

WfExS possible RO-Crate profiles

WorkflowStagingProvenance

Should it align to workflow testing description or result?

Workflow profile plus

- Input
 - Type of input: file, directory, value (atomic value type?)
 - Identifier / URI or CURIE-represented PID
 - Digest? (Naming Things with Hashes rfc6920)
 - Symbolic name in workflow (with dots for nested ones)
 - Cardinality: one, many
 - Resolved / canonical URL used to fetch (without credentials)
 - Controlled access?
 - Metadata????? (contentType, encodingFormat, ..)
- Expected output
 - Type of expected output: file, directory, value
 - Cardinality: one, many
 - Symbolic name in the workflow (if supported by workflow language) or in this profile.
 - Pattern to match output files (if workflow language does not support describing outputs).
- Workflow repository
 - Type of workflow repository: TRS, RO-Crate, git repository, other?????
- Workflow
 - Absolute URL
 - Identifier
 - Relative path
 - Symbolic Revision or commit
 - Effective revision or commit (with hashtag or fingerprint)
 - Type (CWL, Nextflow, etc.)
 - Symbolic Engine restriction
 - Effective Engine version
- Abstract workflow representation
 - In CWL
 - In GraphML / dot / PNG SVG
- Container technology
 - Type of container technology: Docker, Singularity, Podman, etc...
- Container
 - Link to container technology

- Symbolic name (URI format)
 - Canonical name (with hashtag or fingerprint)
- Licence of this RO-Crate.
- Attribution of this RO-Crate.
- Recommended Licence for results when it is going to be executed.
- Recommended Licence for metadata (provenance RO-Crate) when it is going to be executed.
- When was staged
- Who staged it?
- Staging environment?

WorkflowExecutionProvenance

Should we use different datasets to group different kind of materialized knowledge? (ro-crate metadata, root/dataset/contextual entity)

Should it align to workflow testing result?

WorkflowStagingProvenance profile plus

- MaterializedInput
 - Link to Input (when it is an explicit input)
 - Is an implicit input?
- Generated output
 - Link to expected output
 - Identifier / URI or CURIE-represented PID
 - Digest? (Naming Things with Hashes rfc6920)
 - Resolved / canonical URL used to push (without credentials)
 - Controlled access?
 - Metadata????? (contentSize, encodingFormat, ..)
- Workflow
 - programming language
- Abstract workflow execution representation
 - In CWL
 - In GraphML / dot / PNG SVG
- Used Container technology (could differ from the expected one)
- Fetched Container
 - Link to used container technology
 - Link to container
 - Canonical name (with hashtag or fingerprint)
- Workflow execution
 - When was executed
 - Technical monitoring: real time, cpu time, memory peaks, working directory disk usage.
 - Did it finish properly?

- Execution restrictions: max memory, max cores, max real time, ...
- Native workflow execution reports:
 - Linked workflow execution.
 - Native log file, native report, ...
- Workflow step
 - Linked workflow execution
 - Used container
 - Technical monitoring: number of repeats of the step, memory, CPU, disk usage of each one of the steps.
- Attribution
 - Identifier of who launched the execution.
 - Digital Signature?