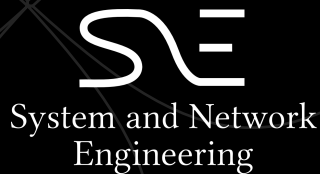


# COMMIT/



## WS-VLAM workflow management System



Adam Belloum  
Institute of Informatics  
University of Amsterdam  
a.s.z.belloum@uva.nl



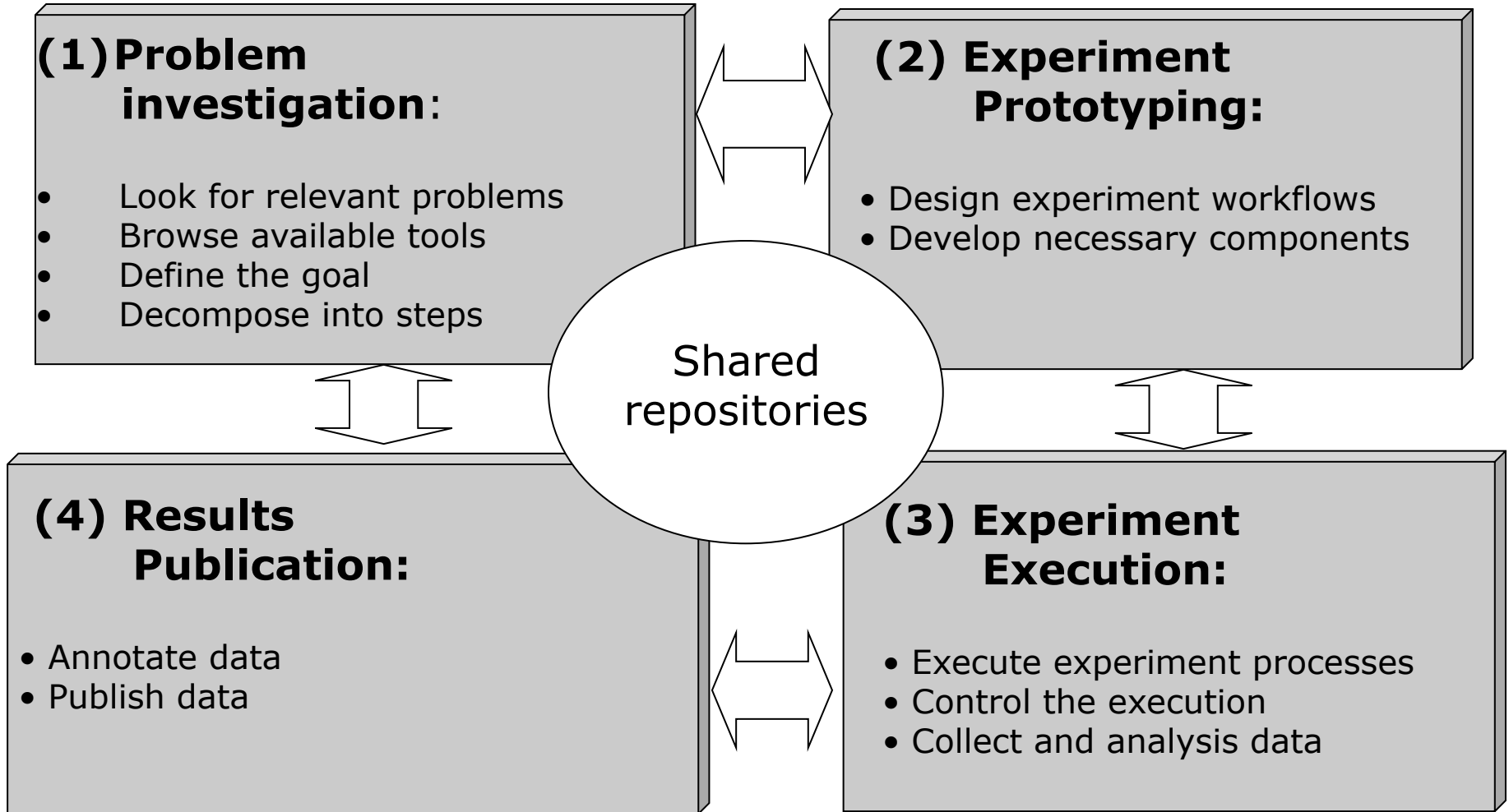
*EGI Technical Forum, 17-21 September 2012 Prague , 2012*

UvA



UNIVERSITEIT VAN AMSTERDAM

# Complex Scientific experiments model



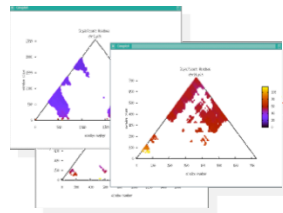
**Collaborative e-Science experiments: from scientific workflow to knowledge sharing** A.S.Z. Belloum, Vladimir Korkhov, Spiros Koulouzis, Marcia A Inda, and Marian Bubak JULY/AUGUST, IEEE Internet Computing, 2011

# Results Publication

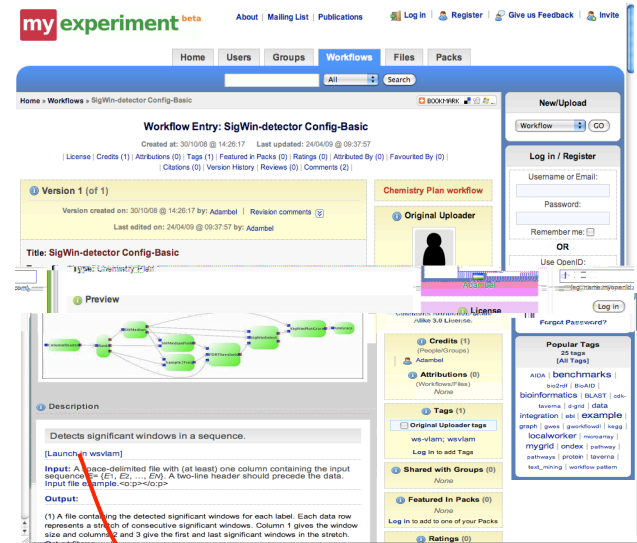
- Workflow can be invoked from other systems
- Workflow can be made available to entire community (using Web 2.0 approach)



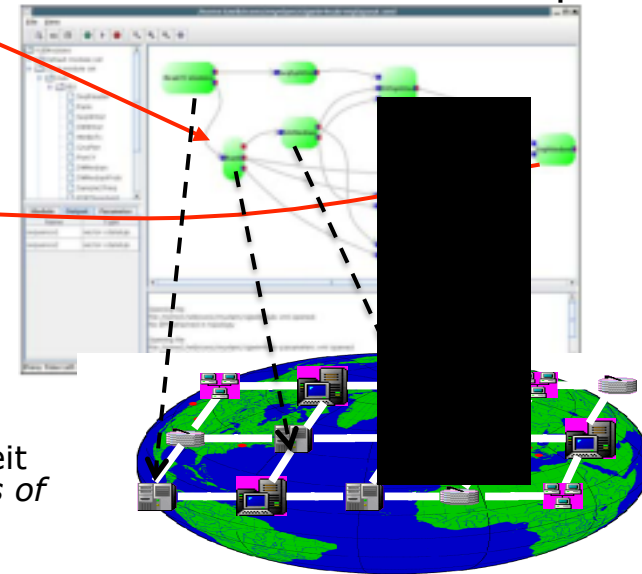
Human transcriptome map



DNA curvature of the *Escherichia Coli* chromosome



WS-VLAM composer



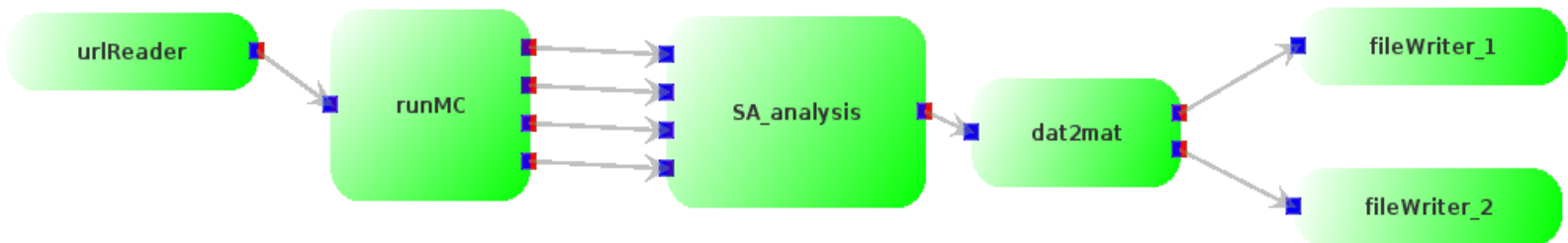
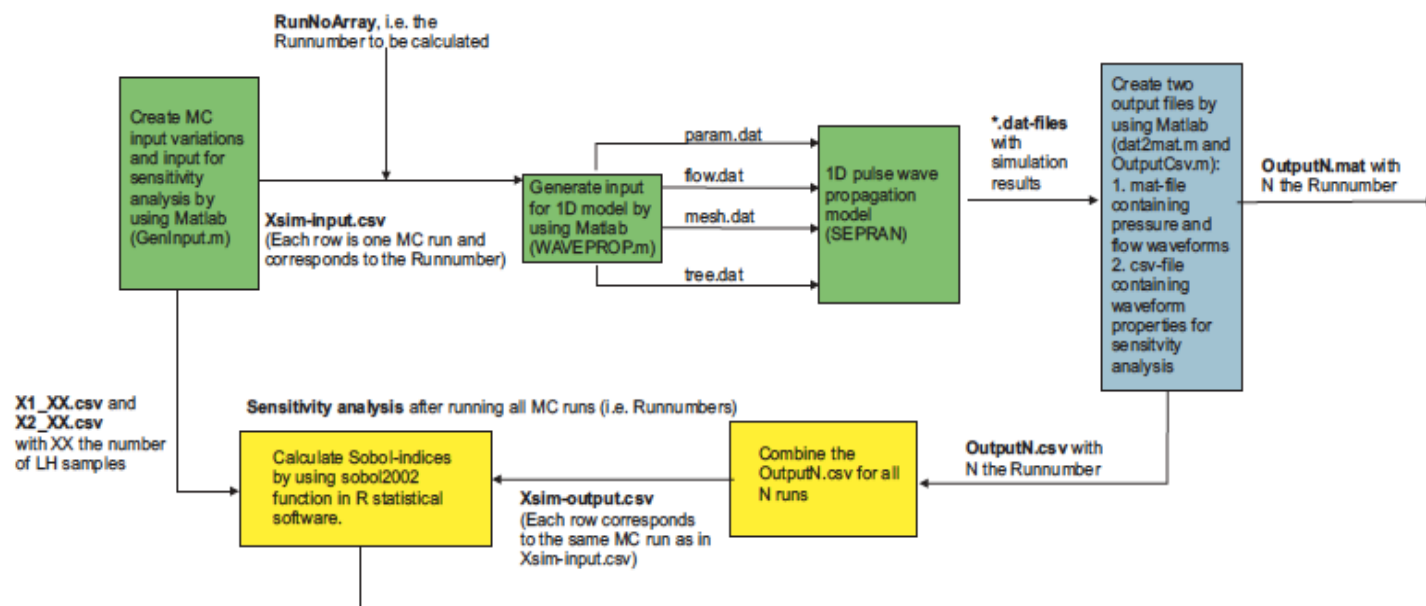
Marcia A Inda, Marinus F van Batenburg, Marco Roos, Adam SZ Belloum, Dmitry Vasunin, Adianto Wibisono, Antoine HC van Kampen and Timo M Breit  
 SigWin-detector: a Grid-enabled workflow for discovering enriched windows of genomic features related to DNA sequences, BMC Research Notes 2008,

# List of applications developed using WS-VLAM

- sigWin detector *[Micro-Array Dept-UvA]*
- *Affymetrix Permutation* *[Micro-Array Dept-UvA]*
- *Omnimatch* *[UU/Leiden]*
- wave propagation *[TUE ]*
- Blast *[AMC ]*
- gut microbiota *[CWI]*
- Smart Infrastructure *[SNE-UvA]*
- Dynamic network control *[SNE-UvA]*
- GridSFEA, *[TU Munchen]*

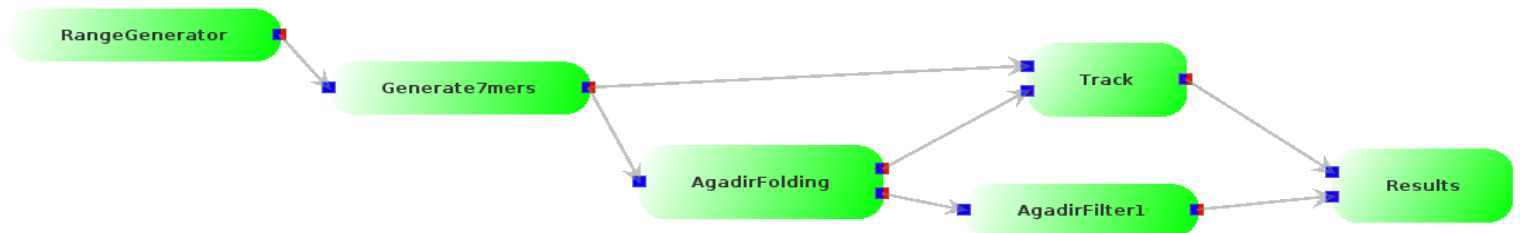
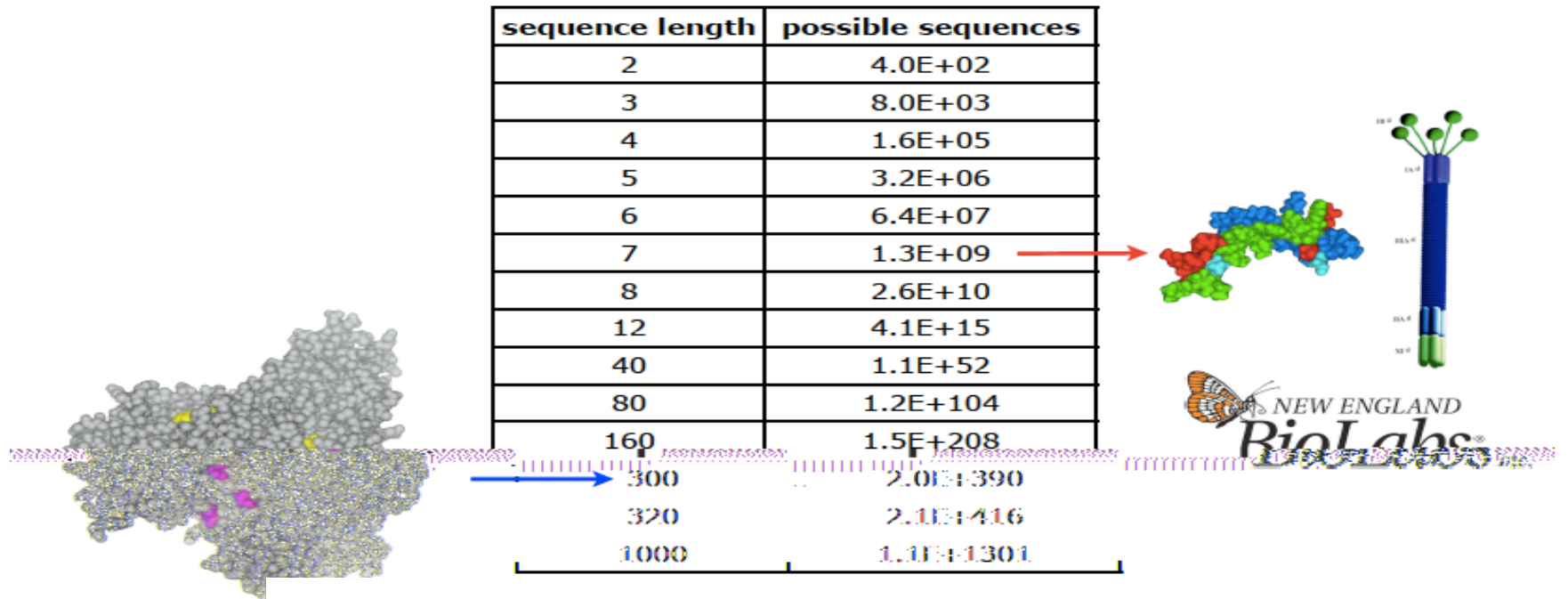
More applications [www.science.uva.nl/~gvlam/wsvlam/Applications](http://www.science.uva.nl/~gvlam/wsvlam/Applications)

# Sensitivity analysis of cardiovascular models

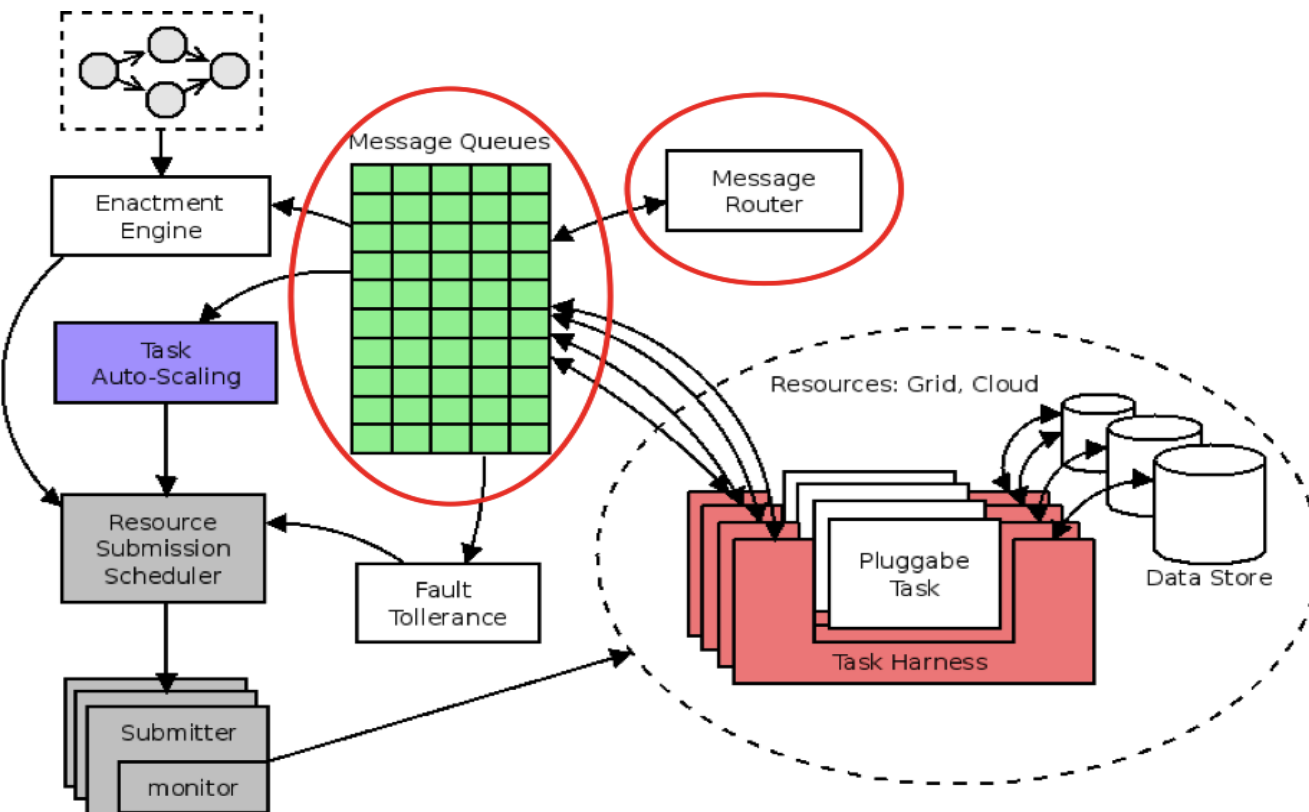


# Protein Folding

## Sequenomics: Mapping protein folding across sequence space



# Message Broker



Message broker plays a pivotal role in the system

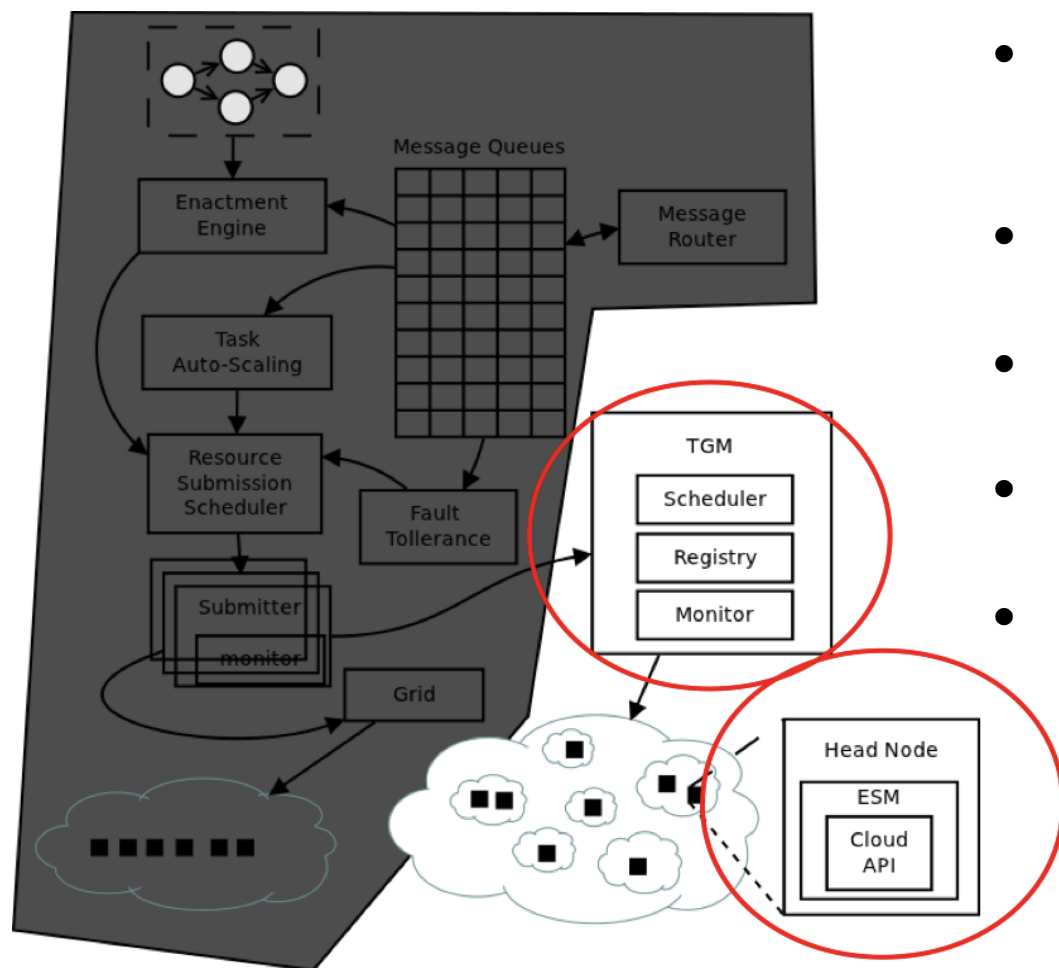
Message broker acts as a data buffer

Communicating tasks are **time decoupled**

Through queue sharing we can achieve scaling

Tasks **communicate** through messaging where messages contain **references** to actual data

# Extension to support Cloud resources



- Managing creation and deletion of clusters within multiple clouds
- Manages scheduling Vms to ideal cloud e.g. cost
- CAdapters abstract various cloud APIs
- TGM supports dynamic cloud addition and removal of clouds
- Each cluster head node is prepped with a controller program ESM and manages local Vms



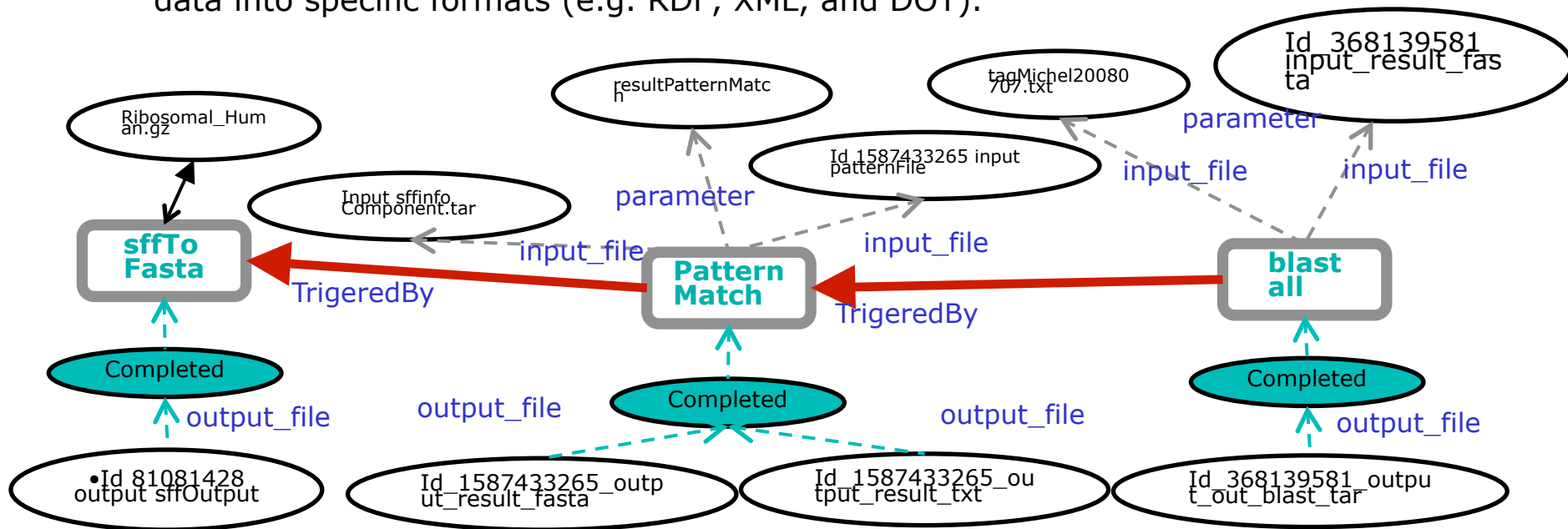
# History-tracing XML (FH Aachen)

- provides data/process provenance following an approach that
  - maps the workflow graph to a layered structure of an XML document.
  - This allows an intuitive and easy processable representation of the workflow execution path,
  - which can be, eventually, electronically signed.

```
<patternMatch>
  <events>
    <PortResolved> provenance
data</PortResolved>
    <ConDone>provenance data
      </ConDone>
    ...
  </events>
  <fileReader2>
    <events> ... </events>
    <sign-fileReader2> ...
      </signfileReader2>
  </fileReader2>
  <sffToFasta>
    Reference
  </sffToFasta>
  <sign-patternMatch> ...
    </sign-patternMatch>
</patternMatch>
```

# OPM 1.1

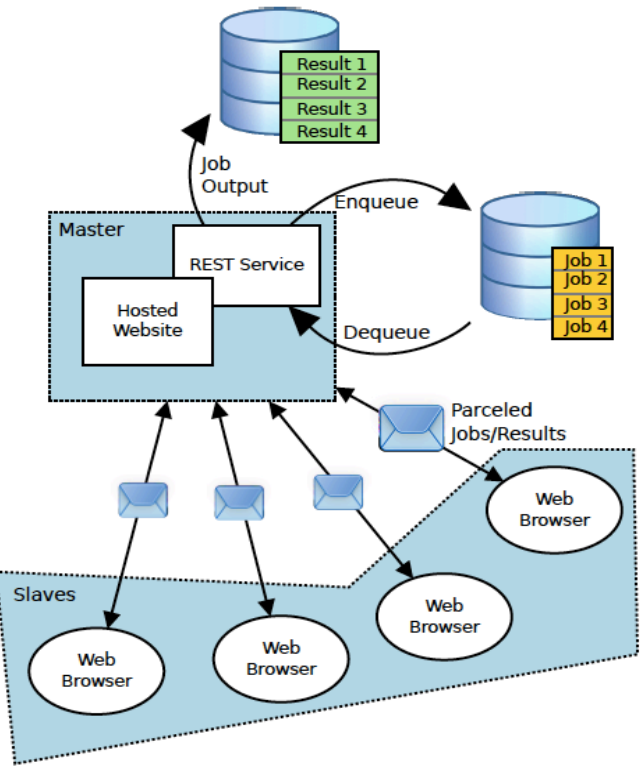
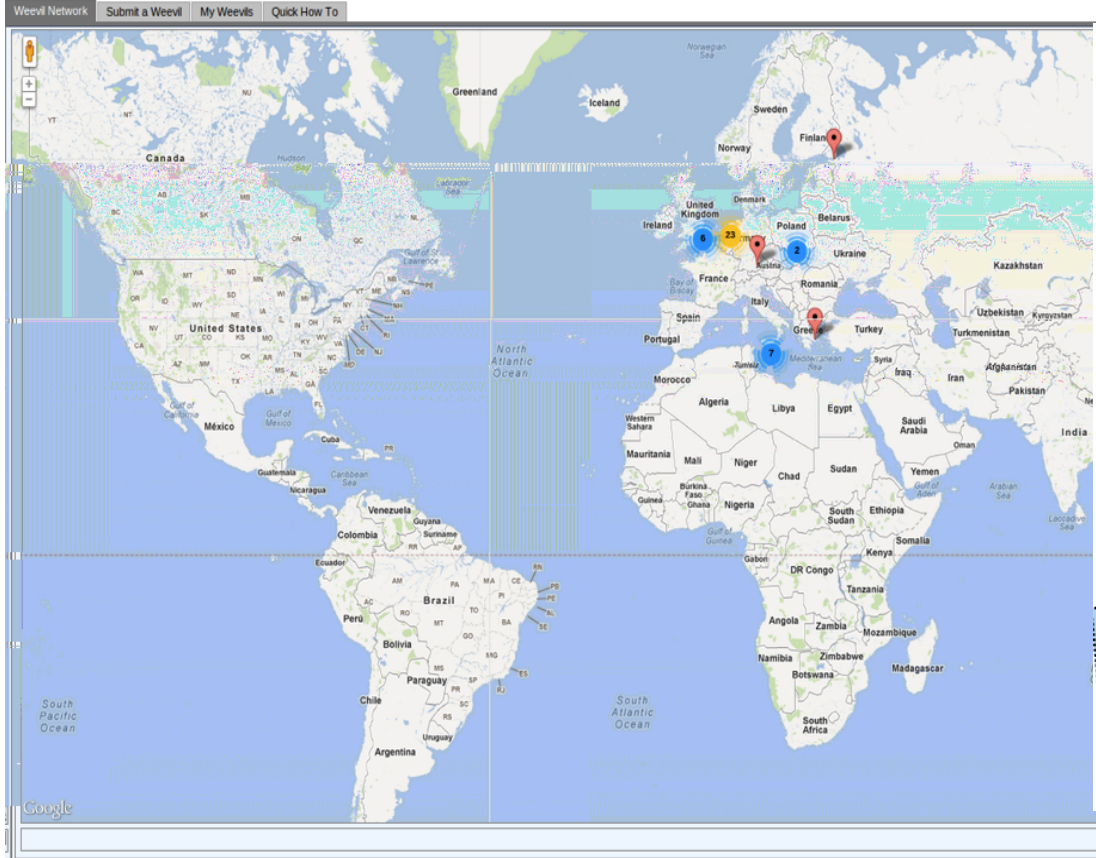
- PLIER is an implementation of the OPM 1.1 specifications.
- PLIER API
  - provides a set of functions to build, store, and share workflow experiments as graphs.
  - implements a relational database as back-end storage that captures the concepts of the OPM model, using the Java Persistence API (JPA 2.0) and Hibernate.
  - provides specific interfaces, using JDO 3.1, to transform, or serialize, the provenance data into specific formats (e.g. RDF, XML, and DOT).



# Browser computing resources

## Weevil Scout

Version: 0.2a BogoMFlops: 1076.4  
CLUSTER SIZE: 42 STATE: WORKING  
COMPLETED JOBS: 1125 DONATED TIME: 5341.543s

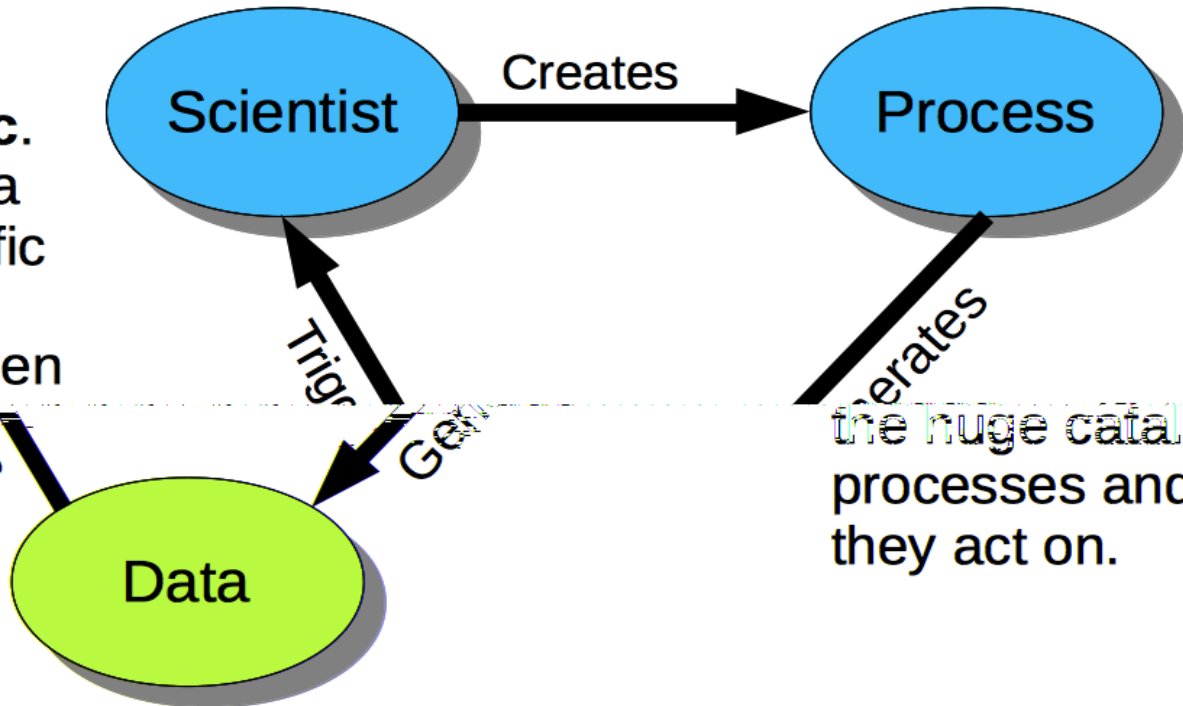


<http://elab.lab.uvalight.net/~weevil/>

CONTACT

# Cycle of research

E-science is becoming increasingly **data-centric**. Processes acting on data give **meaning** to scientific data thus a **semantic** relationship exists between



the huge volume of the data

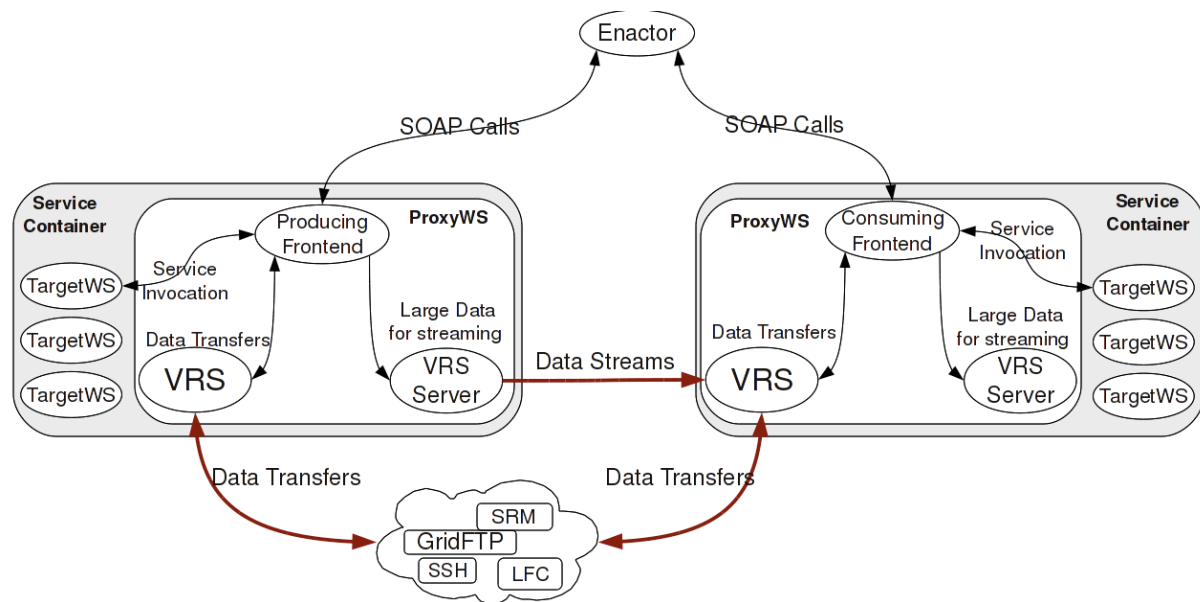
the huge catalog of processes and they act on.

relationship between the many processes and the data within a specific domain form an **ecosystem**. Ecosystems consist of **workflows** within a domain.

The **symbiotic** relationship between processes and data give rise to a dynamic system that incorporates all

# Usage of Web Services in e-science

- In service orchestration, **all data is passed to the workflow engine** before delivered to a consuming WS
- Data transfers are made through **SOAP**, which **is unfit for large data transfers**

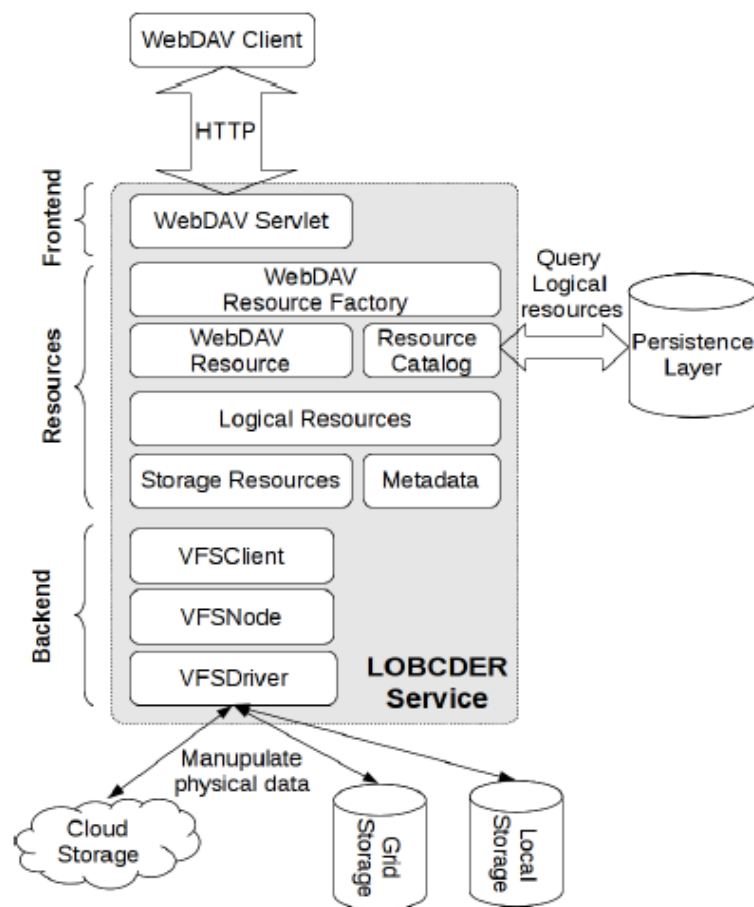


## Enabling web services to consume and produce

**large distributed** datasets Spiros Koulouzis, Reginald Cushing, Konstantinos Karasavvas, Adam Belloum, Marian Bubak to be published JAN/FEB, IEEE Internet Computing, 2012

# Large Object Cloud Data storage fedeRation

- Service for accessing large binary objects stored in various framework and providers



# References

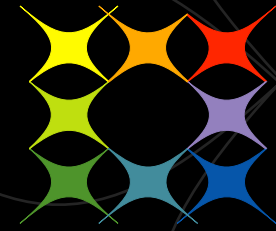
1. A. Wibisono, R. Koning, P. Grosso, **A.S.Z. Belloum**, M. Bubak and C. de Laat, *Building an Adaptive Online Ontology Instance Editor*, Concurrency and Computation: Practice and Experience (2012): 16 Jul. 2012, doi:10.1002/spe.2141.
2. S. Koulouzis, R. Cushing, K. Karasavvas, **A.S.Z. Belloum**, M. Bubak, *Enabling Web Services to consume and produce large distributed datasets* IEEE Internet Computing, vol. 16, no. 1, pp. 52-60, January/February, 2012, doi:10.1109/MIC.2011.138.
3. S. Koulouzis, R. Cushing, D. Vasunin, **A.S.Z. Belloum** and M.T. Bubak, *Cloud Federation for Sharing Scientific Data*, In Proceedings of the 8<sup>th</sup> IEEE International Conference on eScience (eScience 2012) Chicago, Illinois, 8-12 October 2012.
4. Reginald Cushing, **Adam S.Z. Belloum**, V. Korkhov, D. Vasyunin, M.T. Bubak, C. Leguy, *Workflow as a Service: An Approach to Workflow Farming*, In Proceeding of the 3rd international workshop on Emerging computational methods for the life sciences (ECMLS '12), doi:10.1145/2483954.2483960.
5. R. Cushing, S. Koulouzis, **A.S.Z. Belloum**, M.T. Bubak, *Service Level Management for Executable Papers* Euro-Par 2011: Parallel Processing Workshops, Lecture Notes in Computer Science vol. 7156, 2012, pp. 116-123, doi: 10.1007/978-3-642-29740-3\_14.
6. Szeplieniec, J. Kocot, T. Schaaf, O. Appleton, M. Heikkurinen, **A.S.Z. Belloum**, J.S. - Fernandez, M. Metzker, *On Importance of Service Level Management in Grids*, In Proceedings of EuroPar 2011: Parallel Processing Workshops, Lecture Notes in Computer Science, vol. 7156, 2012, pp. 64-75, doi: 10.1007/978-3-642-29740-3\_9.

# COMMIT

<http://www.commit-nl.nl/>



<http://www.science.n/~gvlam/wsvlam/>



vl-e

<http://www.vle.nl/>