Publicação de Proveniência em Workflows na Web Semântica

Rachel Castro¹, Renan Souza¹, Vítor Silva¹, Kary Ocaña², Daniel de Oliveira³ e Marta Mattoso¹



¹COPPE-Universidade Federal do Rio de Janeiro (UFRJ) ²Laboratório Nacional de Computação Científica (LNCC) ³Universidade Federal Fluminense (UFF)



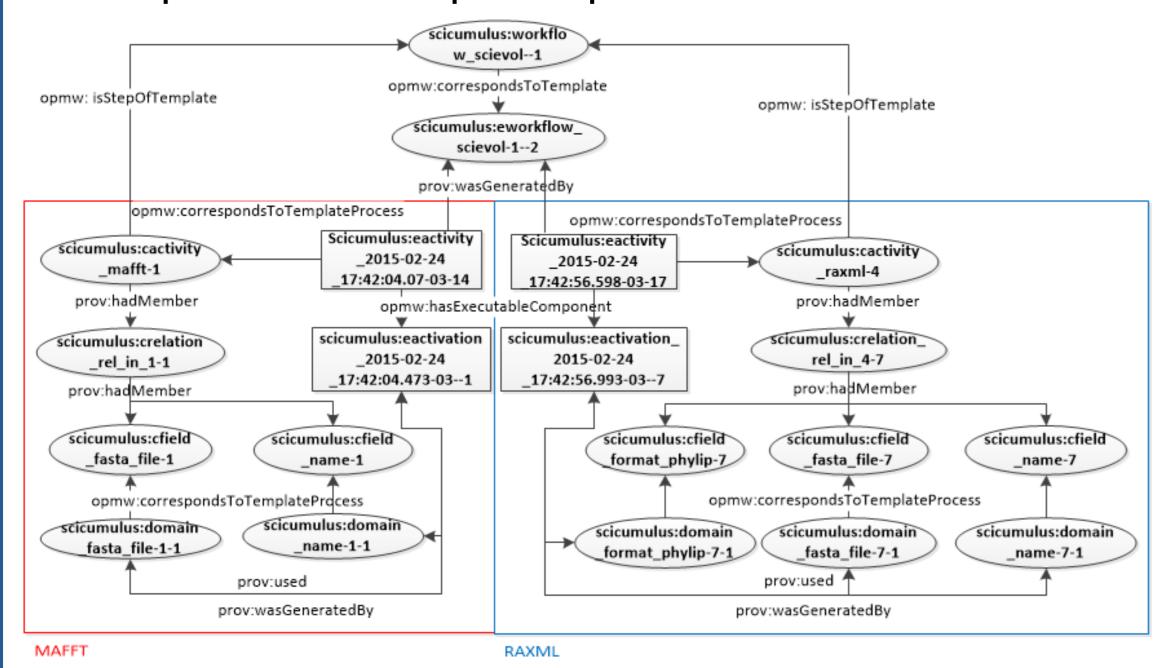
rachelcastro@poli.ufrj.br, {renanfs, silva, marta}@cos.ufrj.br, kary@lncc.br , danielcmo@ic.uff.br

Introdução

- Dados de proveniência de *workflows*:
 - √ Garantem a reprodução de simulações computacionais
 - ✓ Comunmente armazenada com restrição ao acesso
 - ✓ Publicação vantajosa → expansão da capacidade analítica
- Metodologia de publicação na Web Semântica

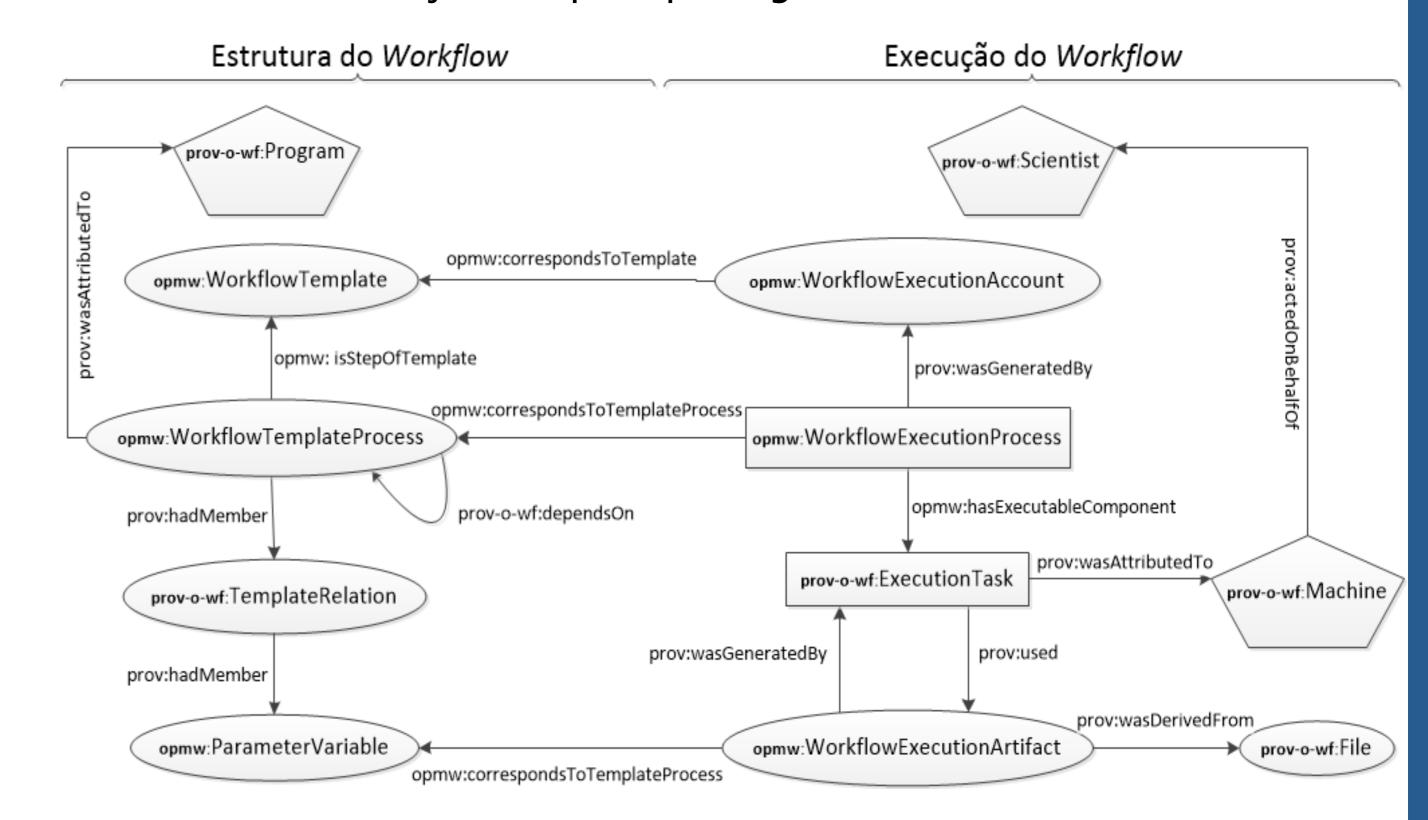
Estudo de Caso

- SGWfC paralelo: SciCumulus
- Dados de proveniência do workflow SciEvol:
 - ✓ Treze atividades constituintes
- Repositório de Triplas: Open Link Virtuoso



Ontologia Proposta: PROV-O-Wf

- PROV-O utilizado como metamodelo
- Reutilização das classes do OPMW
- Propriedades de objeto:
 - ✓ Definidas na OPMW e na PROV-O
 - ✓ dependsOn → dependência entre atividades
- Diferencial:
 - ✓ Abrange SGWfC que apoiam execuções em paralelo
 - ✓ Identificação dos principais agentes

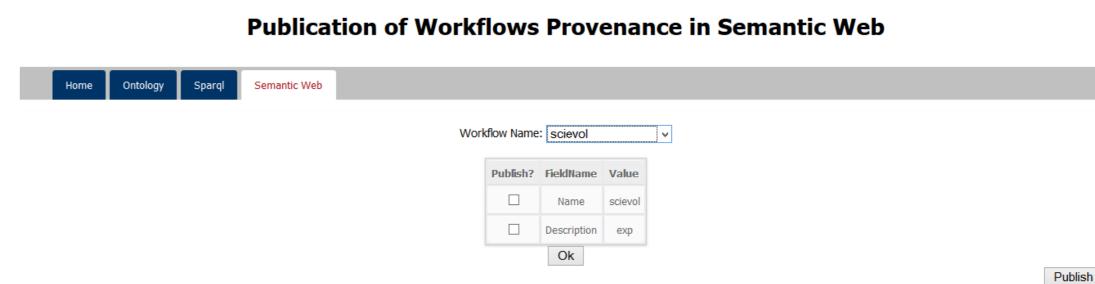


Análise do Domínio, Engenharia da Ontologia, Triplificação e Publicação

- Análise do Domínio:
 - ✓ Simulações computacionais modeladas como workflows, executadas em ambiente de PAD
 - ✓ Coleta de proveniência através de SGWfC paralelos
 - ✓ Domínio: dados de proveniência de workflows gerados por um SGWfC capaz de coletar dados em granularidade fina
- Engenharia da Ontologia:
 - ✓ Reutilização de ontologias como característica importante
 - ✓ Ontologias IntelLeo e MyGrid → não utilizadas
 - ✓ PROV-O e OPMW: independência de domínio e padrões W3C
 - ✓ Modelo de dados PROV-Wf
- Triplificação e Publicação:
 - ✓ Dados de proveniência armazenados em um SGBDR → processo ETL

Domain Analysis Data Selection Planning Domain Understanding Enable Public Access Public SPARQL Endpoint Publication & Applications Visahups Visualizations Visahups Visualizations Parameter of the Domain Domain Analysis Ontology Engineering Inferences Machine Readable Graph of Triples Databases Linkage to the LOD cloud Triplification & Linkage ETL RDF HTTP URIS

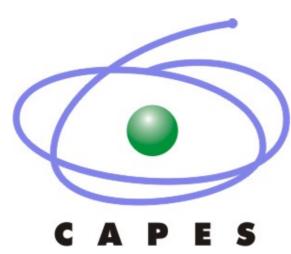




Conclusão

- Desenvolvimento de ontologia:
 - ✓ Reutiliza ontologias bem difundidas
 - ✓ Considera dados de proveniência de uma granularidade fina
 - ✓ É independente de domínio e de SGWfC
 - ✓ Abrange SGWfC paralelos e que executam em ambiente de PAD







Referências

Costa, F., Silva, V., Oliveira, D., Ocaña, K., Ogasawara, E., Dias, J. and Mattoso, M. (2013) "Capturing and Querying Workflow Runtime Provenance with PROV: a Practical Approach", In: Proceeding of the Joint EDBT/ICDT 2013 Workshops, New York, USA.

Souza, R., Cottrell, L., White, B., Campos, M. L. and Mattoso, M. (2014) "Linked Open Data Publication Strategies: Application in Networking Performance Measurement Data", In: 2nd ASE International Conference on Big Data Science and Computing, Stanford, CA, USA.

Moreau, L., Missier, P. and Belhajjame, B. (2013) "The PROV Data Model and Abstract Syntax Notation".