

MoBiFlow: Principles and Design of a Workflow System for Molecular Biology

Markus Held, *Bundesanstalt für Finanzdienstleistungsaufsicht, Germany*

Wolfgang Küchlin, *Eberhard Karls Universität Tübingen, Germany*

Wolfgang Blochinger, *Universität Stuttgart, Germany*

ABSTRACT

Web-based problem solving environments provide sharing, execution and monitoring of scientific workflows. Where they depend on general purpose workflow development systems, the workflow notations are likely far too powerful and complex, especially in the area of biology, where programming skills are rare. On the other hand, application specific workflow systems may use special purpose languages and execution engines, suffering from a lack of standards, portability, documentation, stability of investment etc. In both cases, the need to support yet another application on the desk-top places a burden on the system administration of a research lab. In previous research the authors have developed the web based workflow systems Calvin and Hobbes, which enable biologists and computer scientists to approach these problems in collaboration. Both systems use a server-centric Web 2.0 based approach. Calvin is tailored to molecular biology applications, with a simple graphical workflow-language and easy access to existing BioMoby web services. Calvin workflows are compiled to industry standard BPEL workflows, which can be edited and refined in collaboration between researchers and computer scientists using the Hobbes tool. Together, Calvin and Hobbes form our workflow platform MoBiFlow, whose principles, design, and use cases are described in this paper.

Keywords: Bioinformatics, Biological Problem Solving Environments (PSEs), Molecular Biology, Web-Based Problem Solving Environments, Workflow Management Systems

INTRODUCTION

Biological Problem Solving Environments (PSEs) are frequently provided as web applications. In the trivial case, such a PSE provides a single service or a limited set of services as a web application. Biologists “connect” these services manually by cutting and pasting data between web sites. Since this process is error-

prone and hard to retrace, a plethora of different bioinformatics workflow systems have emerged (Taylor, Deelman, Gannon, & Shields, 2006; (Tiware, & Sekhar, 2007). Increasingly, web portals are used to access and to share scientific workflows (De Roure, Goble, & Stevens, 2007; Christie & Marru, 2007). Some portals provide workflow construction facilities via Java Web Start (Christie & Marru, 2007; Sipos & Kacsuk, 2005), while true web-based workflow construction tools lack sophisticated user experience

DOI: 10.4018/jssmet.2011100107

(Carrere & Gouzy, 2006; Bartocci, Corradini, Merelli, & Scortichini, 2007). Often, scientific workflow management systems are provided as desktop applications with rather complex user interfaces (Oinn et al., 2004; Shah et al., 2004).

Similar to programming languages, the complexity of the user interface and the workflow notation remains an area of conflict. Supporting a large set of different workflows and different types of services increases the complexity of a workflow language, thus reducing usability by domain experts. In contrast, high user-friendliness limits the possible set of workflow patterns and access to arbitrary services. Domain experts should be able to compose and execute workflows in a domain specific modeling system, and still fall back to a collaboration with software engineers if a more complex workflow model is needed. In this paper we propose a hybrid approach of augmenting an existing collaborative workflow development system with a biology-specific scientific workflow mode. In software engineering it is often recognized that 80% of the useful functionality is provided by 20% of the code (the “Pareto principle”).

We argue that many typical life science workflows can be expressed by a reduced workflow notation. Similar to high-level programming languages, which are compiled to a common machine language, we regard the Business Process Execution Language (BPEL) as the “assembly language” of an arbitrary set of domain-specific workflow languages (OASIS, 2007).

In this article, we give an overview of our previous work on the bioinformatics workflow system MoBiFlow and the software ecosystem on which it is built. The article extends our workshop paper (Küchlin & Held, 2010). MoBiFlow consists of the biology-specific workflow system Calvin and the collaborative workflow development system Hobbes, with additional tools such as the computation of workflow metrics, layout capabilities or Web 2.0 based video conferencing.

In particular, we make the following contributions:

1. We have formalized common e-science processes in biology as a meta-process and we derive requirements for next generation bioinformatics workflow systems.
2. We show how low-level workflow languages can be combined with domain-specific workflow notations to provide ease-of-use and flexibility at the same time.
3. We describe the design principles behind the MoBiFlow system, which represents a solution for high-level and low-level collaborative workflow development and usage.

REQUIREMENTS FOR E-BIOLOGY WORKFLOW SYSTEMS

Computer Use and Infrastructure in Biology Departments

In this section we analyze contemporary e-science usage in biology departments, based on the example of the Center for Plant Molecular Biology (ZMBP) of the University of Tübingen. ZMBP consists of fifteen research groups in four branches. We have cooperated with researchers from the plant physiology branch, which consists of seven research groups.

We held several meetings and interviews with ZMBP research staff to evaluate their requirements for a bioinformatics workflow system. In an agile fashion, we gave demonstrations of our evolving workflow system to staff members with differing levels of expertise in e-science.

In the ZMBP, “*in silico* experiment” web applications are frequently used to prepare “wet” experiments in the laboratory. Existing workflow solutions have been evaluated but failed to satisfy the users, especially due to intricate user interfaces and lack of user interaction. Another important factor for the acceptance of novel software is its availability on the web. On the one hand, biologists are used to work on the web, as described above. On the other hand, in large institutions, departmental poli-

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/mobiflow-principles-design-workflow-system/60736

Related Content

Flexible Provenance Tracing

Liwei Wang, Henning Koehler, Ke Deng, Xiaofang Zhou and Shazia Sadiq (2011). *International Journal of Systems and Service-Oriented Engineering* (pp. 1-20). www.irma-international.org/article/flexible-provenance-tracing/55120

Talent Management Practices and Competitive Advantage: Evidence From the Jordanian Insurance Sector

Yazan Emnawer Al-Haraisa, *Noor Al-Ma'aitah, Khalaf Al-Tarawneh and Ahmad Abuzaid (2021). *International Journal of Service Science, Management, Engineering, and Technology* (pp. 102-115). www.irma-international.org/article/talent-management-practices-and-competitive-advantage/284874

Assistive Technology For Night Blind Tourist: A Mobile Application Approach

ObaidUllah Owais Khan, Kamran Ahsan and Abdul Salam (2014). *International Journal of E-Services and Mobile Applications* (pp. 43-65). www.irma-international.org/article/assistive-technology-for-night-blind-tourist/119704

The Transformative Impact of Contactless Technology on Hospitality: Experiences, Efficiency, and Safety

Varinder Singh Rana, Ashish Raina, Gaurav Bathla, Lojain Aly and Mahendra Khairiya (2026). *Future of Contactless Technology in Hotels and Restaurants* (pp. 113-130). www.irma-international.org/chapter/the-transformative-impact-of-contactless-technology-on-hospitality/398041

Employee Self-Service Portals

Beverly Lloyd-Walker (2010). *Electronic Services: Concepts, Methodologies, Tools and Applications* (pp. 1368-1375). www.irma-international.org/chapter/employee-self-service-portals/44020