

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

<https://github.com/mygrid/scuf12>
<http://taverna.googlecode.com/>

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

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

- Taverna's future workflow and data format
- Goal: Simplify third-party **reading, writing, annotating** and **extending**
- Scuf2: specification, schema, ontology, Java API and conversion tool

the-workflow-bundle.scuf12

mimetype

META-INF/manifest.xml

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/*

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 `xmlns:xsi` 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 `xmlns:xsi`

profile/tavernaWorkbench.rdf

workflow/HelloWorld.rdf

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 Scufi2 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.

Workflow

4Ever

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

myGrid

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

Funded by: EPSRC EP/G026238/1; European Commission's 7th FWP FP7-ICT-2007-6 270192; FP7-ICT-2007-6 270137