

vl·e



virtual laboratory for e·science

Dynamic workflow

Adam Belloum

Systems and Networking Engineering research group

Institute of Informatics

University of Amsterdam

UvA



UNIVERSITEIT VAN AMSTERDAM



Outline

- Definition of workflows (**Business, Scientific, Dynamic**)
- Make the case for workflow and Dynamic workflow
- Described two different ways support Dynamic workflows

Business Workflows

- “The **automation** of a business process, in whole or parts, where **documents**, **information** or **tasks** are passed from one participant to another to be processed, according to a set of **procedural rules** “
 - Workflow Management Coalition

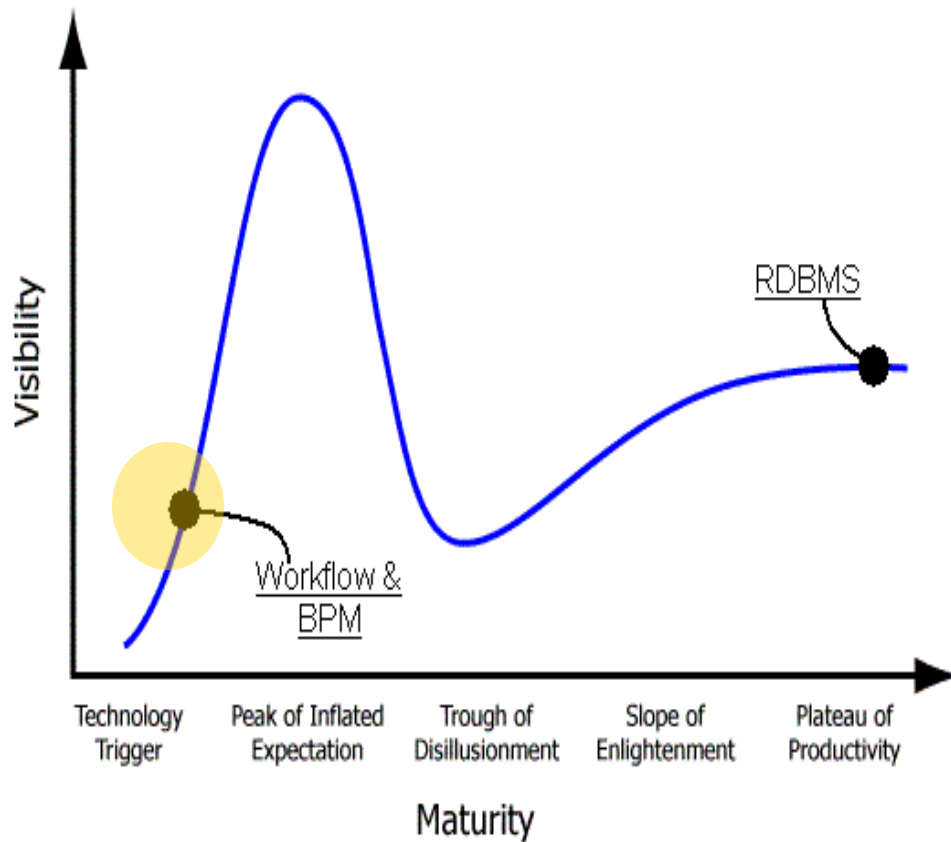


Scientific Workflows

- “These are networks of **analytical steps** that may involve, e.g., database access and querying steps, **data analysis** and **mining** steps, and many other steps including computationally **intensive jobs** on high performance cluster computers.”

Bertram Ludäscher et al. (Kepler project)

Workflow Hype Curve



- “When talking about an **RDBMS** in a software development team most people will **get the picture** and shake their heads slightly up and down **confirming they understand what you're saying.**”
...

The state of workflows



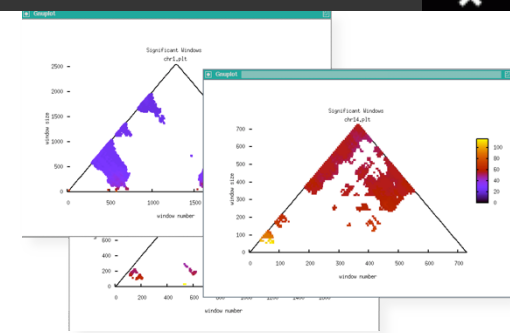
The case for **workflows**

- **Capturing knowledge/enforce best practice**
 - Capture business process based on the company policy
 - Capture best practices of scientist, expert from a specific domain.
- **Incorporate human decision in the process**
 - There are cases that can not be automated both in business and scientific workflow
- **Easy development**
- **Increase the re-usability**

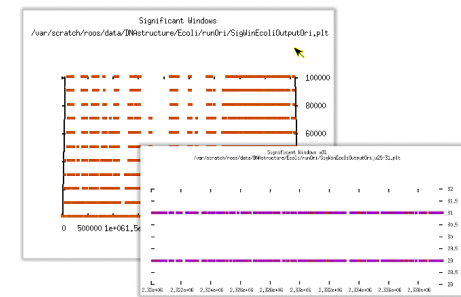
Business workflow ...



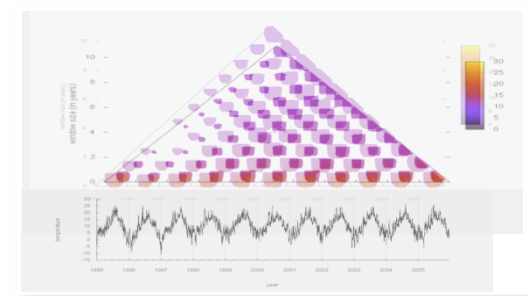
Scientific workflow



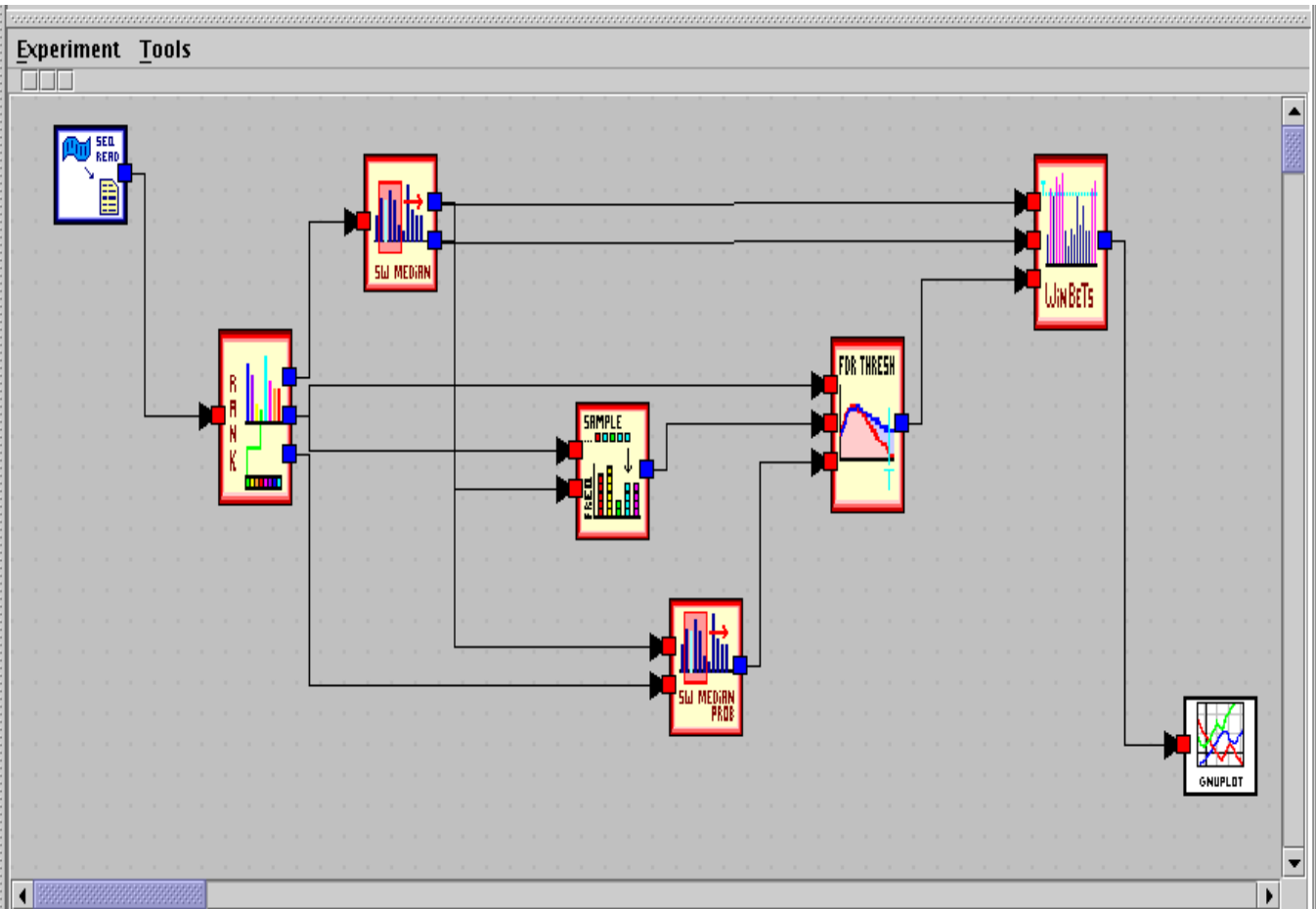
Human transcriptome map

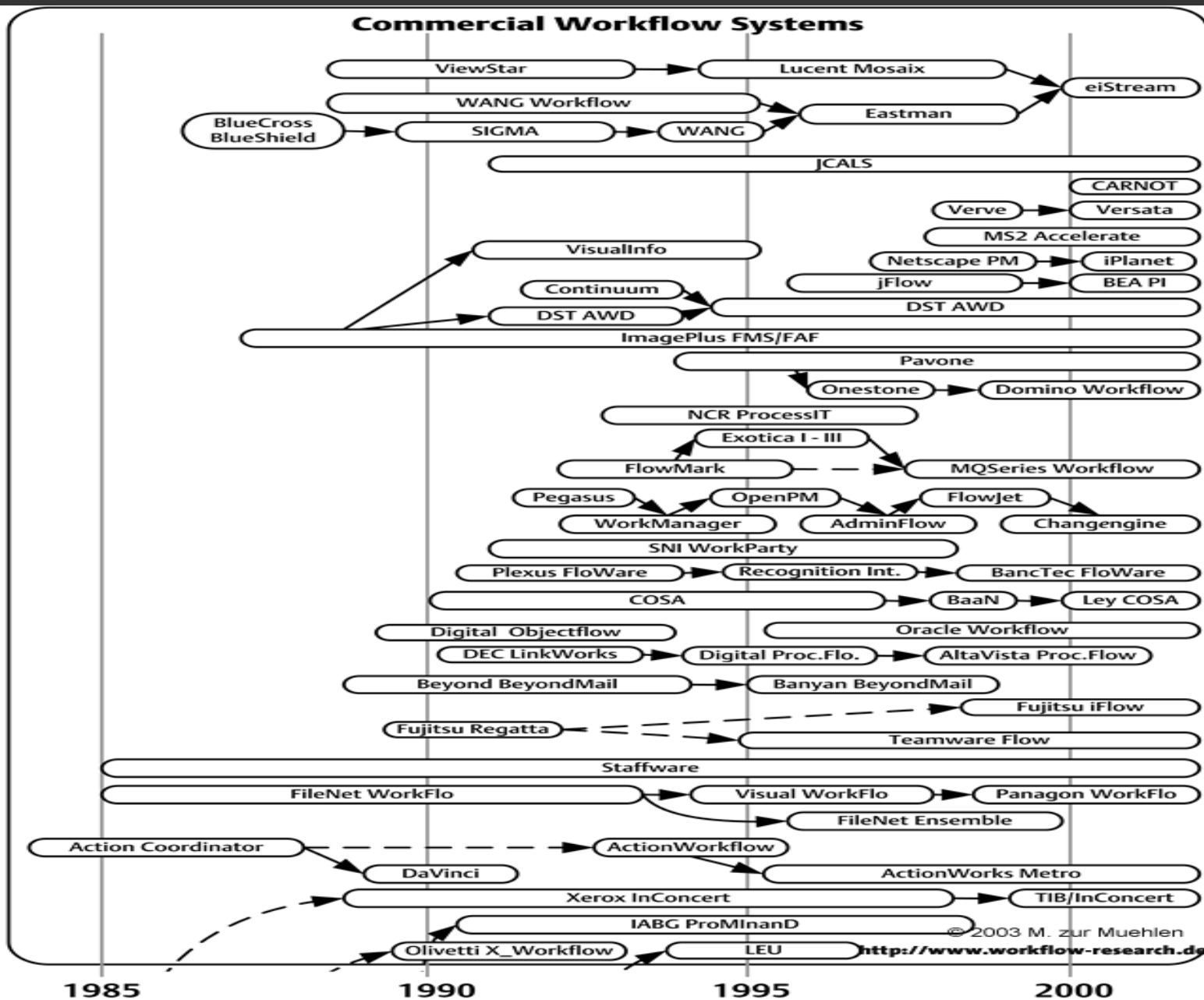


DNA curvature of the *Escherichia coli* chromosome



Temperature in Amsterdam





Scientific workflow management systems

- Askalon: <http://www.dps.uibk.ac.at/projects/askalon>
- Gridbus: <http://gridbus.csse.unimelb.edu.au/workflow>
- ICENI: <http://www.lesc.ic.ac.uk/iceni>
- Karajan: <http://www-unix.globus.org/cog/java>
- Kepler: <http://kepler-project.org>
- Pegasus: <http://pegasus.isi.edu>
- Taverna: <http://taverna.sourceforge.net>
- Triana: <http://www.trianacode.org>
- WS-VLAM: <http://www.science.uva.nl/~gvlam/ws-vlam>



The case for **Dynamic** workflows

- Different **alternatives** to proceed with the workflow and there is no way to decide which one is best at **design time**
- **Repeat** a sub-workflow until we get a certain condition is reached
- The **basic** structure or **semantics** changes. This may be because of an automated planner changing the workflow or optimized it on the fly based on prior execution knowledge.
- **Failure** in one part of the workflow may cause the whole workflow to fail if there is no alternative branch.

Dynamic workflow in a few words ...

- **WFMC** or the e-Science definition

+

- The ability to **adapt automatically** or **via user input** at **run time** to changes without **compromising**, business logic, performance, safety etc

Manual changes (User in the loop)

- User stepping in and changing the flow on the fly.
 - Can be achieved by providing computational **steering** capabilities
- “Computational steering is a valuable mechanism for scientific investigation in **parameters** of a running program can be **altered** and the results visualized immediately”



Automatic changes (Autonomic Behavior)

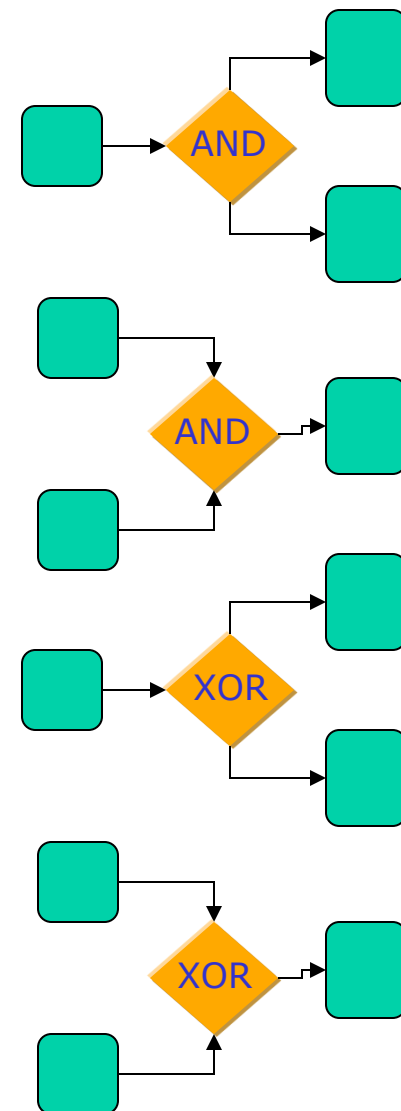
- The change of the flow is done on the fly by the system itself.
- Can be achieved by providing **control flow** capabilities

20 Control-flow patterns, are described in the workflow patterns page

www.workflowpatterns.com





Basic Control Patterns

- Parallel Split
 - execute activities in parallel
- Synchronization
 - synchronize two parallel threads of execution
- Exclusive Choice
 - choose one execution path from many alternatives
- Simple Merge
 - merge two alternative execution paths
- More control patterns can be found on the workflow patterns page



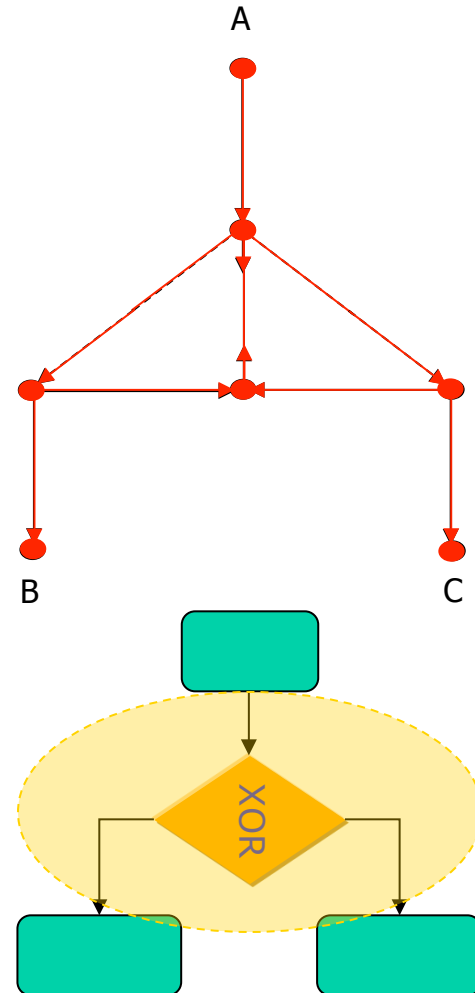
	If	Switch	loops
ApacheAnt	Y	Y	N
Askalon	Y	*	Y
DAGMAN	N	N	N
GrADS	N	N	N
Gridbus	N	N	N
ICENI	Y	*	Y
Karajan	Y	Y	Y
Kepler	Y	Y	N
Pegasus	N	N	N
Taverna	N	N	N
Triana	Y	N	Y
Unicore	Y	N	Y

REO approach

- *Reo* is an **exogenous** coordination language based on a calculus of **connector** composition.
- Atomic connectors are called **channels**.
- Every channel represents a primitive interaction (protocol), explicitly defined as a binary constraint.
- **Channel composition yields more complex interaction protocols**, represented as more complex constraints.
- *Reo* connectors are **dynamically reconfigurable**.
- Synchronous channel
 - write/take 
- Synchronous drain: two sources
 - write/write
- Synchronous spout: two sinks
 - take/take 
- Lossy synchronous channel
 - 
- Asynchronous FIFO1 channel
 - write/take 

Exclusive Router using Reo connector

- Synchrony and exclusion propagate through **synchronous segments** of a circuit.





Conclusion

- **Semantics** can play an important to support developing more dynamic workflow management systems
- **Interaction** with activities in workflow, steering workflows themselves and changing them during runs
- **How much** can or should be **automated**.

References

1. Workflow management coalition <http://www.wfmc.org/>
2. The *Workflow Patterns Initiative* <http://www.workflowpatterns.com>
3. Workflow Research www.workflow-research.de/Research/index.html
4. Composition by Anonymous Third Parties by Farhad Arbad, CWI <http://www.cwi.nl/htbin/reo/view>
5. Dynamic workflows and User Steering <http://vtcpc.isi.edu/wiki/images/e/e3/Dynamic1.pdf>

References

1. 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.
2. 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.
3. 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.
4. 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/>