



Other SCAPE outputs available

William Palmer, Peter May & Carl Wilson

SCAPE Information Day

British Library, UK, 14th July 2014

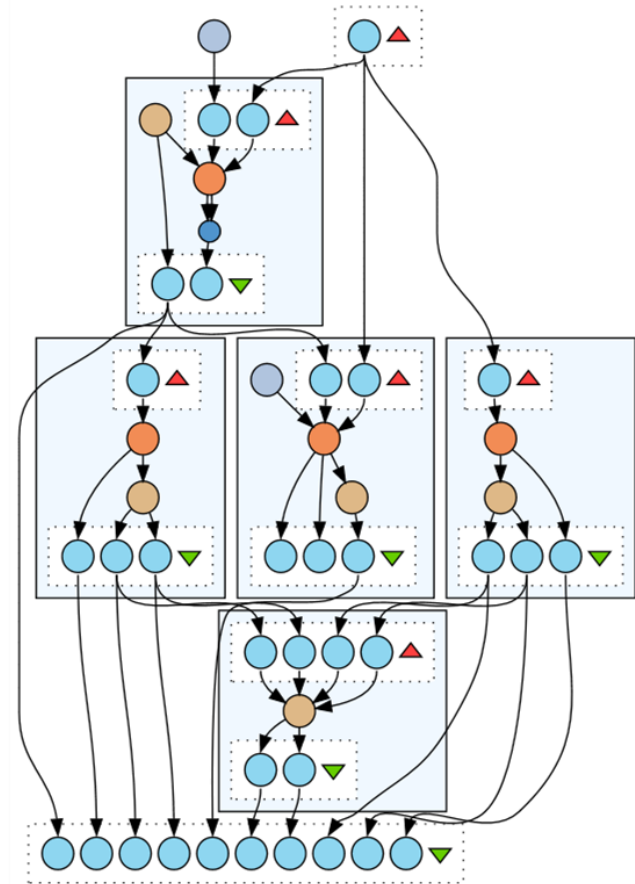
- Digital Repository
- RODA: Repository of Authentic Digital Objects
 - Delivers functionality of the main units of the OAIS model
 - Uses Fedora Commons (2/3) /EAD/PREMIS/...
- Fedora 4 (in development, beta status)
 - Improved scalability
 - Flexible storage options
 - Better linked-data capabilities
- RODA & Fedora 4 implement standard SCAPE APIs for other tools/services to interact with a repository

- Preservation Planning
- Walks you through making preservation plans
- Use components^[1] from Component Catalogue to test/design preservation workflows
- For example, automatically design an image migration workflow

[1] <http://openplanets.github.io/scape-component-profiles/>

Plan: Image migration to TIFF

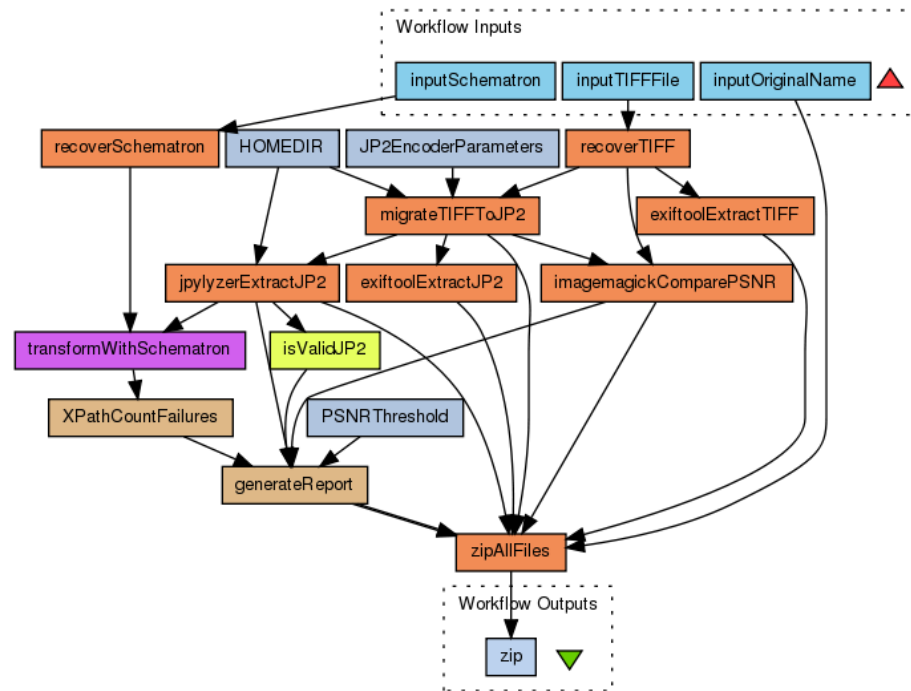
Imagemagick - image/* ~ image/tiff



- “Preservation Watch” service
- Ideally use a central instance, aggregating preservation watch information from several repositories
- Detect risks and opportunities
- Uses plugins for easy integration of new information sources
- Users can create triggers for when risks or opportunities occur
- Notifies users when events occur such as format obsolescence, or new preservation tools are available
- Compare repository information against others’



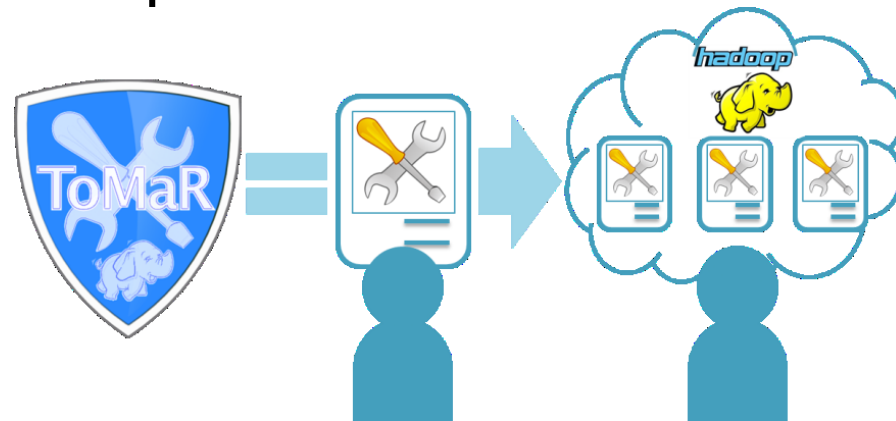
- An open-source Workflow Management System
- Graphically design and execute workflows*
 - *command line and server execution environments available
- Integrates with the Component Catalogue at MyExperiment.org
- ...



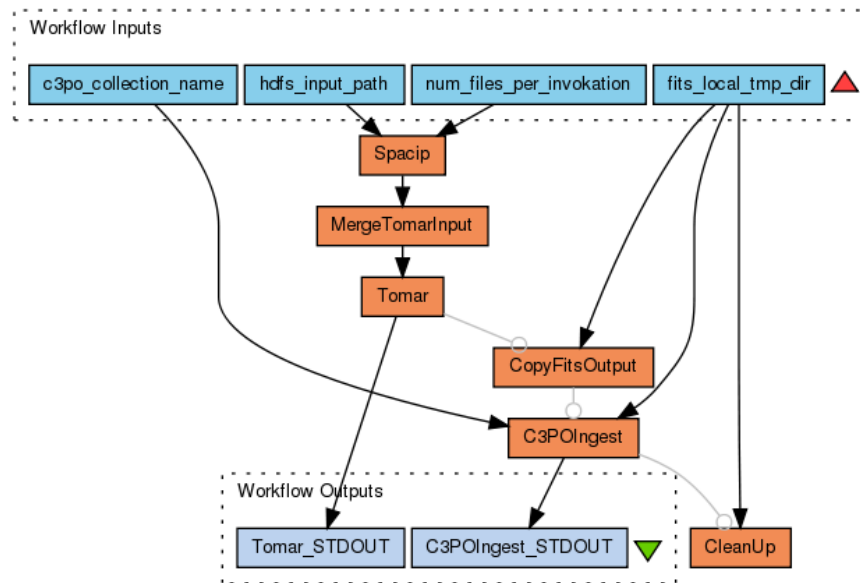
- Uses SCAPE tool specification files “toolspecs”
- Standardise tool invocation to standard type:
 - QA / Characterisation/ Migration / etc
- Create Debian packages for standardised invocation
- Create Taverna Components for upload to the Component Catalogue (on MyExperiment)
- Component Profiles:
<http://www.myexperiment.org/tags/3693.html>
- Components:
<http://www.myexperiment.org/tags/3214.html>

ToMaR - Demo

- “To MapReduce”: wrap command line applications for parallel execution in a Hadoop MapReduce job
- (Re)uses SCAPE tool specification “toolspecs” (as used by the ToolWrapper)
- Can parallelise standalone tools that cannot otherwise be integrated within Hadoop/MapReduce
- Can be used to run a step in a Taverna workflow (see demo)

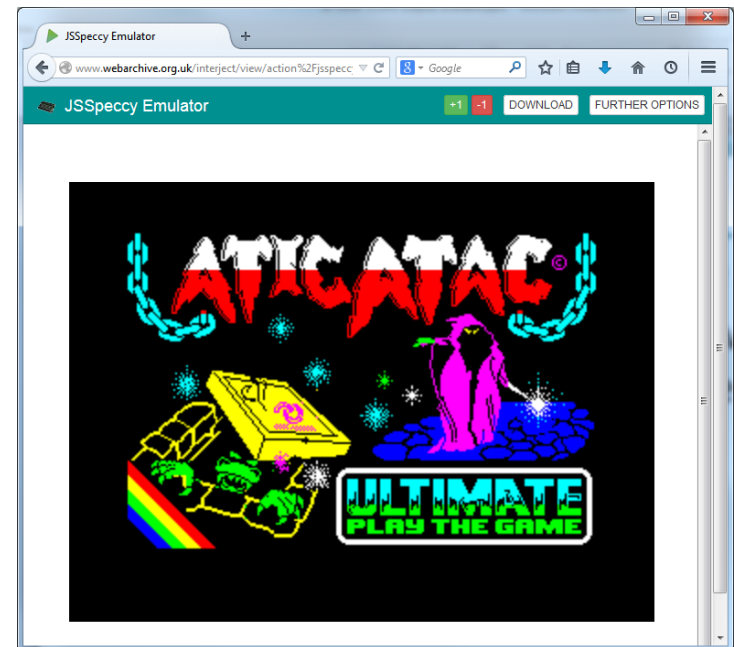


- Hadoop based web archive record processing
 - ARC to WARC migration
 - Identify files with Droid
 - Identify files with Tika
- Different approach to identification/characterisation from Nanite



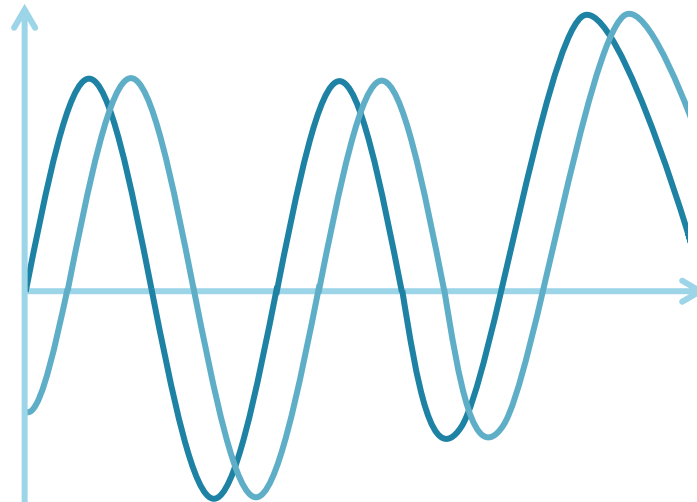
Interject - Demo

- Sits on top of the UK Web Archive
- Some [older] content may not be viewable with contemporary software/hardware
- Therefore, provide:
 - Emulation-on-access/demand
 - Migration-on-access/demand
- Rate the access: +1/-1 buttons
- Potentially use ratings to inform future preservation actions



- Compares web pages using two approaches:
 - Visual
 - Structural (HTML etc)
- Similarity scores are produced
- Can be used for QA, comparing archive copy with live site

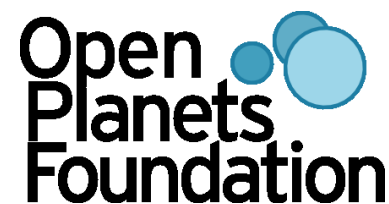
- Suite of tools for:
 - Detecting overlap in two audio files
 - Finding occurrences of a smaller WAV in a bigger one (and pre-index files for analysis)
 - Find the similarity between two audio files



SCAPE tools are published as open source software.

Tools and services from SCAPE are sustained by

- Open Planets Foundation -
address core digital preservation
challenges and engage with the community



- COPTR -
Community Owned digital Preservation
Tool Registry

