Aplicação de Ontologias de Proveniência em *Workflows* Científicos: um Mapeamento Sistemático*

Luiz Gustavo Dias[1], Bruno Lopes[1], Daniel de Oliveira[1]

[1]Instituto de Computação – Universidade Federal Fluminense (UFF)

lgdias@id.uff.br, {bruno,danielcmo}@ic.uff.br

Resumo. *Experimentos científicos modelados como workflows são executados por complexos mecanismos chamados de Sistemas de Gerência de Workflows (SGWf). Existem diversos SGWfs com seus prós e contras, porém todos compartilham diversas características como por exemplo, a necessidade de fornecer apoio para os cientistas analisarem seus dados. Os dados de proveniência tem um papel importante no fornecimento das informações necessárias em diferentes etapas experimentais. Desta forma, o presente trabalho tem como objetivo mapear e caracterizar abordagens que utilizam uma das quatro ontologias de proveniência selecionadas, analisando fatores como adequabilidade, requisitos de execução e arquitetura. Após o estudo, percebeu-se que as ontologias de proveniência podem ser aplicadas em diferentes etapas do ciclo de vida do workflow científico, mas principalmente na fase de análise.*

1. Introdução

A utilização de simulações computacionais complexas com o objetivo de apoiar experimentos em diversos domínios da ciência se tornou uma realidade na última década [Mattoso et al. 2010, Atkinson et al. 2017]. Modelar, executar e principalmente analisar o resultado de um experimento científico são tarefas de realização complexa, especialmente em experimentos computacionalmente intensivos compostos de diversas atividades (*workflows*), que manipulam um grande volume de dados, complexos e não estruturados, (*i.e., big data*) [Jagadish et al. 2014].

Existem diversas aplicações que já apoiam as tarefas diárias dos cientistas, principalmente no que tange a composição e a execução desses *workflows*. Os *workflows* são geralmente modelados, executados e monitorados por mecanismos complexos chamados Sistemas de Gerência de *Workflows* (SGWf), que apoiam a especificação em termos de artefatos executáveis. De acordo com [Liu et al. 2015], SGWfs bem conhecidos são o Pegasus, Swift/T, e o SciCumulus. Além dos SGWf, os *Scientific Gateways* [Karasavvas et al. 2012, Gesing et al. 2018, Gesing et al. 2017] são complexos sistemas de informação que visam integrar diversas abordagens existentes que apoiam a composição e execução de *workflows* em ambientes distribuídos como nuvens, grades computacionais e *clusters* de computadores. Além dos SGWfs e dos *Scientific Gateways* existem soluções mais pontuais como o SciManager [Ramos et al. 2016], que apoia as etapas do ciclo de vida do *workflow* tanto na fase de composição, quanto de execução, e análise [Mattoso et al. 2010].

Para que um experimento seja considerado de fato científico, o *workflow* associado a ele deve ser passível de reprodução sob as mesmas condições, mesmo que seja executado por terceiros. Desta forma, os descritores associados ao experimento, como por exemplo, sua definição, os dados consumidos e os produzidos durante a sua execução são fundamentais para que o mesmo seja considerado válido, consistente e ainda, capaz de ser reproduzido por terceiros.

*Os autores agradecem ao CNPq, CAPES e FAPERJ por financiarem esse trabalho.

Esta categoria de descritores denomina-se metadados de proveniência [Freire et al. 2008]. Todas as abordagens anteriormente citadas, além da execução de *workflows*, capturam metadados de proveniência.

Apesar de representarem um avanço, muitos desses sistemas não apoiam de forma efetiva a análise e consulta sobre os dados de proveniência capturados [de Oliveira et al. 2018]. Tais consultas e análises acabam sendo pontuais e dependem de conhecimento prévio do cientista sobre a estrutura do banco de dados de proveniência (*i.e., schema*). Idealmente, os cientistas deveriam consultar tais dados sem conhecimento prévio do *schema* do banco de dados e também inferir conhecimento não trivial sobre esses dados.

De forma a minimizar a complexidade na representação de dados de proveniência bem como em sua análise, estruturas ontológicas vêm se mostrando viáveis. O benefício do uso de ontologias na ciência e na indústria pode ser justificado por sua estrutura hierárquica, composta por classes e instâncias que se relacionam para representar o domínio [Bada et al. 2004]. O perfil de relacionamento possibilitado por ontologias, além de formalidade, proporciona bom desempenho quando aplicadas a conjuntos de dados escaláveis, como por exemplo bases de dados astronômicas e geológicas.

Embasado no exposto, o presente artigo tem como objetivo a construção de um mapeamento para analisar fatores arquiteturais e de aplicação de ontologias no contexto de proveniência de dados científicos. As pesquisas que compuseram o corpus foram obtidas através do indexador *Web Of Science* (`https://login.webofknowledge.com`) e analisadas de acordo com dez características: tipo de pesquisa, ferramenta aplicada, padrão de proveniência, domínio de aplicação, padrão de consulta, tecnologias utilizadas para implementação, nível de reprodutibilidade do estudo, tipo de arquivo de entrada, tipo de arquivo de saída, e tecnologias de armazenamento.

O presente artigo se encontra organizado em 4 seções além desta introdução. Na Seção 2 apresentamos uma breve fundamentação teórica. Na Seção 3 apresentamos a estrutura do mapeamento sistemático realizado e na Seção 4 os resultados obtidos. Finalmente, na Seção 5 apresentamos a conclusão desse artigo.

2. Fundamentação Teórica

Workflows científicos são abstrações utilizadas para mapear o encadeamento de aplicações que formam um fluxo coerente e são aplicados em diferentes domínios científicos [Mattoso et al. 2010]. No contexto da ciência podem ser definidos como um conjunto de atividades conectadas que descrevem um experimento científico, e ao contrário de *workflows* de negócio, são centrados no processo de transformação dos dados [Davidson and Freire 2008]. De forma resumida, as etapas do ciclo do vida de um *workflow* científico, podem ser definidas como: **[Composição]:** idealização de um *workflow* para a execução de um experimento. **[Instanciação]:** processo de preparo para aplicar o *workflow* a um ambiente e contexto. **[Execução]:** execução do *workflow* projetado e instanciado, as atividades que compõem o *workflow* são executadas, produzem dados que são reusados por outras atividades dentro do fluxo. **[Análise]:** análise dos resultados obtidos com a execução do experimento, pode ser realizada em tempo de execução, *i.e., runtime*, ou após a execução, *i.e., post-mortem.*

A proveniência por sua vez, diz respeito ao histórico dos registros de determinado contexto, e no que tange ferramentas científicas está associada a metadados que justifiquem e expliquem resultados e etapas de derivação [Sheikh et al. 2018, Freire et al. 2008]. Uma característica importante da proveniência é o seu nível de detalhamento, denominado granula-

ridade. A granularidade possibilita maior filtragem e manipulação de dados, além de empregar maior flexibilidade a sistemas, visto que é responsável pela precisão na especificação de dados/metadados de proveniência. Quanto a sua caracterização, a proveniência pode ser do tipo prospectivo ou retrospectivo [Freire et al. 2008]. A proveniência prospectiva se refere aos metadados da especificação do *workflow* ou da ferramenta, e é definida como a captura da especificação das atividades computacionais correspondentes a etapas que devem ser executadas. A proveniência retrospectiva por sua vez, diz respeito ao processo de captura de metadados relacionados ao ambiente e execução de um *workflow* junto aos módulos que compõem o sistema de *workflow* ou a aplicação, que por sua vez justifica determinado conjunto de dados [Freire et al. 2008].

Dados de proveniência podem ser representados em diferentes estruturas, dentre elas em forma de ontologias. Uma ontologia é definida como um conjunto de conceitos organizados por classes, semanticamente relacionados por componentes denominados propriedades. Tal característica viabiliza o reuso dos dados, visto o alto nível de detalhamento possibilitado por propriedades [Mizoguchi 2004]. Nessa pesquisa foram analisados trabalhos que aplicaram as ontologias *Open Provenance Model* (OPM — `https://openprovenance.org/opm`), PROV-O (`https://www.w3.org/TR/prov-overview/`), *Dublin Core* (DC — `http://dublincore.org/`) e *Provenance Authoring and Versioning* (PAV — `https://pav-ontology.github.io/pav/`). Tais ontologias foram selecionadas para o estudo visto sua ampla utilização no que tange dados de proveniência, metadados, e organização do conhecimento.

3. Mapeamento Sistemático

A execução do estudo foi baseada no protocolo descrito na subseção a seguir composto por: Objetivo, Questão de Pesquisa, Critérios de Análise, Fonte de Pesquisa, *String* de Busca, e Critérios de Seleção [Kitchenham 2004].

3.1. Planejamento e Execução

A estrutura utilizada para a realização do presente mapeamento foi composta por três módulos, sendo eles o planejamento, a execução e a análise de resultados. A respeito da etapa de planejamento, foi utilizado o protocolo a seguir.

Objetivo: Identificar como ontologias são utilizadas no contexto de proveniência e *workflows* científicos.

Definição da questão de pesquisa: Tal etapa constou em definir a questão de pesquisa a ser respondida, desta forma tem-se que a questão geral foi definida como: **"QP: Quais os cenários de aplicação das ontologias analisadas, no âmbito de proveniência e *workflows* científicos?"**.

Definição de critérios de análise: Para que a questão de pesquisa fosse respondida, os estudos foram analisados de acordo com dez critérios: tipo de pesquisa, ferramenta aplicada, padrão de proveniência, domínio, linguagem de consulta, requisitos de implementação, reprodutibilidade do estudo, tipos de arquivo de entrada, tipos de arquivo de saída e armazenamento de proveniência.

Definição da fonte de pesquisa: Etapa designada à definição da base para coleta de dados. Como resultado da etapa, ficou definida a base *Web of Science*.

Definição da *string* de busca: Etapa destinada à listagem dos termos utilizados para a busca estudos relacionados ao tema. Desta forma foram selecionados trabalhos que continham em seu título ou resumo os termos "*provenance authoring and versioning*", "*dublin*

core metadata initiative", "PROV-O", e "OPM". Sendo assim foi utilizada a seguinte *string* de busca: TÍTULO:(provenance authoring and versioning) OR TÍTULO: (prov-o) OR TÍTULO: (open provenance model) OR TÍTULO: (dublin core metadata initiative) AND TIPOS DE DOCUMENTO: (Article OR Abstract of Published Item OR Book Chapter).

Definição de critérios de seleção: Etapa voltada à realização da triagem secundária, com objetivo de definir critérios de inclusão e exclusão de pesquisas da amostra. Foram definidos como critérios de inclusão:

- CI1: A pesquisa deve ser escrita em inglês;
- CI2: Os resultados da *string* de busca devem ser relacionados a problemática da pesquisa;

Foi definido como critério de exclusão:

- CE1: A pesquisa não cumpre pelo menos um critério de inclusão.

A busca dos trabalhos no repositório foi realizada em agosto/setembro de 2018 e após submeter a *string* de busca indexador, foram retornadas 21 pesquisas. Após filtragem inicial com base no protocolo, quatro artigos foram eliminados da amostra, e após leitura dos mesmos, o número foi reduzido a 15 artigos significativos. O processo de análise dos estudos constou em desenvolver uma estrutura tabular e a partir da leitura identificar determinados itens que a preencheriam. Após a análise, os estudos foram agrupados tendo como padrão a linha do tempo da publicação, desenvolvida de forma crescente. A relação completa dos trabalhos analisados está disponível em: `https://goo.gl/URZRHp`.

4. Resultados

Cenário de Aplicação de Ontologias para Proveniência (QP): A partir dos resultados encontrados, pode-se afirmar que ontologias para proveniência podem ser utilizadas em diferentes etapas do ciclo de vida de um *workflow*, principalmente nas etapas de composição [Hoekstra and Groth 2014, de Oliveira et al. 2012] e análise [Simmhan et al. 2011, Simmhan and Barga 2011, Schreiber et al. 2012]. Os benefícios relacionados à fase de composição, podem ser comprovados, tendo como base a variedade de agentes, atividades e instâncias envolvidos no processo de desenvolvimento de um experimento científico, onde diferentes componentes podem ser equivalentes, possibilitando assim o desenvolvimento de *wokflows* variantes. Aplicadas na etapa de análise, ontologias permitem que mecanismos de inferência sejam agregados no processo de verificação dos resultados. Além de fornecer informações sobre como os dados se relacionam, o que/quem os gerou, suas versões, dentre outras características, ontologias podem simplificar o processo de entendimento das transformações de dados que ocorrem durante a execução do experimento. Isso é possível porque além dos dados resultantes, ontologias também possibilitam relacionar os resultados a metadados e agentes.

Domínio de Aplicação e Estrutura Aplicada: Após a avaliação notou-se que quando levada em consideração a linha do tempo de publicação das pesquisas, existe tendência em tornar as soluções específicas de domínio. Pesquisas mais recentes [Feng 2013, Ciccarese et al. 2013, Jing 2015, Wu and Treloa 2015] em sua maioria estão relacionadas a soluções específicas de domínio, enquanto pesquisas mais antigas objetivam documentar estruturas [Moreau et al. 2011, Groth and Moreau 2011], propor mapeamentos entre estruturas [Miles 2011], e propor/avaliar ferramentas multidomínio [Schreiber et al. 2012, Hoekstra and Groth 2014, Simmhan and Barga 2011]. Existem pesquisas publicadas em

períodos iniciais que também propõem soluções de domínio [Simmhan et al. 2011, Pan et al. 2011], entretanto em menor quantidade quando comparadas a períodos mais recentes. O motivo para tal especificação pode ser justificado pela constante necessidade de sanar problemas pontuais de domínio, como a normalização de dados heterogêneos advindos de fontes distintas do domínio geoespacial [Frank and Zander 2016], e acurácia nos resultados experimentais tratando-se principalmente de bases de dados escaláveis como no caso de bases de dados astronômicas [Feng 2013].

Relacionado a ontologia utilizada, levando em consideração a ordem cronológica das publicações, nota-se que apesar do modelo PROV ser mais robusto e mais recente que o modelo OPM, e de existir maior volume de soluções que fazem uso dessa ontologia publicadas em períodos mais recentes [Jing 2015, Hoekstra and Groth 2014, Wu and Treloa 2015], o modelo OPM também é utilizado [Pan et al. 2011]. Quanto à ontologia PAV, percebeu-se que além de ser aplicável de forma independente de domínio, a mesma pode ser utilizada de forma a complementar a ontologia PROV. Apesar de agregar muitas vantagens ao aplicar elementos PAV, uma desvantagem encontrada foi a necessidade de um especialista de domínio, uma vez que é necessário procedimento de triagem e análise do dado a ser inserido na estrutura. No caso da DC, a *string* de busca aplicada não listou pesquisas que abordassem sua aplicabilidade a proveniência fonte de pesquisa utilizada.

Linguagens para Consultas, Implementação, Armazenamento e Reprodutibilidade: Sobre a metodologia utilizada durante o processo criativo das soluções, tem-se que a grande maioria dos estudos analisados não explicitam linguagens relacionadas à implementação, técnicas de inferência ou armazenamento de proveniência [Kwasnikowska and Van den Bussche 2008, Miles 2011, Moreau et al. 2011, Pan et al. 2011, Groth and Moreau 2011, Ciccarese et al. 2013, Hoekstra and Groth 2014, Wu and Treloa 2015, Moreau et al. 2008, Kwasnikowska et al. 2015], o que pode ser justificado pelo perfil teórico da grande maioria das pesquisas avaliadas.

Nos casos em que haviam essas informações, foram utilizadas no processo de construção de consultas para inferência, dois tipos distintos de linguagens, a linguagem relacional SQL [Simmhan and Barga 2011, Simmhan et al. 2011, Feng 2013, Jing 2015], e a linguagem baseada em grafos *Gremlin* [Schreiber et al. 2012]. A partir disso notou-se que o paradigma de armazenamento da informação não impacta diretamente no uso das estruturas no contexto da proveniência, uma vez que os recursos utilizados nestas pesquisas variam entre os tipos de armazenamento e linguagem de consulta.

No tipo de pesquisa e informações sobre linguagens de programação, consulta e armazenamento, tem-se que quase 50% não forneciam este tipo de informação [Pan et al. 2011, Groth and Moreau 2011, Ciccarese et al. 2013, Hoekstra and Groth 2014, Wu and Treloa 2015]. As linguagens utilizadas para implementação foram Java, C e .NET [Simmhan and Barga 2011, Simmhan et al. 2011, Feng 2013, Jing 2015].

Input e Output: No que tange informações de entrada e saída das ferramentas identificadas no estudo, tem-se a utilização dos formatos XML [Simmhan and Barga 2011, Groth and Moreau 2011, Feng 2013, Ciccarese et al. 2013, Jing 2015, Hoekstra and Groth 2014], CSV [Simmhan et al. 2011] e RDF [Feng 2013, Ciccarese et al. 2013, Hoekstra and Groth 2014]. De forma análoga aos outros itens avaliados

no mapeamento, nem todas as pesquisas possuíam informações dessa natureza, tendo em vista o perfil da pesquisa. Levando em consideração trabalhos que explicitavam tais características, tem-se que o formato XML teve predominância tanto em arquivos de entrada quanto saída, o que pode ser justificado pela portabilidade fornecida pelo formato, estrutura e nível de detalhamento da informação disposta em arquivos que fazem uso de tal extensão.

5. Conclusão

Este artigo descreve um mapeamento sistemático de ontologias para proveniência. Sua contribuição científica baseada no planejamento do protocolo é relacionada aos resultados do mapeamento e na resposta da questão de pesquisa. Após o levantamento possibilitado pela aplicação de uma string de busca composta por quatro estruturas ontológicas à base de dados *Web of Science*, a qual retornou pesquisas relacionadas a apenas três ontologias (OPM, PROV e PAV). Foram classificados, analisados e comparados 15 artigos representativos.

Percebeu-se que a quantidade de pesquisas teóricas sobressaem pesquisas experimentais, e que existem diversas propostas de mapeamento entre estruturas do tipo OPM e PROV. Conclui-se também que é possível tornar a ontologia PROV mais robusta, fazendo uso da ontologia PAV. Quando levados em consideração o ciclo de vida de *workflows* e as etapas de composição e análise, tem-se que a maioria das soluções desenvolvidas aplicam ontologias na fase de análise, e que o processo de definição da ontologia a ser utilizada, deve ser avaliada levando em consideração por exemplo, a natureza do problema a ser resolvido.

Nota-se ainda que informações relacionadas às características analisadas não são descritas nas pesquisas analisadas, o que impacta de forma direta ou indireta na reprodutibilidade dos estudos. Percebeu-se que o uso de ontologias pode agregar diversas vantagens a experimentação científica principalmente em cenários heterogêneos. Isso porque através do relacionamento de elementos proporcionados por ontologias, faz-se possível por exemplo, a manipulação de dados de entrada e saída de diferentes tipos e por diferentes módulos, independente da etapa do ciclo de vida do *workflow*, possibilitando a orquestração de serviços compatíveis aos dados. Outra vantagem está relacionada ao nível de granularidade de metadados de proveniência. Tal benefício pode ser justificado pelos padrões de propriedades e elementos propostos por diferentes tipos de ontologia, que podem ser mais simplificados como o padrão OPM, e mais robustos como a ontologia PROV-O, que pode relacionar dados ainda mais específicos quando agregada a ontologia PAV, que adiciona metadados de proveniência autoria e versionamento ao contexto.

Referências

Atkinson, M. P., Gesing, S., Montagnat, J., and Taylor, I. J. (2017). Scientific workflows: Past, present and future. *Future Generation Comp. Syst.*, 75:216–227.

Bada, M., Stevens, R., Goble, C., Gil, Y., Ashburner, M., Blake, J. A., Cherry, J. M., Harris, M., and Lewis, S. (2004). A short study on the success of the gene ontology. *Web Semantics: Science, Services and Agents on the World Wide Web*, 1(2):235–240.

Ciccarese, P., Soiland-Reyes, S., Belhajjame, K., Gray, A. J., Goble, C., and Clark, T. (2013). Pav ontology: provenance, authoring and versioning. *Journal of biomedical semantics*, 4(1):37.

Davidson, S. B. and Freire, J. (2008). Provenance and scientific workflows: challenges and opportunities. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1345–1350. ACM.

de Oliveira, D., Ogasawara, E. S., Dias, J., Baião, F. A., and Mattoso, M. (2012). Ontology-based semi-automatic workflow composition. *Journal of Information and Data Management*, 3(1):61–72.

de Oliveira, W. M., de Oliveira, D., and Braganholo, V. (2018). Provenance analytics for workflow-based computational experiments: A survey. *ACM Comput. Surv.*, 51(3):53:1–53:25.

Feng, C.-C. (2013). Mapping geospatial metadata to open provenance model. *IEEE transactions on geoscience and remote sensing*, 51(11):5073–5081.

Frank, M. and Zander, S. (2016). Smart web services for big spatio-temporal data in geographical information systems. In *SALAD@ ESWC*.

Freire, J., Koop, D., Santos, E., and Silva, C. T. (2008). Provenance for computational tasks: A survey. *Computing in Science & Engineering*, 10(3).

Gesing, S., Dooley, R., Pierce, M. E., Krüger, J., Grunzke, R., Herres-Pawlis, S., and Hoffmann, A. (2018). Gathering requirements for advancing simulations in HPC infrastructures via science gateways. *Future Generation Comp. Syst.*, 82:544–554.

Gesing, S., Wilkins-Diehr, N., Dahan, M., Lawrence, K. A., Zentner, M. G., Pierce, M. E., Hayden, L., and Marru, S. (2017). Science gateways: The long road to the birth of an institute. In *50th Hawaii International Conference on System Sciences, HICSS 2017, Hilton Waikoloa Village, Hawaii, USA, January 4-7, 2017*.

Groth, P. and Moreau, L. (2011). Representing distributed systems using the open provenance model. *Future Generation Computer Systems*, 27(6):757–765.

Hoekstra, R. and Groth, P. (2014). Prov-o-viz-understanding the role of activities in provenance. In *International Provenance and Annotation Workshop*, pages 215–220. Springer.

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., and Shahabi, C. (2014). Big data and its technical challenges. *Commun. ACM*, 57(7):86–94.

Jing, N. (2015). A prov-o based approach to web content provenance. In *Logistics, Informatics and Service Sciences (LISS), 2015 International Conference on*, pages 1–6. IEEE.

Karasavvas, K., Wolstencroft, K., Mina, E., Cruickshank, D., Williams, A. R., Roure, D. D., Goble, C. A., and Roos, M. (2012). Opening new gateways to workflows for life scientists. In *HealthGrid Applications and Technologies Meet Science Gateways for Life Sciences, Proceedings of HealthGrid 2012, Amsterdam, The Netherlands, 21-23 May 2012.*, pages 131–141.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26.

Kwasnikowska, N., Moreau, L., and Bussche, J. V. D. (2015). A formal account of the open provenance model. *ACM Transactions on the Web (TWEB)*, 9(2):10.

Kwasnikowska, N. and Van den Bussche, J. (2008). Mapping the nrc dataflow model to the open provenance model. In *International Provenance and Annotation Workshop*, pages 3–16. Springer.

Liu, J., Pacitti, E., Valduriez, P., and Mattoso, M. (2015). A survey of data-intensive scientific workflow management. *Journal of Grid Computing*, 13(4):457–493.

Mattoso, M., Werner, C., Travassos, G. H., Braganholo, V., Ogasawara, E. S., de Oliveira, D., da Cruz, S. M. S., Martinho, W., and Murta, L. (2010). Towards supporting the life cycle of large scale scientific experiments. *IJBPIM*, 5(1):79–92.

Miles, S. (2011). Mapping attribution metadata to the open provenance model. *Future Generation Computer Systems*, 27(6):806–811.

Mizoguchi, R. (2004). Tutorial on ontological engineering part 2: Ontology development, tools and languages. *New Generation Computing*, 22(1):61–96.

Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., et al. (2011). The open provenance model core specification (v1. 1). *Future generation computer systems*, 27(6):743–756.

Moreau, L., Freire, J., Futrelle, J., McGrath, R. E., Myers, J., and Paulson, P. (2008). The open provenance model: An overview. In *International Provenance and Annotation Workshop*, pages 323–326. Springer.

Pan, J., Lenhardt, C., Wilson, B., Palanisamy, G., Cook, R., and Shrestha, B. (2011). Geoscience data curation using a digital object model and open-source frameworks: Provenance applications. In *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pages 3815–3818. IEEE.

Ramos, L., Ocaña, K., and Oliveira, D. (2016). Um sistema de informação para gerência de projetos científicos baseados em simulações computacionais. In *Proceedings of the XII Brazilian Symposium on Information Systems*, pages 216–223. ACM.

Schreiber, A., Ney, M., and Wendel, H. (2012). The provenance store proost for the open provenance model. In *International Provenance and Annotation Workshop*, pages 240–242. Springer.

Sheikh, U., Khan, A., Ahmed, B., Waheed, A., and Hameed, A. (2018). Provenance inference techniques: Taxonomy, comparative analysis and design challenges. *Journal of Network and Computer Applications*.

Simmhan, Y. and Barga, R. (2011). Analysis of approaches for supporting the open provenance model: A case study of the trident workflow workbench. *Future Generation Computer Systems*, 27(6):790–796.

Simmhan, Y., Groth, P., and Moreau, L. (2011). Special section: The third provenance challenge on using the open provenance model for interoperability. *Future Generation Computer Systems*, 27(6):737–742.

Wu, M. and Treloa, A. (2015). Metadata in research data australia and the open provenance model: A proposed mapping. In *21st International Congress on Modelling and Simulation, Gold Coast, Australia*.