

## APAPER REVIEW: THE MPO SYSTEM FOR AUTOMATIC WORKFLOW DOCUMENTATION

JOURNAL	AUTORES	LINK - SCOPUS	COMPILADO POR
Fusion Engineering and Design	G. Abia, E.N.Coviello, S.M.Flanagan et al.	<a href="https://goo.gl/TupakW">https://goo.gl/TupakW</a>	Luiz Gustavo Dias - UFF

### RESUMO

Dados de experimentos de grande escala são produtos que podem ser utilizados para aplicações críticas, entretanto sua importância não está na quantidade e sim no que é possível com seu uso. Desta forma, o artigo apresenta um sistema de metadados, proveniência e ontologia (mpo) que pode automatizar a documentação de *workflows* científicos e suas informações. Baseado em metadados, fornece informações explícitas sobre as relações entre os passos do *workflow*.

### CONCEITOS MPO

O sistema MPO tem como objetivo documentar atividades de pesquisa científica rastreando *workflows*. Isso objetiva a facilitação na criação de metadados, auxiliando por exemplo na identificação das pessoas como em questões como Quem, O que, Quando, Como, e Porque,

Para criar um modelo para captura e gerenciamento de metadados em um sentido mais amplo, projetos MPO definem vários conceitos de conveniência:

1. Data Object: um pedaço de informação que está relacionada a uma atividade ou pesquisa, pode ser um grande conjunto de dados, um único valor, ou um artigo científico. Pode ser consumido ou produzido por um código de computador. Pode ser armazenado em um arquivo ou em uma base de dados.
2. Activity: qualquer coisa que cria, move ou modifica dados de uma forma para outra. Uma atividade pode consumir e produzir dados.
3. Connection: Link entre múltiplas atividades ou data objects.
4. Workflow: Uma série de atividades e data objects conectados, que são organizados na forma de um gráfico acíclico direcionado (DAG) que consiste em um fluxo de uma direção que permite loops. Um *workflow* mostra os passos de um experimento e a relação entre seus elementos. Data objects, activities e connections constroem blocos de um *workflow*.
5. Collection: Um grupo de entidades relacionadas. Pode conter um número de data objects e *workflows*. Uma coleção pode incluir outras coleções.
6. Metadata: um texto associado a um objeto, atividade workflow ou coleção.
7. Comment: informação textual que pode estar relacionado a um data object, activity, workflow, ou collection.

De forma adicional, MPO define outros dois conceitos: Proveniência e ontologia:

1. Proveniência: informações sobre a origem dos data objects, bem como todas as transformações sofridas por eles. Sempre que um workflow é executado, uma instância dessa execução é gerada. Essa instância representa a proveniência, que inclui data objects e parâmetros utilizados como entrada.
2. Ontologia: A ontologia e a estrutura que captura os termos que descrevem propriedades de objetos em um domínio específico. Pode referenciar a um vocabulário controlado ou estrutura de classificação. O principal objetivo das ontologias é permitir uma forma de pesquisa das informações básicas sobre as informações do domínio, bem como sua proveniência.

## ARQUITETURA MPO

Nesta seção os autores conceituam os componentes da arquitetura MPO e as relacionam. Desta forma não será analisada neste review, tendo vista o objetivo.

## EXPERIENCIA DE IMPLANTAÇÃO

Após testar MPO os autores ressaltam que muitas alterações foram realizadas desde sua versão 1.0, resultando em aprimoramentos notáveis. Os comandos possibilitados pela ferramenta são adaptados para *workflows* e produção automática de metadados durante a realização do experimento. Os metadados gerados e o grafo resultante da execução compõem um documento final bastante rico.

## CONCLUSÃO

O sistema MPO automatiza a documentação de fluxos de trabalho científicos e informações associadas e o faz de forma independente, além de possibilitar a análise de metadados documentados, bem como apresentá-los. Além do mais, pode ser usado além de forma individualmente como de forma colaborativa com outros pesquisadores.

## ANÁLISE SEGUNDO O LEITOR

1. Qual tipo de proveniência abordada no trabalho?  
**A pesquisa analisa uma ferramenta específica para proveniência de dados em workflows por meio do sistema MPO, e a partir do trecho “Based on recorded metadata, it provides explicit information about the relationships among the elements of workflows in notebook form augmented with directed acyclic graphs”, assume-se que o tipo de proveniência abordado é do tipo prospectiva.**
2. Qual tipo de ontologia utilizada no trabalho?  
**Não é citado. Entretanto, por possuir foco em utilizar a ontologia MPO para capturar proveniência de workflows, assume-se que a pesquisa aborda ontologias de tarefa.**
3. Qual a principal vantagem em se utilizar ontologia no contexto da pesquisa?  
**MPO não é utilizado apenas por pesquisadores individuais, mas também por grupos de pesquisa, e sua estrutura favorece o compartilhamento e colaboração entre seus componentes.**
4. Questões de granularidade são abordadas no artigo?  
**Não é citado.**