

# Ontologias e Proveniência em *Workflows* Científicos

---

Luiz Gustavo Dias

Universidade Federal Fluminense  
Instituto de Computação  
Pós-Graduação em Computação  
E-science

# Agenda

1. Introdução
2. Fundamentação Teórica

Workflows

Proveniência

Ontologias

3. Trabalhos Relacionados
4. Método
5. Resultados

Tipo de Proveniência

Tipo de Ontologia

Benefícios da estrutura ontológica e granularidade

6. Conclusões e Trabalhos Futuros

# Introdução

---

- *Workflows* científicos são compostos por diversos módulos que manipulam dados e parâmetros de entrada, gerando novos dados que servirão de entrada para outros módulos em um fluxo.
- Entretanto na prática percebe-se que apenas compartilhar *workflows* não garante seu sucesso [1].

- Paralelo a isto tem-se modelos e métodos baseados em ontologias, que voltados a *workflows* apoiam a preservação e reprodução de experimentos.
  - *wfdesc*: utilizada para descrever a estrutura do *workflow*;
  - OBI (*Ontology for Biomedical Investigation*): provê termos relacionados a investigações no contexto da biomedicina [2].

- O trabalho objetivou analisar pesquisas de um repositório que abordam a aplicabilidade de ontologias para proveniência em experimentos científicos baseados em *workflow*.

# Fundamentação Teórica

---

O ciclo de vida de um *workflow* científico possui quatro estágios básicos segundo [6]:

- Projeto: Consiste na idealização de um *workflow*;
- Instanciação: Consiste ao processo de preparo para aplicar o *workflow* à uma tarefa específica. É importante salientar que neste estágio é importante indicar entradas e parâmetros para sua execução;
- Execução: Diz respeito a execução do *workflow* projetado e instanciado. As sub-tarefas que compõem o *workflow* são executadas, produzem dados que são reusados por outras sub-tarefas dentro do fluxo, com a finalidade de concluir o experimento;
- Análise Pós-Execução: Etapa que consiste na análise dos resultados obtidos com a execução do experimento;



- A proveniência diz respeito ao histórico dos registros de determinados contextos. No que tange *workflows* científicos, está relacionada a metadados que justifiquem e expliquem resultados, e etapas de derivação [9, 3].
  - Proveniência prospectiva: especificação de tarefas computacionais correspondentes a etapas do *workflow* gerar determinado resultado ou classe de resultados [3].
  - Proveniência retrospectiva: metadados que justifiquem a derivação de determinado objeto ou conjunto de dados [3].

- Ontologia pode ser entendida como um conjunto de conceitos e suas relações, que visam representar determinado domínio de maneira formal (eliminando inconsistências), possibilitando a compreensão do contexto por humanos e computadores [7].
  - **Ontologias de domínio** (*domain ontology*): definem determinado contexto bem como atividades e atores relacionados a ele, define um vocabulário voltado ao compartilhamento dentro de um domínio específico cujos principais objetivos são o compartilhamento e o reuso.
  - **Ontologias de tarefa** (*task ontology*): fornece um vocabulário aplicável a questões relacionadas a uma tarefa específica, que pode ou não ser independente de domínio.

# Trabalhos Relacionados

---

- O trabalho de Perez et al. [8] visa realizar um levantamento dos principais sistemas de captura de proveniência com dois objetivos principais: identificar suas principais características, e determinar quais são os mais utilizados e suas características comuns;
- Lopez et al. [5] por sua vez, realizam um mapeamento sistemático com a finalidade de analisar práticas em *Software Product Line* (SPL) tendo em vista as vantagens que o uso dessa prática impacta no produto;

# Método

---

- Foi definida a abordagem qualitativa, uma vez que não foram utilizados testes estatísticos no processo de análise;
- Mapeamento sistemático;



**Figura 1:** Método para a execução do estudo. Fonte: Kitchenham e Charters (2007)[4].

- **Definir o objetivo:** Identificar como pesquisadores utilizam ontologias no contexto da proveniência em experimentos científicos baseados em *workflows* tendo em vista o tipo de proveniência;
- **Definir questões de pesquisa:**
  - Q1: Qual o tipo de proveniência abordada na pesquisa analisada?
  - Q2: Qual o tipo de ontologia aplicada na pesquisa analisada?
  - Q3: Quais as principais vantagens em se utilizar ontologias no contexto de proveniência em experimentos baseados em *workflows* na pesquisa analisada?
  - Q4: Qual o nível de granularidade de proveniência abordada na pesquisa analisada?



- **Definir fonte de pesquisa:** Repositório Scopus<sup>1</sup>;
- **Definir *string* de busca:** ABS ( ontology AND workflow AND provenance ) AND ( LIMIT-TO ( PUBYEAR , 2017 ) OR LIMIT-TO ( PUBYEAR , 2016 ) );

---

<sup>1</sup><https://www.scopus.com>

- **Definir critérios de seleção:**
  - I1: A pesquisa deve ser escrita em inglês;
  - I2: A pesquisa deve ser fornecida de forma gratuita;
  - I3: Os termos da string de busca devem ser relacionados na problemática da pesquisa;

# Resultados

---

# Resultados: Tipo de Proveniência

- Informação implícita;
- 50% dos trabalhos analisados abordam a proveniência retrospectiva;
  - 3/4 publicados em 2016;
  - Heterogeneidade de dados - Frank et al.;
  - Sistema acoplável para sistemas de análise de dados - Wu et al.;
  - Ciência dos materiais - Merkys et al.;
- 25% abordam a proveniência prospectiva;
  - Sistemas de documentação de *workflow* - Abla et al.;
  - PovCaRe - Valdez et al.;
- 25% abordam ambas as proveniências;
  - Coordenar *web services* - Zhang et al.;
  - Domínio Geoespacial;
  - Relacionamento de dependências (hard. Soft. data) - Miksa et al.;

# Resultados: Tipo de Ontologia

- Informação implícita;
- Ontologias de tarefa > ontologia de domínio em 2016;
  - Stillerman, Alba, Wu / Frank
- Ontologias de domínio > ontologias de tarefa em 2017;
  - Merkys, Valdez, Zhang / Miksa

# Resultados: Benefícios da estrutura ontológica e granularidade

- Relacionamentos;
- Formalidade;
- Estrutura Hierárquica;
- Granularidade pouco abordada - Wu et al.

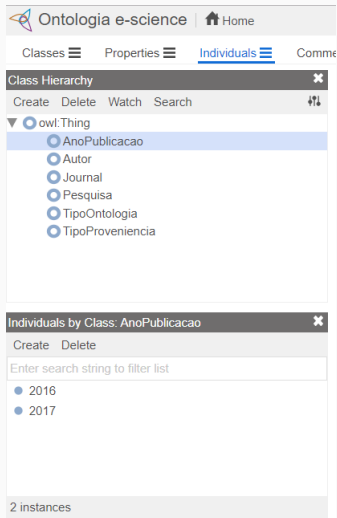
## Conclusões e Trabalhos Futuros

---

- Enquanto estudos menos recentes são focados na manipulação de proveniência retrospectiva, estudos mais recentes focam na proveniência prospectiva e em ambas.



# Conclusões e Trabalhos Futuros



**Figura 2:** Conceitos da ontologia



**Figura 3:** Repositório

-  K. Belhajjame, J. Zhao, D. Garijo, M. Gamble, K. Hettne, R. Palma, E. Mina, O. Corcho, J. M. Gómez-Pérez, S. Bechhofer, et al.  
**Using a suite of ontologies for preserving workflow-centric research objects.**  
*Web Semantics: Science, Services and Agents on the World Wide Web*, 32:16–42, 2015.
-  R. R. Brinkman, M. Courtot, D. Derom, J. M. Fostel, Y. He, P. Lord, J. Malone, H. Parkinson, B. Peters, P. Rocca-Serra, et al.  
**Modeling biomedical experimental processes with obi.**  
*Journal of biomedical semantics*, 1(1), 2010.
-  J. Freire, D. Koop, E. Santos, and C. T. Silva.  
**Provenance for computational tasks: A survey.**  
*Computing in Science & Engineering*, 10(3), 2008.
-  B. Kitchenham and S. Charters.  
**Guidelines for performing systematic literature reviews in software engineering, 2007.**