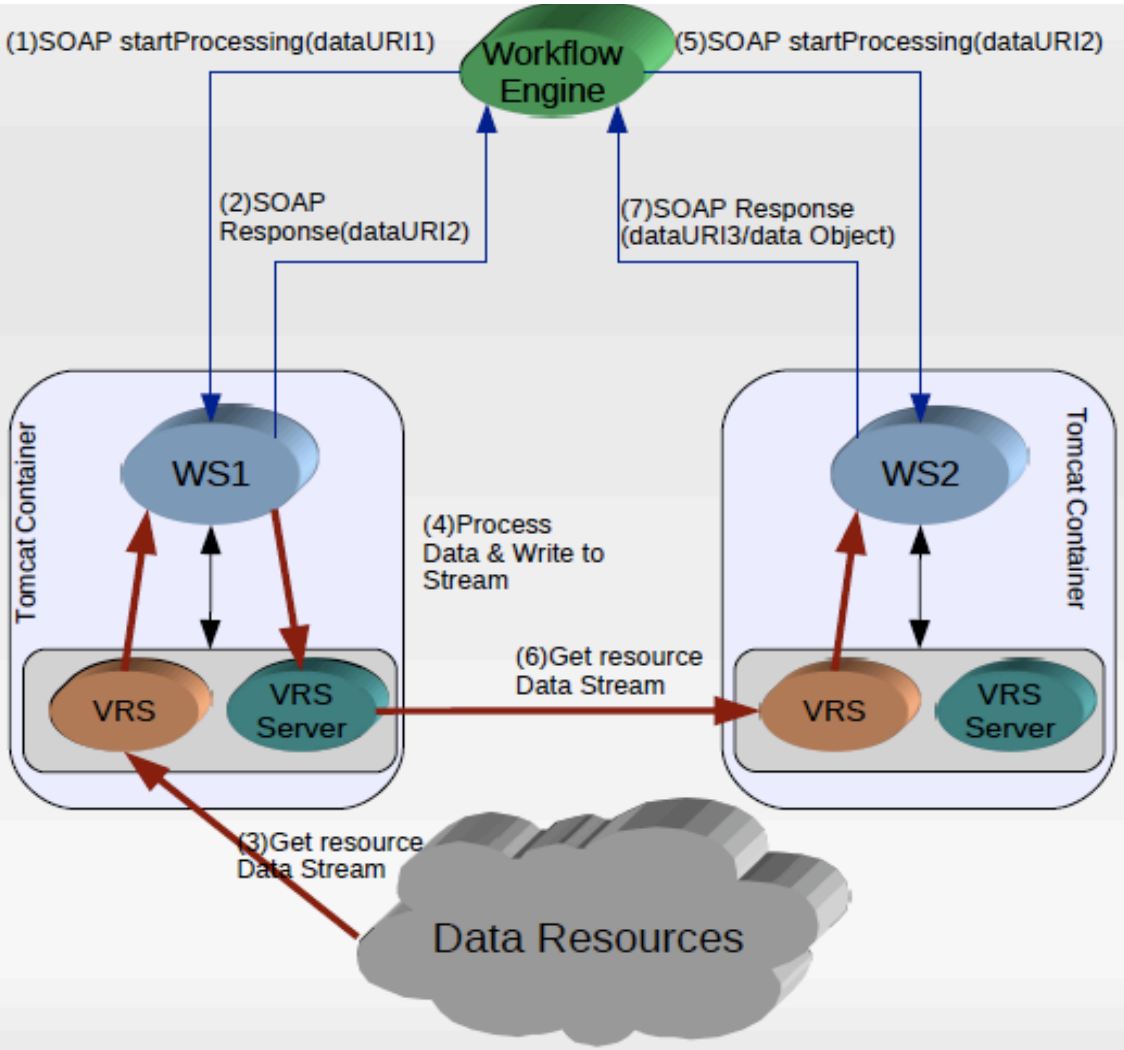
- Introduce a new execution model for data driven workflows.
- It will enable e-Science applications to meet increasing data demands.
  - Provide scalable data transports for web services.
  - Bring data to and from legacy web services
  - Develop an API that web services can use.



(*) Web Service and Data centric workflows



(*) Web Service and Data centric workflows



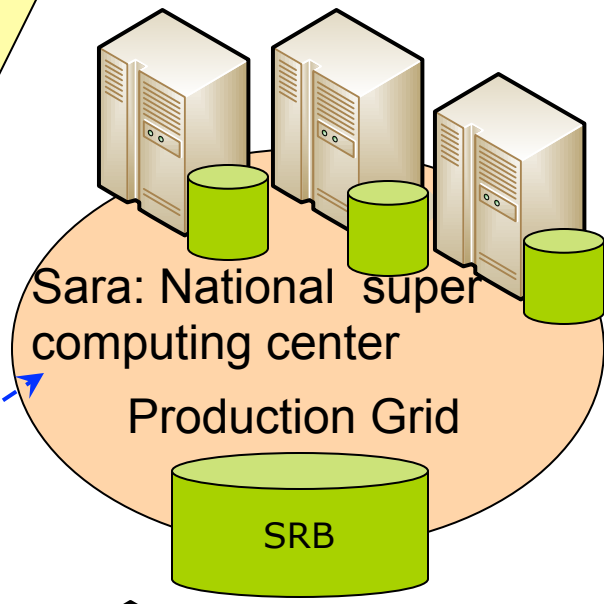
S. Koulouzis, E. Meij, M. S. Marshall, and A. Belloum, ``Enabling data transport between web services through alternative protocols and streaming" eScience, IEEE International Conference on e-Science, vol. 0, pp. 400–401, 2008. 14

Example of Complex eScie

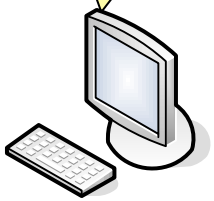
WSRF Services
- WS-VLAM engine
- workflow component repository

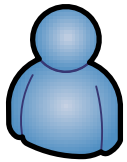
VLe Studio
- WS-VLAM composer
- VBrowser
- Semantic tools

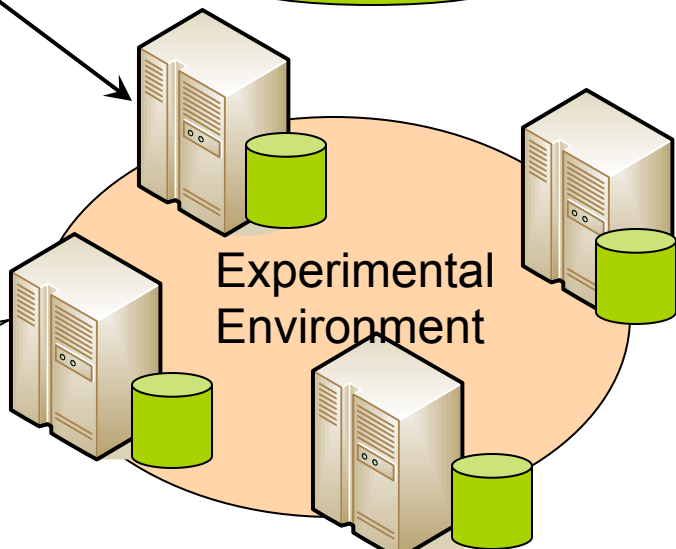
SAW: **S**emantic **A**nnotation for **W**orkflow
CLAMP: Connecting **LA**nguage for **M**odules & **P**rograms
HAMMER: Hybrid-b**A**sed **M**atch**M**aker for e-Science **R**esources



Server host

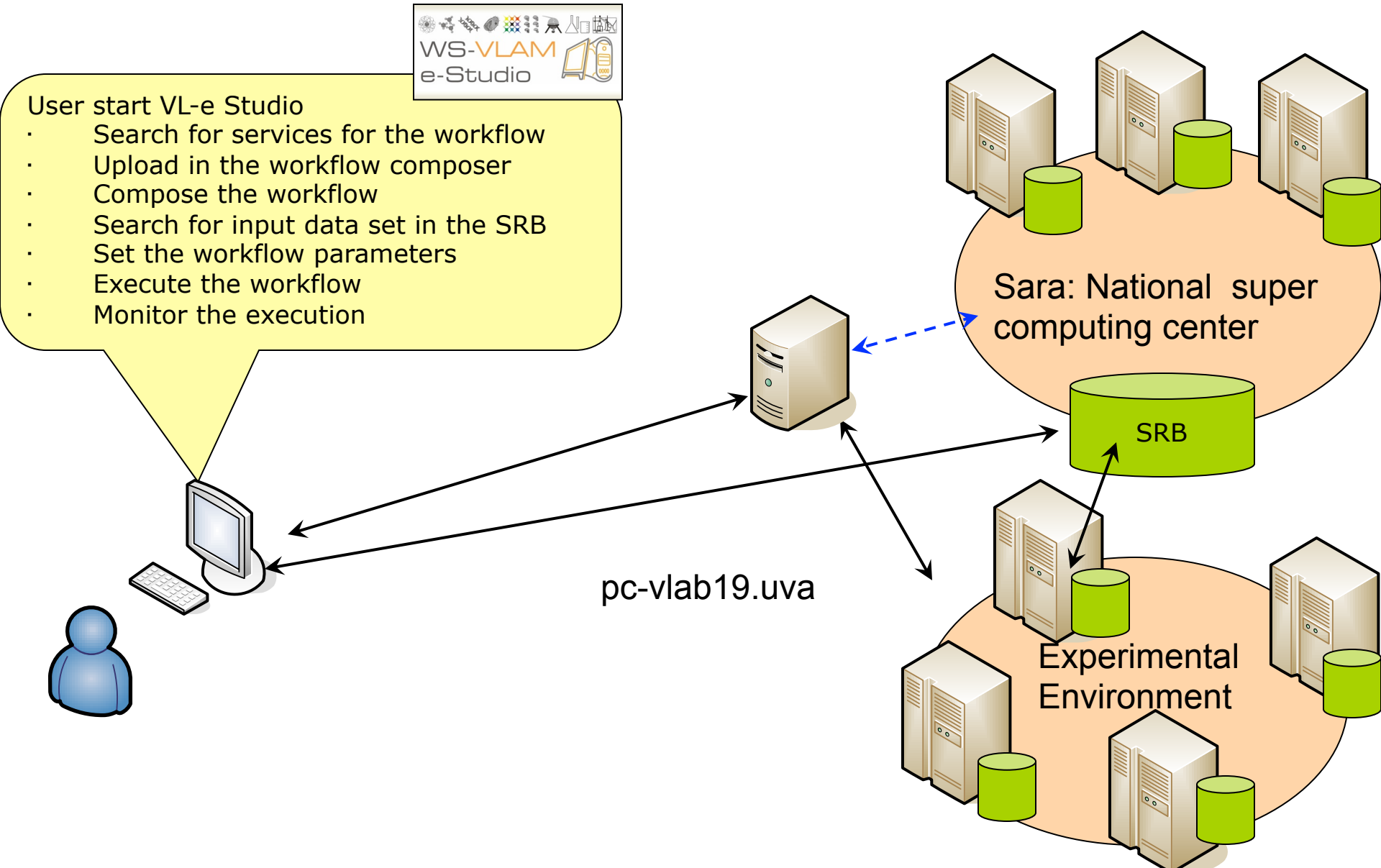






Computing Nodes
- Workflow components
- Grid Middleware → GT4

Example of usage Scenario …



References

1. **A.S.Z Belloum**, Z. Zhao, and M. Bubak, *Workflow systems and applications* Future Generation Computer System. 25 (5): 525-527 (2009), doi:10.1016/j.future.2008.10.011.
2. M.A. Inda, M.F van Batenburg, M. Roos, **A.S.Z Belloum**, D. Vasunin, A. Wibisono, A. HC van Kampen, and T.M. Breit, *SigWin-detector: a Grid-enabled workflow for discovering enriched windows of genomic features related to DNA sequences.* BMC Research Notes 2008, 1:63 doi:10.1186/1756-0500-1-63.
3. Z. Zhao, **A.S.Z. Belloum**, and M. Bubak. *Support for Cooperative Experiments in VL-e: From Scientific Workflow to Knowledge Sharing*, In Proceedings of the IEEE 4th International Conference on e-science and Grid Computing (eScience 2008), pp. 329-330, 7-12 Dec. 2008, doi:10.1109/eScience.2008.120.
4. A. Wibisono, Z. Zhao, **A.S.Z. Belloum**, and M. Bubak, *A Framework for Interactive Parameter Sweep Applications,* In Proceedings of the 8th International Conference on Computational Science, Part III, pp. 481-490, 2008, doi:10.1007/978-3-540-69389-5_55.
5. Z. Zhao, **A.S.Z. Belloum**, C. De Laat, P. Adriaans, and L.O. Hertzberger, *Distributed execution of aggregated multi domain workflows using an agent framework,* In Proceedings of the *IEEE Congress on* Services, pp. 183-190, 9-13 July 2007, doi:10.1109/SERVICES.2007.30.
6. V. Korkhov, A. Wibisono, D. Vasyunin, **A.S.Z. Belloum**, *Interactive Dataflow Driven workflow Engine,* Scientific Programming, vol.15 (3), pp. 173-188 (2007), ISSN 1058-9244.
7. V. Korkhov, A. Wibisono, D. Vasyunin, **A.S.Z. Belloum**, C. De Laat, P. Adriaans, L.O. Hertzberger*, WS-VLAM: Towards a Scalable Workflow System on the Grid*, In Proceedings of the 2nd workshop on Workflows in support of large-scale science, in conjunction with HPDC 2007, June 25 2007, pp. 63-68, doi: 10.1145/1273360.1273372
8. A. Wibisono, D. Vasyunin, V. Korkhov, **AS.Z. Belloum**, *WS-VLAM: a GT4 based workflow management system*, In Proceedings of the International Conference on Computational Science (ICCS 2007), Lecture Notes in Computer Science, vol. 4489-2007, pp. 191-198, doi:10.1007/978-3-540-72588-6_34.
9. Ketan C. Maheshwari , Sílvia D. Olabarriaga , Charl P. Botha , Jeroen G. Snel , Johan Alkemade , **Adam S.Z. Belloum**, *Problem solving environment for medical image analysis*, in CBMS 2007, In Proceedings of the 20th *IEEE International Symposium on Computer-Based Medical Systems (CBMS '07)*, pp.165-170, 20-22 June 2007, doi:10.1109/CBMS.2007.87

vl·e

http://www.vl-e.nl/