Wf4Ever: Advanced Workflow Preservation Technologies for Enhanced Science

STREP FP7-ICT-2007-6 270192

Objective: ICT-2009.4.1b — "Advanced preservation scenarios"

D2.2v1 Design, implementation and deployment of workflow lifecycle management components - Phase II

Deliverable Co-ordinator: **XX**

Deliverable Co-ordinating Institution: **University of Manchester**

Other Authors: **XX**

This deliverable describes the second phase of delivery of workflow lifecycle management components. It includes a description of the Research Object Model, which facilitates interoperation between components; an initial Research Object Storage and Retrieval Service; RO Manager command line tool; and a definition of a model for workflow abstraction.

Document Identifier:	Wf4Ever/2012/D2.2v1/0.1	Date due:	July 31, 2013
Class Deliverable:	Wf4Ever FP7-ICT-2007-6 270192	Submission date:	June 1, 2013
Project start date	December 1, 2010	Version:	0.1
Project duration:	3 years	State:	Draft
		Distribution:	Public

Wf4Ever Consortium

This document is part of the Wf4Ever research project funded by the IST Programme of the Commission of the European Communities by the grant number FP7-ICT-2007-6 270192. The following partners are involved in the project:

Intelligent Software Components S.A. (ISOCO) – Coordinator Edificio Testa, Avda. del Partenón 16-18, 1^o, 7^a Campo de las Naciones, 28042 Madrid Spain Contact person: Jose Manuel Gómez Pérez E-mail address: jmgomez@isoco.com	**University of Manchester (UNIMAN)** School of Computer Science Oxford Road, Manchester M13 9PL United Kingdom Contact person: Carole Goble E-mail address: carole.goble@manchester.ac.uk
Universidad Politécnica de Madrid (UPM) Departamento de Inteligencia Artificial, Facultad de Informática. 28660 Boadilla del Monte. Madrid Spain Contact person: Oscar Corcho E-mail address: ocorcho@fi.upm.es	**Instytut Chemii Bioorganicznej PAN - Poznan Supercomputing and Netowrking Center (PSNC)** Network Services Department Ul Z. Noskowskiego 12-14 61704 Poznań Poland Contact person: Raul Palma E-mail address: rpalma@man.poznan.pl
University of Oxford (OXF) Department of Zoology South Parks Road, Oxford OX1 3PS United Kingdom Contact person: Jun Zhao, David De Roure E-mail address: jun.zhao@zoo.ox.ac.uk david.deroure@oerc.ox.ac.uk	**Instituto de Astrofísica de Andalucía (IAA)** Dpto. Astronomía Extragaláctica. Glorieta de la Astronomía s/n, 18008 Granada Spain Contact person: Lourdes Verdes-Montenegro E-mail address: lourdes@iaa.es
Leiden University Medical Centre (LUMC) Department of Human Genetics Albinusdreef 2, 2333 ZA Leiden The Netherlands Contact person: Marco Roos E-mail address: M.Roos1@uva.nl	

Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed to the writing of this document or its parts:

- iSOCO
- OXF
- PSNC
- UNIMAN
- UPM

Change Log

Version	Date	Amended by	Changes
0.1	01-07-2013	Khalid Belhajjame	Initial outline

Executive Summary

This deliverable describes the second phase of delivery of workflow lifecycle management components. These components are focused around the Wf4Ever Research Object Model (RO Model), which provides descriptions of workflow-centric ROs – aggregations of content. This model is used to structure and describe ROs which are then stored and manipulated by the components of the Wf4Ever Toolkit.

The RO Model provides a framework for describing aggregations of content along with annotations of the aggregated resources, a vocabulary for describing workflows, and a vocabulary for describing provenance. The model underwent few changes in the last year in the light of user comments. We provide here a summary of the new version of the RO model. We also present the components developed for creating and managing Research Objects: the Research Object Storage and Retrieval API (implemented as part of the Research Object Digital Library (RODL)) and a command line tool – the RO Manager. These components and services are also discussed in D1.2v3 (Wf4Ever Sandbox – Phase II), D1.3v1 (Wf4Ever Architecture – Phase II) and D1.4v1 (Reference Wf4Ever Implementation – Phase II).

One of the main development in the last year consists in incorporating research objects within the myExperiment environment to allow scientists who already use myExperiment to create, share and reuse research objects. We discuss the efforts that went into this task, and report on an activity that we conducted to convert all existing Taverna T2 workflows into ROs.

We present advanced management functions that we developed for abstracting and indexing workflows, with the aim of supporting the discovery and reuse of workflows. We present an ontology that we developed for abstracting workflows in terms of motifs that characterize data manipulation and transformation patterns, which we term motifs. We also report on a solution that we developed for indexing workflows based on the services (processes) that they use.

This deliverable should be read in tandem with D1.3v2 (Wf4Ever Architecture – Phase II), D1.4v2 (Reference Wf4Ever Implementation – Phase II), D1.2v3 (Wf4Ever Sandbox – Phase III), D3.2v2 (Design, implementation and deployment of Workflow Evolution, Sharing and Collaboration components – Phase II) and D4.2v2 (Design, implementation and deployment of Workflow Integrity and Authenticity Maintenance components – Phase II) in order to provide a complete picture of the state of the Wf4Ever Phase II components.

Contents

List of Tables

List of Figures

1 Introduction

2 The Research Object Model

This section present the RO ontologies, `ro`, `wfdesc`, `wfprov`, `wf4ever`. In doing so, we will use a UML class diagram to illustrate the classes and properties of such ontologies.

3 Research Object Storage and Retrieval

This section presents the components that constitute the RODL, using a UML class diagram, and show how the user can utilize RODL using a UML sequence diagram.

4 Research Object Manager

This section presents the RO manager architecture, and presents the functionalities it provides using a UML sequence diagram, if that is plausible.

5 Research Object-Enabled myExperiment

This section describes the efforts that went into incorporating research objects within myExperiment. In particular, how the notion of myExperiment pack was used as a starting point to incorporate new features/functionalaities. We will also discuss the diferent iterations that involved Wf4ever and Biovel users in those developments.

6 Workflow Abstraction using Motifs

This section presents the motif ontology, again using a UML class diagram, and provides an example of a workflow that was annotated using the motifs.

7 Workflow Indexation

This section shows how workflows can be indexed using the trie structure. It presents the approach as well as an example workflow that is indexed.

References