



Complex In-silico Experiments in Integrative Biology



Overview

- Project introduction
- IB requirements and context
- Status and plans
- Some specific issues

Project Overview

- EPSRC Best Practice project
- from myGrid ...
 - workflow / information model
- ... to IB
 - computationally demanding domain
- 1 year, started Jan 2005, 2sy
- CCLRC, IT Innovation, Manchester, Oxford

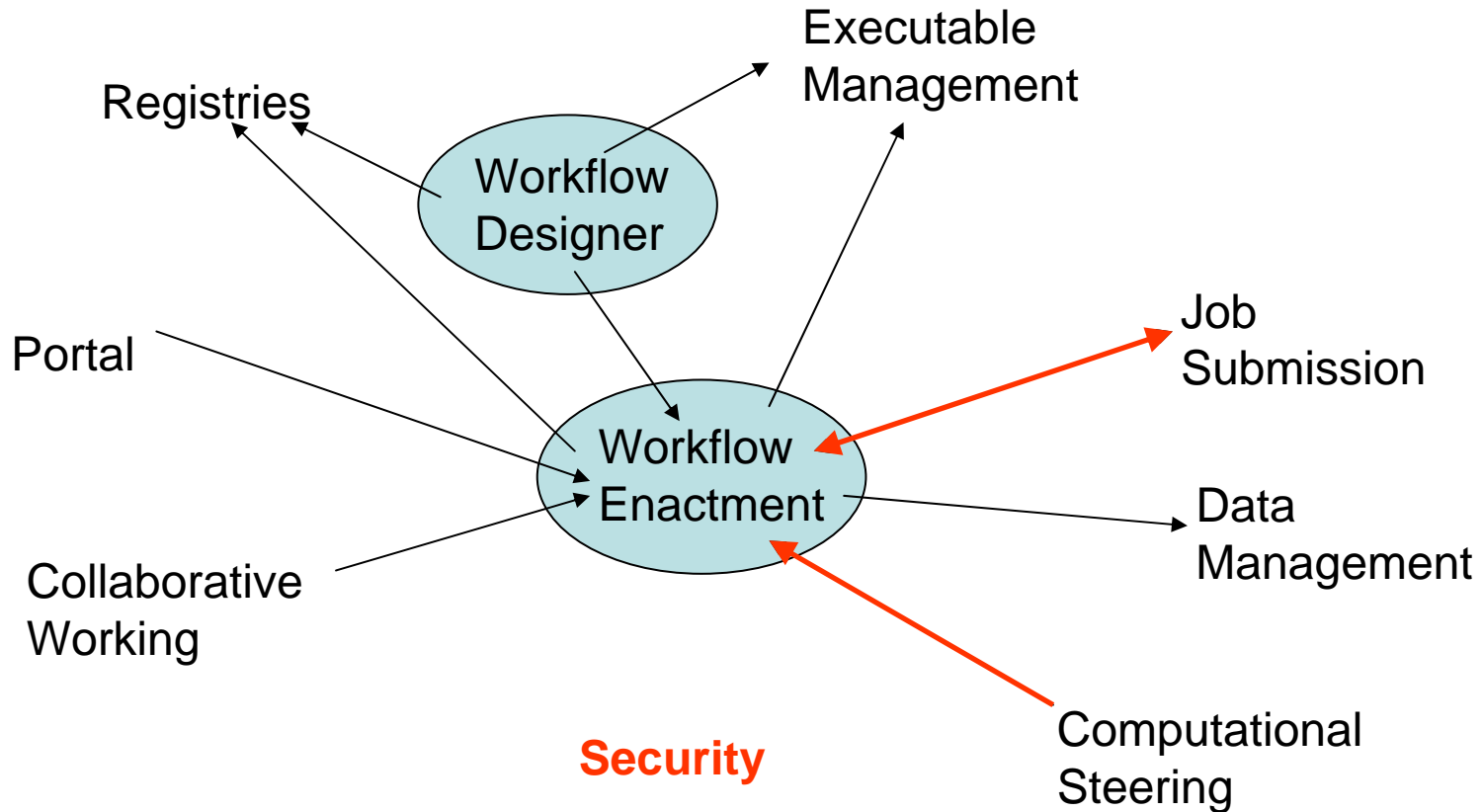




Objectives

- “Extend scientists ability to steer in-silico experiments beyond current computational steering to cover the whole experimental process”
- “Provide the necessary information management to make this useful”

Workflow in the IB Environment





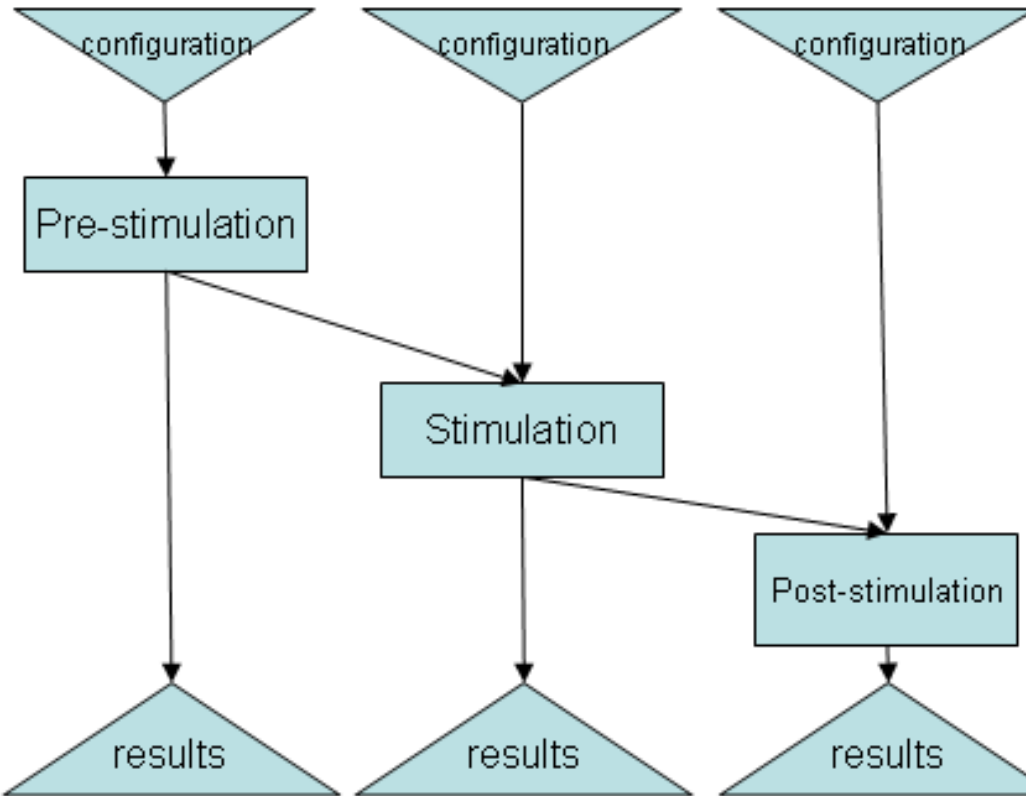
IB workflow characteristics

- moderate workflow complexity
 - some tight coupling, c.f. coupled simulation models
 - mostly loose, linear sequences
- large, long running activities
 - handling and monitoring HPC jobs (batch & interactive)
 - computational steering (of the activity)
- large data flows
 - streaming of data between activities
 - separate data flows from control flows
- dynamic workflows
 - workflow steering (ad hoc workflows)



Prototype workflow

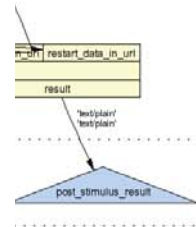
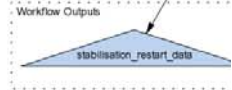
Inputs



stabil

Simulation processors

Outputs



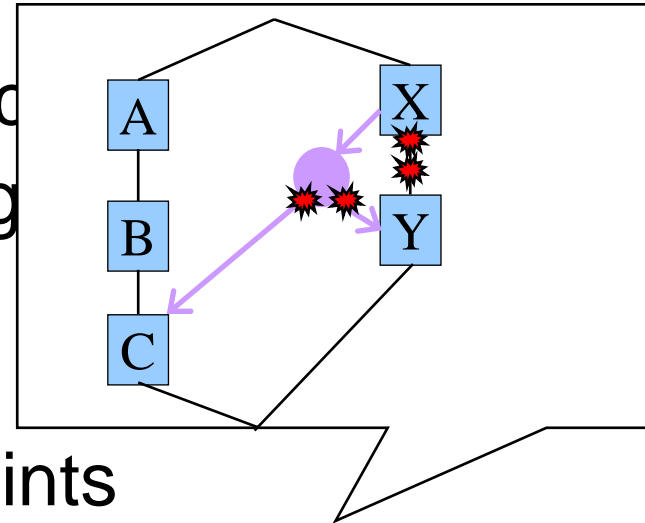


Status

- Initial investigations completed
 - issues to be addressed
 - initial workflows modelled
- Workplan
 - workflow extensions for HPC (ongoing)
 - steering for workflows (ongoing)
 - provenance for steering (starting)
 - annotations
 - integration into IB
 - extract/capture “best practice” for reuse

Steering workflows

- Steering via the Taverna client
 - reconnection to running workflows
 - pause/restart facility
 - setting breakpoints
 - editing data at breakpoints
 - integrity in the face of concurrency
 - impact on provenance
 - invisible (to Taverna) edits
 - LSID versioning (see later)



Data Management

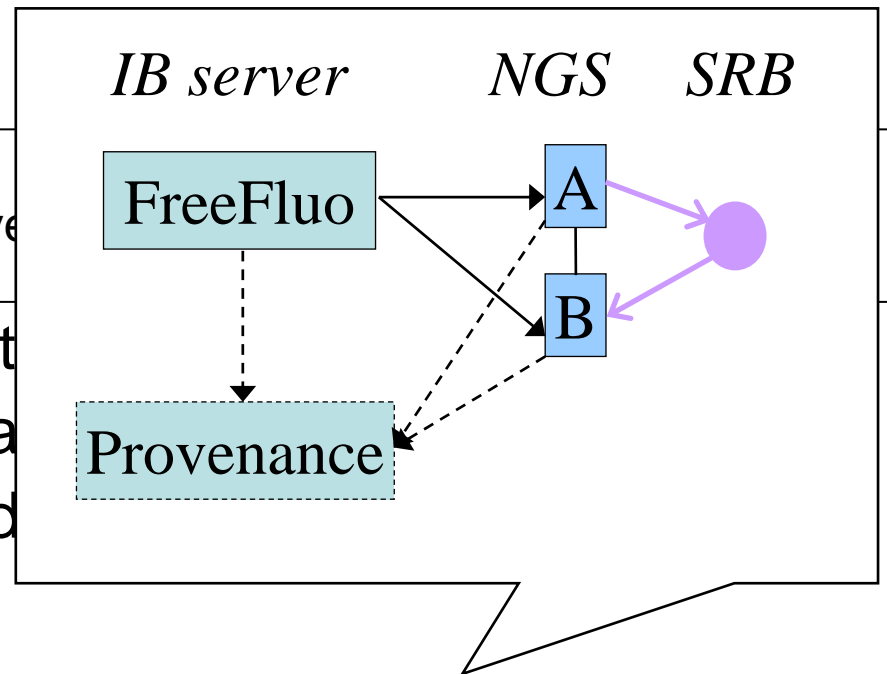
- LSID

urn:lsid:www.integrative

- intermediate result
- local copies / stream
- SRB support required

- Data Marshalling

- integration with SRB
- pass-by-reference (using LSIDs?)
- balance explicit / implicit marshalling



Summary

- **myGrid** provides some quite sophisticated tools
- but **IB** brings in a number of new wrinkles.
- which **myIB** is addressing.

- Thank you

- Questions?