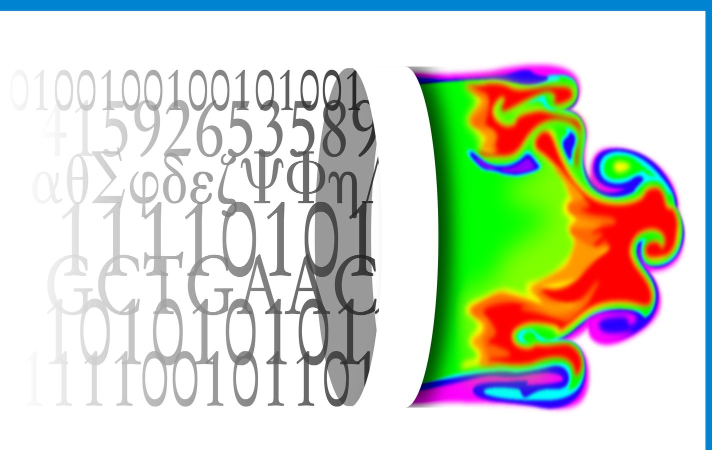


University of Amsterdam

Computational Science

Mastering Complexity in e-Science Experiments

Reginald Cushing, Adam Belloum, Marian Bubak

Informatics Institute, University of Amsterdam, The Netherlands

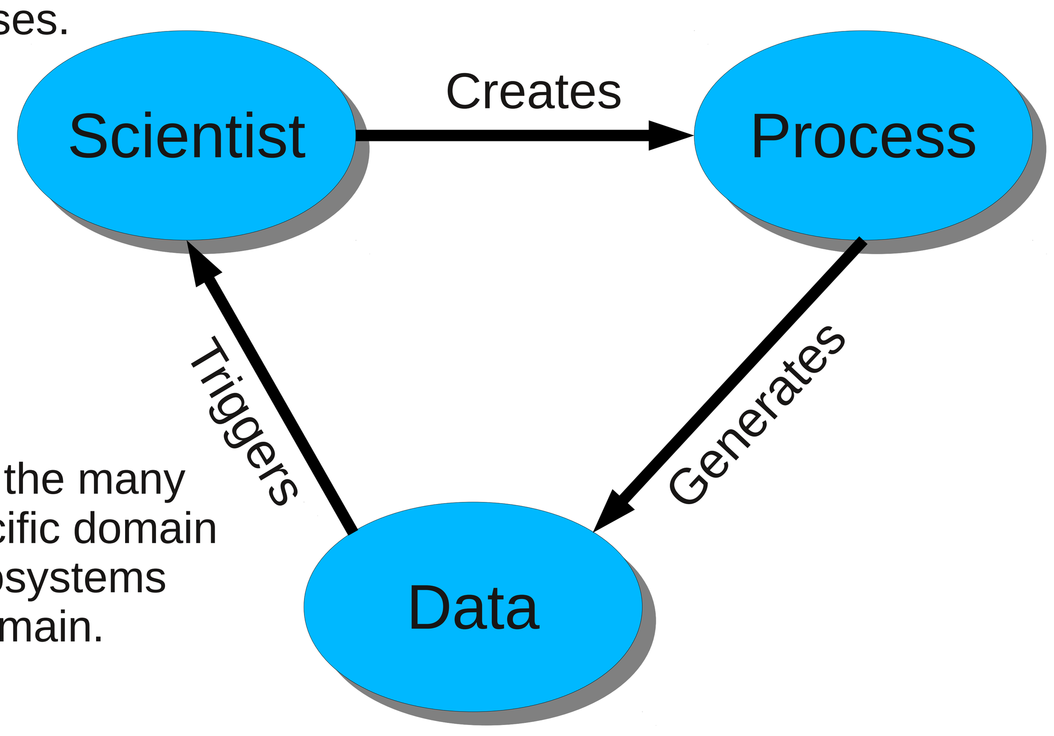
http://staff.science.uva.nl/~gvlam/wsvlam/

1. Motivation

E-Science is in a continuous **cycle** of **evolution**. The evolution cycle involves scientists producing new processes that generate new data which trigger new processes. Through this cycle the domain ecosystem evolves to incorporate new processes and data.

With many processes being developed (e.g. **BioCatalogue** has over **2000** services) and data being produced, ecosystems are expected to grow and thus the complexity of the domain ecosystem also increases.

E-science is becoming increasingly **data-centric**. Processes acting on data give **meaning** to scientific data thus a **semantic** relationship exists between the huge catalogue of processes and the data they act on.



The **symbiotic** relationship between the many processes and the data within a specific domain give rise to a domain **ecosystem**. Ecosystems incorporate all **workflows** within a domain.

2. Objectives

Our main goal is to investigate the challenges in **complex e-Science experiments**. Specifically dealing with:

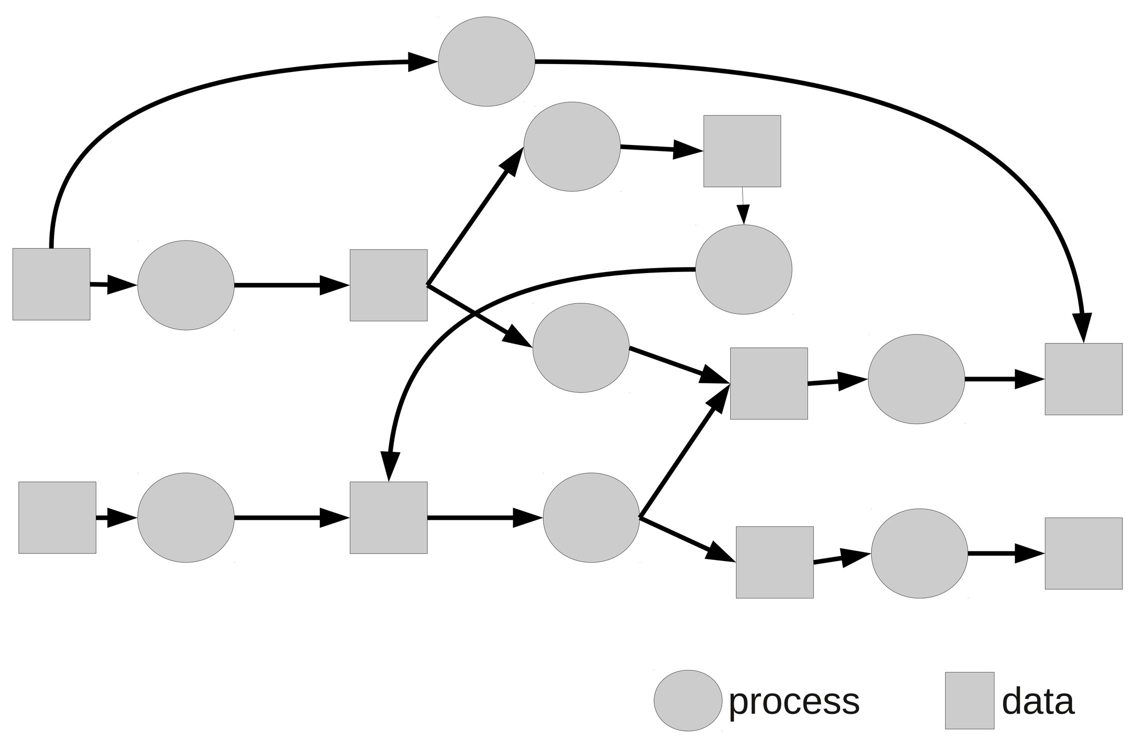
- Integrating **multiple** domain specific **workflows** and processes into domain **ecosystems**
- Scientist-ecosystem **interaction** for learning about new workflows
- **Modeling** complex workflow processes
- Workflow **re-usability** between multiple scientific workflow management systems
- **Provisioning** computing and network resources to sustain CPU and data intensive experiments
- Knowledge **preservation** through ecosystems
- **Optimization** of ecosystems through workflow **provenance** and profiling feedback

3. E-Science Ecosystem

An ecosystem can be viewed as a **semantically** annotated network where processes transform data between inputs and outputs. The network evolves as new processes are added.

With an **evolving** ecosystem, multiple **paths** will eventually exist for the same data transformation.

The ecosystem shows **what** can be done in a specific domain given an input data.



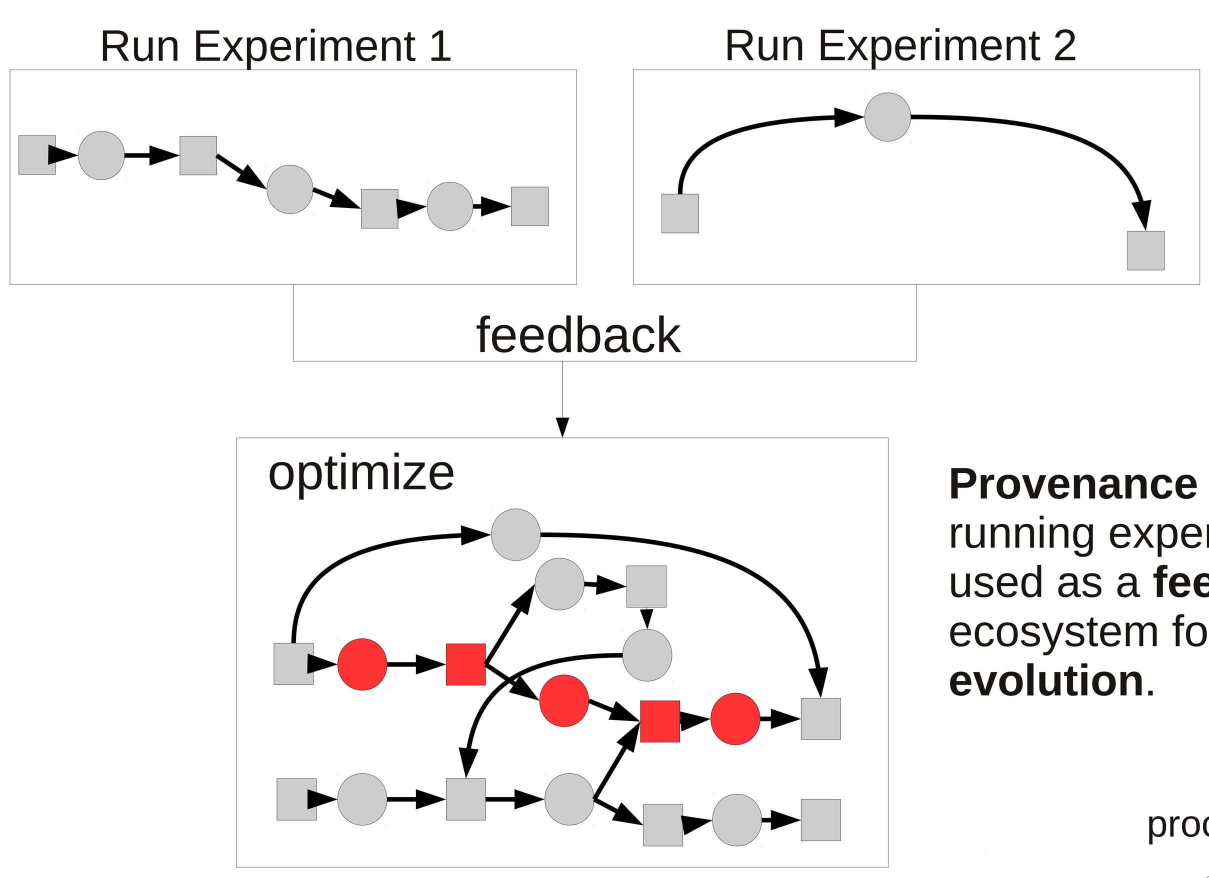
4. Workflow Generation

An **ecosystem** describe **what** can be done within a domain and **how** it can be achieved. In essence the **"how"** is a generated **workflow** out of the ecosystem. **Querying** what can be done we can deduce how the experiment can be achieved(workflow generation).

Common approach to e-Science is for scientists to build their **workflows.** Scientists have to know **what** (the experiment) they want and **how** (the composition) they want it accomplished.

The huge libraries of processes to choose from, **composing** a workflow can be a daunting task. Process catalogues such as BioCatalogue provide searching interfaces to find web services as it states *"Web Services are hard to find"*

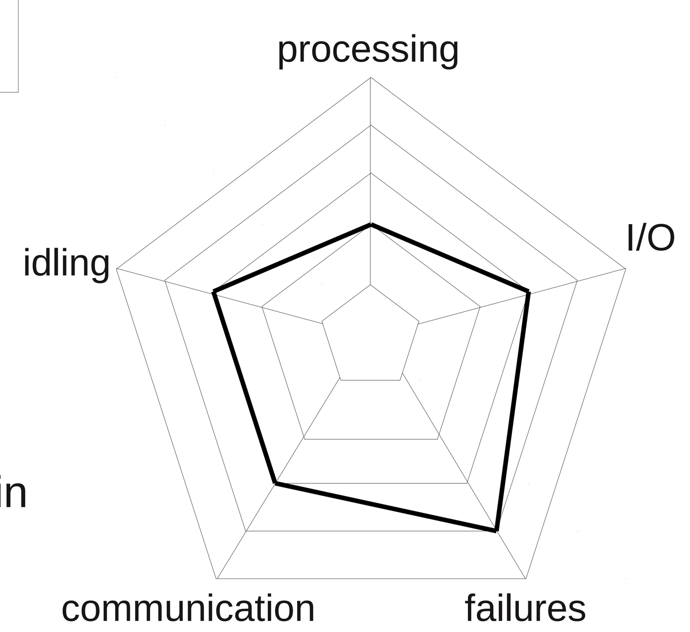
5. Provenance and Profiling Feedback



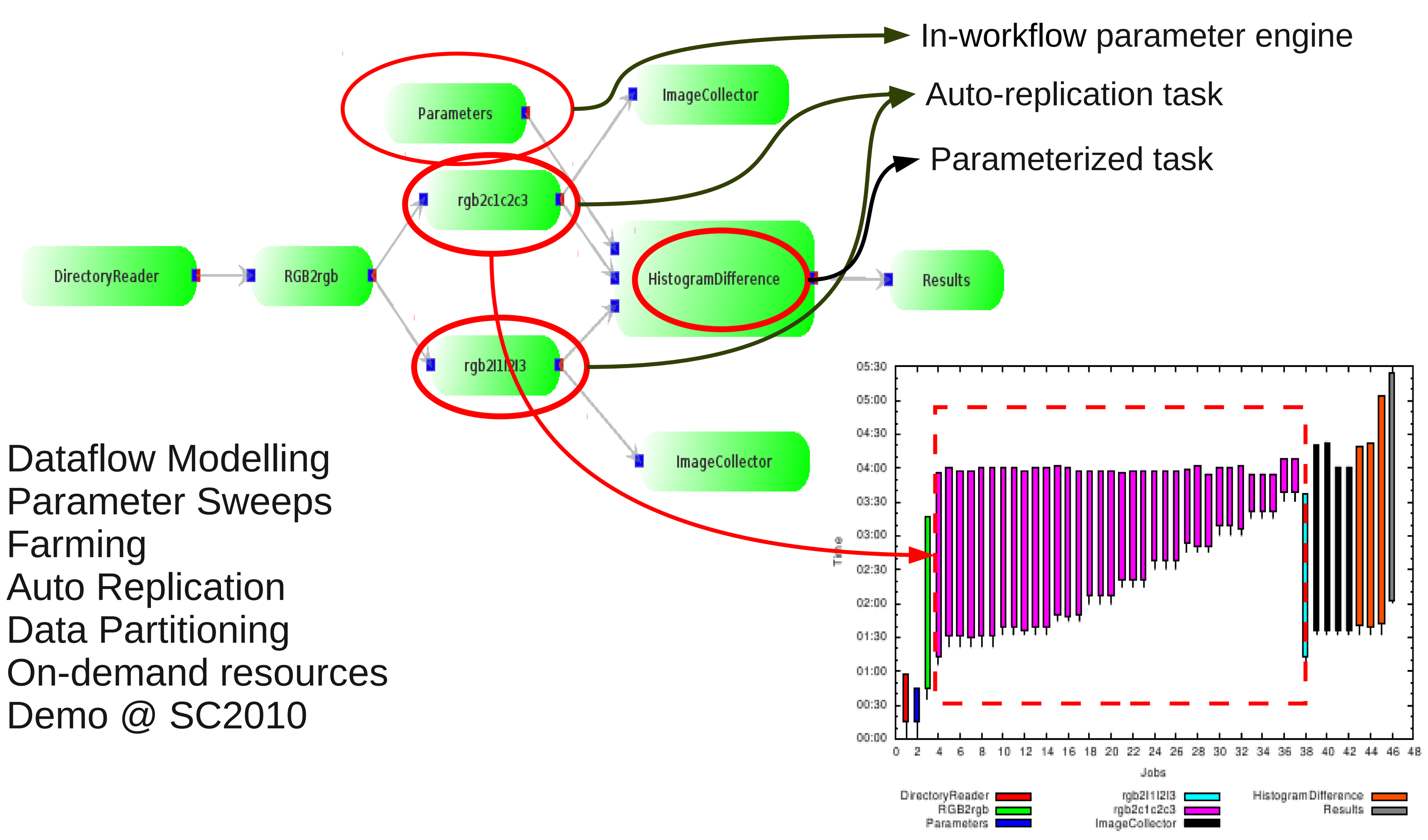
Provenance collected from running experiments can be used as a **feedback** into the ecosystem for further **evolution**.

Process **profiling** can help an ecosystem **mature** by learning about the characteristic of processes within the system.

Maturity is increased by for example systematically removing failing processes in an effort to maintain a stable ecosystem.

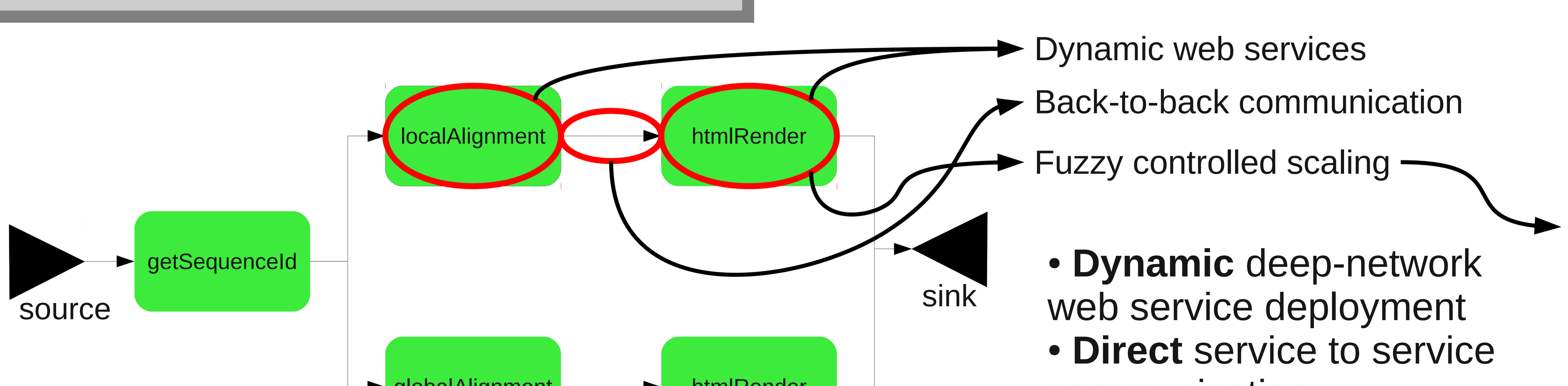


6. Dataflow Workflow Enactment

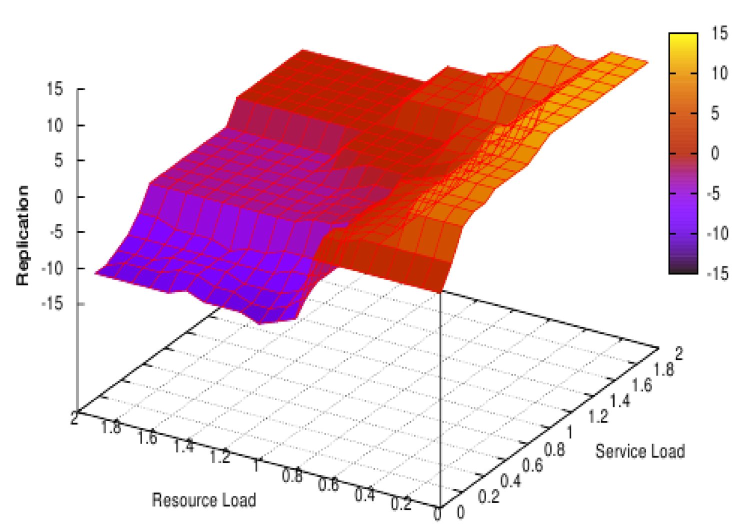


- Dataflow Modelling
- Parameter Sweeps
- Farming
- Auto Replication
- Data Partitioning
- On-demand resources
- Demo @ SC2010

7. Cooperating Web Services



- **Dynamic** deep-network web service deployment
- **Direct** service to service communication
- **Elastic prediction-based** service scaling



Publications

[1] R. Strijkers, R. Cushing, et al *Towards Executable Scientific Publications*, ICCS 2011.

[2] R.Cushing, S.Koulouzis, et al *Data-driven prediction-based auto-scaling framework for scientific workflow tasks on distributed systems*. Submitted to ACM/IFIP/USENIX Middleware 2011.

[3] R.Cushing, S.Koulouzis, et al, *Dynamic Handling for Cooperating Scientific Web Services*, Submitted to IEEE e-Science 2011.

COMMIT CAPACITIES gSLM VPH-Share

Contact Person: Reginald Cushing
Email: R.S.Cushing@uva.nl