Accepted Manuscript

Abstract, link, publish, exploit: An end to end framework for workflow sharing

Daniel Garijo, Yolanda Gil, Oscar Corcho

PII: S0167-739X(17)30027-4
DOI: http://dx.doi.org/10.1016/j.future.2017.01.008
Reference: FUTURE 3290

To appear in: *Future Generation Computer Systems*

Received date: 16 May 2016
Revised date: 1 October 2016
Accepted date: 7 January 2017

Please cite this article as: D. Garijo, Y. Gil, O. Corcho, Abstract, link, publish, exploit: An end to end framework for workflow sharing, *Future Generation Computer Systems* (2017), http://dx.doi.org/10.1016/j.future.2017.01.008

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Abstract, Link, Publish, Exploit: An End to End Framework for Workflow Sharing

Daniel Garijo[1], Yolanda Gil[1] and Oscar Corcho[2]

[1] Information Sciences Institute and Department of Computer Science
University of Southern California
{dgarijo,gil}@isi.edu
[2] Facultad de Informática, Universidad Politécnica de Madrid
ocorcho@fi.upm.es

Abstract

Scientific workflows are increasingly used to manage and share scientific computations and methods to analyze data. A variety of systems have been developed that store the workflows executed and make them part of public repositories. However, workflows are published in the idiosyncratic format of the workflow system used for the creation and execution of the workflows. Browsing, linking and using the stored workflows and their results often becomes a challenge for scientists who may only be familiar with one system. In this paper we present an approach for addressing this issue by publishing and exploiting workflows as data on the Web, with a representation that is independent from the workflow system used to create them. In order to achieve our goal, we follow the Linked Data Principles to publish workflow inputs, intermediate results, outputs and codes; and we reuse and extend well established standards like W3C PROV. We illustrate our approach by publishing workflows and consuming them with different tools designed to address common scenarios for workflow exploitation.

1. INTRODUCTION

Scientific workflows define the "set of tasks needed to manage a computational science process" [Deelman et al., 2009]. They have been used successfully in several domains [Ruiz et al., 2014, Dinov et al., 2009, Wolstencroft et al., 2013] in order to represent, execute, re-run, debug, document and reproduce scientific methods. Scientific workflows are important products of scientific research which should be treated as first-class citizens in cyber-infrastructure [Gil et al., 2007].

There has been great interest in the publication of workflows, particularly to enable reproducibility and shareability of scientific experiments. There are a number of frameworks that allow the inclusion of workflows and codes in scientific articles [Leisch 2007, Mesirov 2010, Falcon 2007]. Workflow repositories like myExperiment [De Roure et al., 2009] and CrowdLabs [Mates et al., 2011] provide mechanisms to publish and search workflows. These repositories support the publication of workflows in their original language specifications. Given the proliferation of workflow systems

Download English Version:

https://daneshyari.com/en/article/4950418

Download Persian Version:

https://daneshyari.com/article/4950418

Daneshyari.com