



Scuf12 – because a workflow is more than its definition

Document Version

Other version

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Soiland-Reyes, S., Williams, A., Owen, S., Withers, D., & Goble, C. (2011). Scuf12 – because a workflow is more than its definition. In *host publication* (pp. 37-37) <http://www.slideshare.net/soilandreyes/2011-0716-scuf12-because-a-workflow-is-more-than-its-definition-bosc-2011>

Published in:

host publication

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.





Scufl2 - because a workflow is more than its definition

The University of Manchester

Project site:
<http://www.taverna.org.uk/>

Source code:
<https://github.com/mygrid/scufl2>
<http://taverna.googlecode.com/>

License: GNU Lesser General Public License (LGPL) 2.1

Stian Soiland-Reyes, Alan R Williams, Stuart Owen, David Withers and Carole Goble
School of Computer Science, University of Manchester, UK

{stian.soiland-reyes, alan.r.williams, stuart.owen, david.withers, carole.a.goble}@manchester.ac.uk

Alpha

- Taverna's future workflow and data format
- Goal: Simplify third-party reading, writing, annotating and extending
- Scufl2: specification, schema, ontology, Java API and conversion tool

"Structured" zip-file (can be unpacked to be exposed on the web)

Self-documenting media type (OpenOffice ODF, ePub OCF, Adobe UCF) - for tools like file(1) and mime magic

Root file: Primary document in the bundle (ePub OCF/Adobe UCF). Alternative representations of same workflow bundle allowed (Turtle, JSON, HTML, etc)

Unique identifier which can be used as prefix for all relative references in bundle

Suggests main workflow and main profile, but executor could (e.g. by parameter) run a different workflow or profile

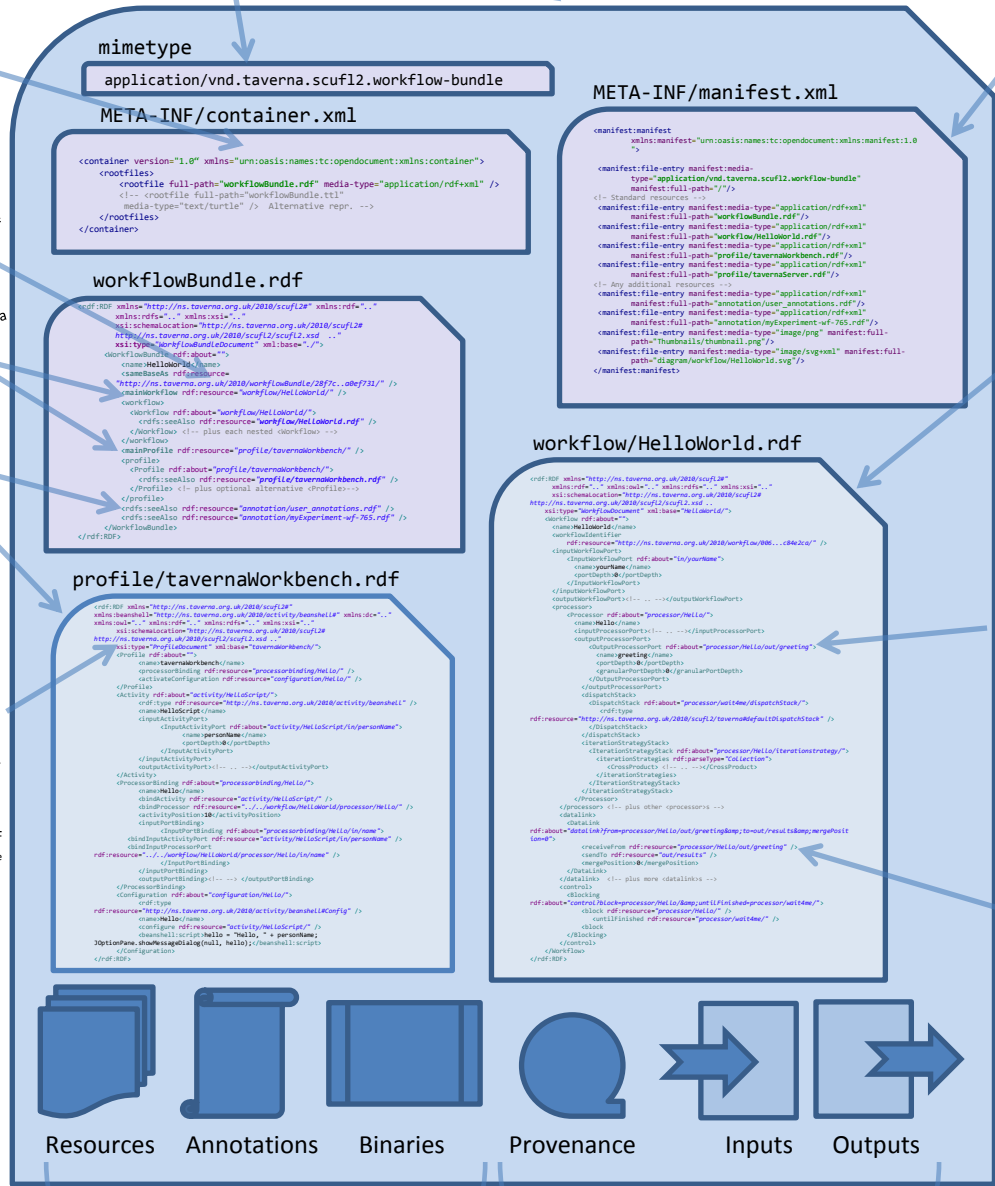
Additional (non-executable) annotations and metadata

Profile gives implementation bindings, alternative profiles can customize execution of workflow steps for different environments (e.g. desktop, server, cloud)

Hybrid of RDF/XML and XML schema - If the xsi:type is given, documents can be parsed and generated as regular XML, using xpath, etc.

The schema ensures the document can still be parsed as RDF/XML as well. Pure RDF writers can omit the xsi:type

the-workflow-bundle.scufl2



Manifest listing all resources in bundle with media types

Definition of workflow structure. Nested workflows are separate resources.

Every workflow part has a URI, allowing deep annotations

Relative references can also be made absolute using the sameBaseAs prefix, RDF clients resolving such URIs at http://ns.taverna.org.uk/2010/workflowBundle/ can be redirected to a generated RDF resource stating what's obvious from the URI.

"Cool URI"-style relative references allow parsers to 'cheat' and pick out processor 'Hello' and port 'greeting' - but only if starts with the keyword paths out/ in/ or processor/

Resources Annotations Binaries Provenance Inputs Outputs

The bundle (as a ZIP file) can be extended to include arbitrary resources. These could be referenced internally by workflow activities, it could be PDFs, spreadsheets, etc.

Wf4Ever: A Research Object (RO) captures enough data and provenance about results to make data and methods reproducible, verifiable, shareable, reusable and repeatable. The Scufl2 workflow run bundle forms such an RO for Taverna workflow results - with the goal of becoming an executable paper.

Embedding provenance, input and output data of a workflow run can provide example use and expected outputs. Changing the root-file to point to the provenance makes the bundle a workflow run bundle, where the workflow definition is secondary.

Similarly a data bundle represents primarily the data, but by including the workflow and the provenance also embed information on how it was generated.



<http://www.wf4ever-project.org/>



<http://www.mygrid.org.uk/>