

Uma Arquitetura para o Registro de Dados de Proveniência em Operações de Caracterização Geológica

Gefersom Lima¹, José Viterbo¹, Clodis Boscarioli² e Edmundo Marassi³

¹Universidade Federal Fluminense (UFF) – Rua Passo da Pátria 156 - Bloco E - 3º andar
CEP: 24210-240 – Niterói, RJ, Brasil

²Universidade Estadual do Oeste do Paraná (UNIOESTE) – Bloco F – Sala 15
CEP: 85819-110 – Cascavel, PR, Brasil

³Petrobras – Petróleo Brasileiro S.A. – Av. República do Chile, 65
CEP 20031-912 – Rio de Janeiro, RJ, Brasil

gefersom@gmail.com, viterbo@ic.uff.br, clodis.boscarioli@unioeste.br,
emarassi@gmail.com

Abstract. *In the prospection of oil reservoirs, geological characterization is a fundamental phase for building models that represent the elements of the petroleum system. The record of provenance data describing the intermediate steps of this process and the data set used in the operations executed, allows the full evaluation of the quality of the resulting characterization. This paper proposes an architecture that supports the collection and recovery of provenance data in applications for geological characterization, and presents a prototype tool implemented as a plug-in for Sigeo3 a geoscience software framework used by Petrobras, the main Brazilian oil company.*

Resumo. *Na prospecção de reservatórios de poços de petróleo, a caracterização geológica é uma fase importante para a construção de modelos que representam os elementos do sistema petrolífero. O registro das informações de proveniência relativas aos passos intermediários deste processo e o conjunto de dados utilizados nas operações realizadas, permite que a qualidade da caracterização resultante seja completamente avaliada e que o processo seja reproduzido sob outras circunstâncias. Este trabalho propõe uma arquitetura que suporta a coleta e recuperação de dados de proveniência em aplicações de caracterização geológica, e apresenta uma ferramenta protótipo implementada como extensão do Sigeo3, um arcabouço de desenvolvimento de software geocientífico utilizada pela Petrobras.*

1. Introdução

A descoberta de uma jazida de petróleo em uma nova área é uma tarefa que envolve longo e dispendioso estudo e análise de dados geofísicos e geológicos das bacias sedimentares. O processo que visa a fornecer informações suficientes para a tomada de decisão de perfurar um poço a partir do conjunto de dados disponíveis é denominado caracterização geológica.

De acordo com Houlding (1995), um processo genérico de caracterização geológica deve conter o seguinte conjunto de etapas básicas:

- **Gerenciamento, Correlação e Integração** de diversas fontes de informação disponíveis de várias regiões investigativas. A mais importante, e potencialmente de maior consumo de tempo, é a correlação espacial das fontes de informações investigativas.
- **Revisão e Análise** das fontes de informações em termos de sua qualidade, suficiência, escala e variação espacial. O objetivo é prover uma apreciação inicial da complexidade e magnitude das condições com as quais o geólogo terá que lidar.
- **Interpretação** da estratigrafia geológica, estrutura e outros fatores relevantes, baseados largamente em observações de características. O resultado é uma discretização da subsuperfície contínua pelo controle das características geológicas. O termo “interpretação” é utilizado devido ao fato de que há informações limitadas que exigem a aplicação da intuição e experiência geológica do geólogo.
- **Análise Espacial** da interpretação e predição da informação, geralmente em termos volumétricos e visualização de procedimentos, e/ou transferência de características resultantes para processos finais relevantes.
- **Predição** da variação espacial de variáveis relevantes, com base em amostras de informações, interpretação geológica e “técnicas de predição” adequadas. O termo “predição” é utilizado já que as verdadeiras condições são raramente determinadas a partir das informações limitadas e as predições são sempre subjetivas a algum grau de incerteza.

Ferramentas computacionais são utilizadas nesse processo, que compreende a execução de vários passos nos quais o geólogo toma decisões que afetam o resultado final. Para (CHENEY et al., 2009), o registro das informações de proveniência relativas aos passos intermediários e o conjunto de dados utilizados nessas operações, permitem que a qualidade da caracterização resultante seja completamente avaliada, o que não poderia ser feito considerando-se apenas o resultado final.

Durante um processo de caracterização geológica há a execução de uma grande quantidade de passos que envolvem uma grande quantidade de informações. Neste cenário, alguns passos podem não ter o resultado esperado pelo geólogo. Neste momento, fica clara a necessidade de análise dos passos anteriores para se verificar se em algum momento houve alguma falha. Ao encontrar o ponto de falha, esta pode ser corrigida, dando-se continuidade ao trabalho de caracterização.

Este aspecto é a principal motivação deste trabalho, ter o registro dos passos executados ao longo do processo de caracterização, com um determinado grau de detalhes que permita identificar o que foi feito anteriormente, torna possível identificar eventuais pontos de falha durante o processo, de forma que um processo possa ser reproduzido, alterando-se apenas os passos desejados.

Algumas das ferramentas de apoio à caracterização geológica como o Petrel (BRYANT e XU 2011) apresentam um histórico de operações sobre os dados que

contém como base as seguintes informações: nome da operação, tipo da operação (podendo ser criação, atualização ou edição), usuário e data e hora. Neste histórico, não há menção de quais dados foram utilizados para a criação de outro dado, quais foram os parâmetros de configuração, algoritmos de interpolação ou cálculo de algum tipo de média e quais informações foram contidas em um dado que sofreram atualizações. Para que essas informações sejam inseridas no histórico do dado, é necessário prover um meio de captura, armazenamento e recuperação das mesmas.

O objetivo deste trabalho é apresentar uma arquitetura de ferramentas para o auxílio do registro, recuperação e armazenamento de informações de proveniência relativas aos artefatos geológicos produzidos durante o processo de caracterização geológica, ou, mais especificamente, para a predição de valores utilizados na criação de mapas de contorno de áreas de interesse.

A arquitetura idealizada incluiu: (i) a proposta de um modelo baseado na ontologia OPM para a representação de dados de proveniência; (ii) a definição de uma metodologia para o registro de dados de proveniência baseada em arquivos externos no formato XML; e (iii), a definição de um conjunto de pacotes para permitir a implementação de ferramentas de coleta e recuperação de dados de proveniência.

Este documento está organizado como segue: Na próxima seção estão apresentados os conceitos básicos relacionados à arquitetura proposta. Na Seção 3, discutimos a arquitetura proposta. A Seção 4 descreve a implementação do protótipo. E, finalmente, na Seção 5, apresentamos nossas conclusões finais.

2. Proveniência de Dados

Proveniência pode ser entendida como o conjunto de processos que originaram um determinado dado (SIMMHAN, PLALE e GANNON, 2005). Aqui, o termo processo designa todas as derivações, conjuntos de dados intermediários e parâmetros de configuração envolvidos da produção dos dados.

Dados de proveniência têm sido utilizados em sistemas computacionais atuais para o acompanhamento de histórico, registro, integridade, autenticidade e recuperação de erros (CHENEY et al., 2009). Muitos sistemas computacionais e usuários utilizam essas informações para auxiliar na correção de erros, depuração, auditoria ou apenas como auxiliares a alguma tarefa como: registros de eventos importantes do sistema, administração do sistema, detecção de invasões, controle de versões de arquivos, históricos de sites navegados mantidos pelos navegadores web.

Uma solução para gerenciamento de dados de proveniência deve compreender três componentes principais (FREIRE et al., 2008):

- Máquina de captura, que realiza a coleta de informações de proveniência;
- Modelo de proveniência, que define a representação das informações de proveniência da abordagem ou a sua modelagem; e,
- Infraestrutura para armazenamento e consulta, que define questões relacionadas à base de dados utilizada como modos de armazenamento, formas de acesso, mecanismo para consultas aos dados de proveniência, etc.

Para este trabalho, o mecanismo utilizado foi o de processo (baseia-se no fato de que todo serviço ou processo deve documentar a si mesmo em relação à proveniência de dados), já que, para um sistema construído baseado em pontos de extensões (*plugins*), pode-se ter a liberdade de gerenciar os dados de proveniência de acordo com os requisitos de cada ponto de extensão e também de acordo com as necessidades do usuário.

Neste tipo de captura cada processo ou atividade deve dispor de recursos que permitam a coleta dos dados de proveniência. A vantagem é que cada atividade ou processo pode controlar a granularidade e a seleção dos dados de proveniência a serem coletados (MENDES, 2011). Além disso, também é possível coletar informações sobre os dados produzidos e intervenções feitas pelo usuário em tempo de execução.

Vários modelos de proveniência já foram propostos e todos compartilham dois tipos essenciais de informações: processo e dependências de dados. Os modelos de proveniência tendem a variar de acordo com o domínio e as necessidades dos usuários nas situações em que são empregados.

Dentre estes, neste trabalho foi adotado o modelo de proveniência denominado *Open Provenance Model* (OPM), proposto por Moreau et al. (2011), que tem por objetivo capturar as dependências causais entre os artefatos, processos e agentes. O OPM é estruturado em torno de um grafo direcionado, no qual os nós correspondem a algum tipo de entidade e as dependências são demonstradas pelas arestas do grafo. A escolha do OPM em comparação aos demais modelos, deve-se ao fato de que suas características se alinham aos princípios deste trabalho.

O modelo OPM tem por base três tipos de entidades (ou nós do grafo):

- Artefato (*Artifact*): Estado imutável, o qual pode ter representação física em um corpo físico ou uma representação digital em um sistema computacional.
- Processo (*Process*): Ação ou série de ações realizadas sobre um artefato ou causadas por ele na qual se resulta na criação de um novo artefato.
- Agente (*Agent*): Entidade contextual que age como um catalisador sobre um processo, habilitando, facilitando, controlando e afetando sua execução.

Muitas abordagens para armazenamento e consulta de dados de proveniência utilizam diferentes modelos de dados, podendo ser baseados em Web Semântica, a exemplo de RDF e OWL, ou em dialetos em XML (FREIRE et al., 2008). Os dados de proveniência podem ser armazenados em tabelas de bancos de dados relacionais ou em arquivos. A vantagem do sistema de arquivos é que não é necessário que haja uma infraestrutura adicional para o armazenamento das informações de proveniência. Porém, uma base de dados relacional pode oferecer uma abordagem de armazenamento de dados de proveniência centralizada e eficiente a qual um grupo de usuários pode compartilhar.

A definição de estruturas para o armazenamento e recuperação de informações de proveniência muitas vezes é afetada por requisitos não-funcionais como restrições de acesso, ambiente de execução, desempenho e facilidade de uso. Algumas abordagens fazem uso de arquivos persistidos no próprio sistema de arquivos local ou em repositórios em bases de dados. O que geralmente determina se um objeto estará ou não

no banco de dados é o seu tamanho em megabytes, por exemplo, volume sísmico, cujo valor em gigabytes pode chegar às centenas.

Como exemplo de sistemas que empregam o mecanismo de persistência baseados em arquivos podem ser citados o PASOA (MOREAU et al., 2011) e o Kepler (LUDÄSCHER et al., 2005). O projeto PASOA (*The Provenance-Aware Service Oriented Architecture*) desenvolveu uma arquitetura de proveniência que se baseia em serviços individuais para gravar sua própria proveniência de dados, focada em experimentos executados em grades computacionais, fazendo uso de serviços web para as chamadas das aplicações presentes na grade. O sistema não modela a noção de *workflow*, em vez disso, captura assertivas que refletem relações entre os serviços e os dados, inferindo a proveniência de uma tarefa ou um produto de dados pela combinação destas assertivas e recursivamente seguindo os relacionamentos que estas representam. Kepler é um Sistema Gerenciador de *Workflow* Científico (GASPAR et al 2010) que armazena os dados de proveniência prospectiva numa Linguagem de Modelagem de Marcação (*Modeling Markup Language - MoML*) formato 1, um dialeto XML para representação de workflows.

Em relação a sistemas que empregam a abordagem de uso de bancos de dados tem-se como exemplo os sistemas Taverna (OINN et al. 2007) e Redux (BARGA e DIGIAMPIETRI, 2008). Taverna é um SGWfC utilizado no projeto myGrid, cujo objetivo é alavancar tecnologias da Web Semântica e ontologias disponíveis para bioinformática para simplificar o processo de análise dos dados. Neste sistema o próprio workflow é responsável pela captura dos dados de proveniência. As informações de proveniência prospectiva são armazenadas no formato XML e as de retrospectiva são armazenadas em uma base de dados relacional. Redux é um sistema que estende o *Windows Workflow Foundation Engine* com o intuito de capturar dados de proveniência do workflow de forma transparente. Todos os dados de proveniência (tanto a do tipo prospectiva como a retrospectiva) no Redux são armazenados em uma base de dados relacional que pode ser consultada através de SQL. Neste sistema também se tem a possibilidade de se realizar uma consulta que fornece todos os passos executados para a derivação de um determinado produto de dados, o qual é denominado de *provenance query* (consulta de proveniência).

É importante definir os mecanismos para acessar as informações de proveniência usando abordagens que façam uso de características de interfaces gráficas, nas quais os dados são visualizados a partir de diferentes formas, dependendo apenas das necessidades dos usuários do sistema. Para a persistência dos dados de proveniência foi escolhido um modelo baseado em documentos semiestruturados utilizando a linguagem de marcação XML. A escolha de um modelo de persistência de dados em XML se deve à aderência deste tipo de abordagem aos requisitos funcionais e não-funcionais, pelo uso de pouco espaço de armazenamento, pelo acesso às informações contidas no arquivo estruturado em formato XML ser rápido, portabilidade, entre outros fatores.

2. Modelo de Proveniência

Segundo Thomas (2001), a caracterização geológica preocupa-se com a predição da variação espacial de uma ou mais variáveis geológicas. O objetivo final é identificar as

condições geológicas e geofísicas que apresentem a possibilidade de se encontrar jazidas de hidrocarbonetos de petróleo.

Durante um processo de caracterização geológica há a execução de uma grande quantidade de passos que envolvem muitas informações. Neste cenário, alguns passos podem não ter o resultado esperado pelo geólogo, evidenciando a necessidade de análise dos passos anteriores para verificar se em algum momento houve alguma falha. Ao encontrar o ponto de falha, esta pode ser corrigida, dando continuidade ao trabalho de caracterização. Este aspecto é a principal motivação deste trabalho, o registro dos passos executados ao longo do processo de caracterização, com um grau de detalhes que permita identificar o que já foi feito, torna possível identificar eventuais pontos de falha durante o processo. Desta forma, um processo pode ser reproduzido, alterando-se apenas os passos desejados.

De acordo com as características do processo de caracterização geológica discutidas neste capítulo, foram selecionadas as seguintes informações para o registro de proveniência no escopo deste trabalho:

- Nome do método de gridagem escolhido, que tem efeito direto sobre a qualidade do mapa gerado, o termo “qualidade” aqui significando que quanto mais o mapa produzido se aproximar dos dados reais mais “qualidade” o mesmo possui, ou seja, o mapa gerado possuirá menos distorções das em relação aos dados reais.
- Sistema de coordenadas associado aos valores observados em uma área a ser gridada, que define a localização de um mapa no planeta Terra. Um mapa só pode ser corretamente avaliado se a região que o mapa representa for conhecida.
- Identificação do usuário que realizou a criação do mapa. Um usuário mais experiente em relação à construção de mapas poderá escolher tanto um algoritmo como parâmetros que sejam mais adequados à região para a qual o mapa será gerado. Esta escolha é subjetiva, ou seja, somente a partir da observação do mapa gerado com os dados originais é que se pode saber se a escolha foi correta ou não.
- Identificação do artefato original utilizado, informação de suma importância, pois sem o conhecimento de qual fonte de dados o mapa foi gerado não é possível avaliar sua qualidade.
- Nome do processo utilizado para a criação do mapa. Um mapa pode ser criado pela combinação de outros mapas ou a partir de um conjunto de pontos de uma determinada área, onde a cada coordenada 2D é associado um valor que representa alguma variável geológica, como a porosidade de uma rocha reservatório.

A combinação de mapas não foi retratada neste trabalho, mas para uma breve descrição os mapas produzidos são chamados de mapas de contorno e a sua criação se dá através da aplicação de operações aritméticas sobre os nós de dois grids de mesma dimensão. Estas operações podem ser de subtração, multiplicação e adição e produzem outro mapa. Para detalhes sobre a criação de mapas de contorno o trabalho de (JONES, HAMILTON e JOHNSON, 1987).

A criação de mapas neste trabalho supõe que somente um conjunto de pontos é utilizado para a criação de mapas. Uma observação importante, é que as informações relativas à resolução do grid, quantidade de linhas e colunas do mesmo não foram escolhidas para o conjunto de dados de proveniência a serem capturados, pois comumente possuem valores fixos para qualquer mapa gerado.

3. A Arquitetura Proposta

A Figura 1 apresenta um diagrama que representa as relações entre os pacotes que compõe a arquitetura proposta: ‘Core’, ‘Persistencia’, ‘Captura’ e ‘Apresentacao’.

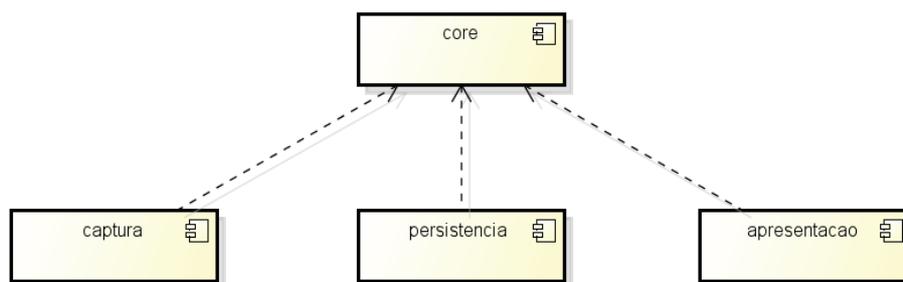


Figura 1. Pacotes que compõem a arquitetura proposta

No pacote ‘Core’ residem os componentes básicos para o funcionamento da aplicação, contendo também o Modelo de proveniência, que representa os dados de proveniência manipulados, conforme descrito na seção anterior. O pacote ‘Persistencia’ é responsável pela persistência dos dados de proveniência. Os dados a serem persistidos são representados de acordo com o modelo de proveniência presente na camada core. As tarefas de leitura e gravação dos dados são de responsabilidade deste pacote.

O pacote ‘Captura’ contém os componentes que são utilizados para a captura dos dados de proveniência, através do monitoramento das atividades da aplicação, e o pacote ‘Apresentacao’ tem a responsabilidade de apresentar as informações de proveniência ao usuário. A apresentação dos dados de proveniência visa a fornecer uma visualização gráfica dos dados de proveniência capturados facilitando a análise do histórico de um objeto

O mecanismo de proveniência fornece um registro dos dados de proveniência, gerados a partir da interação do usuário com o sistema em tempo de execução. A cada iteração do usuário com o sistema são capturados dados de proveniência que possibilitem o reconhecimento de como um artefato foi gerado e as informações utilizadas para tal. Após a finalização da captura dos dados de proveniência relativos à criação do mapa, os mesmos são inseridos e persistidos no objeto final que representa o mapa para que possam posteriormente ser consultados. Esta característica está presente para evitar perda de informações de proveniência devido a algum erro no sistema que afete o seu funcionamento.

A recuperação dos dados de proveniência pode ser feita acessando-se diretamente as informações persistidas em arquivo no formato XML. Os usuários que não conhecem tal formato podem ter dificuldades para entender seu conteúdo, principalmente quando a quantidade de dados de proveniência encontrados no arquivo é

muito grande. Nesses casos, o mais indicado é que os dados sejam visualizados por meio da interface gráfica disponibilizada pelo sistema.

Para a representação gráfica da estrutura persistida com as informações de proveniência, é gerada uma conversão das informações de proveniência coletadas em uma linguagem de marcação para a especificação DOT, uma linguagem textual para descrição de grafos com base em uma notação de fácil compreensão e processável por ferramentas específicas da Petrobras.

4. O Sistema Desenvolvido

A partir desta arquitetura foi desenvolvida uma ferramenta protótipo como extensão (*plug-in*) para o Sigeo3, um arcabouço de desenvolvimento de software geocientífico utilizado pela Petrobras. Utilizando-se o protótipo criado, realizaram-se testes simulando a realização do processo de caracterização geológica para o registro e posterior visualização das informações de proveniência.

As imagens apresentadas na Figura 2 têm como objetivo demonstrar como os dados de proveniência são importantes para se verificar a qualidade atingida na geração de um mapa geológico de uma área de interesse. A Figura 2a representa um mapa onde são plotados os pontos onde foram realizadas medições de uma determinada variável geológica (profundidade). Este conjunto de dados serviu de entrada para o processo de caracterização geológica.

As Figuras 2b e 2c representam dois mapas distintos obtidos a partir do processo de caracterização geológica tendo como base as medições representadas na Figura 2a. Naquelas figuras, os valores plotados foram calculados usando as técnicas de predição selecionadas. Para as áreas em branco informam que não há informações disponíveis. Os pontos em cores mais intensas (quentes) representam valores de profundidade maiores e cores menos intensas (frias) são relativas a valores menores de profundidade. As profundidades utilizadas têm os seus valores iniciados em 0 (zero), que é o menor valor, e seguem adiante para valores negativos, que representam maiores profundidades, ou seja, uma profundidade de valor -1000 é maior que uma profundidade de valor -500.

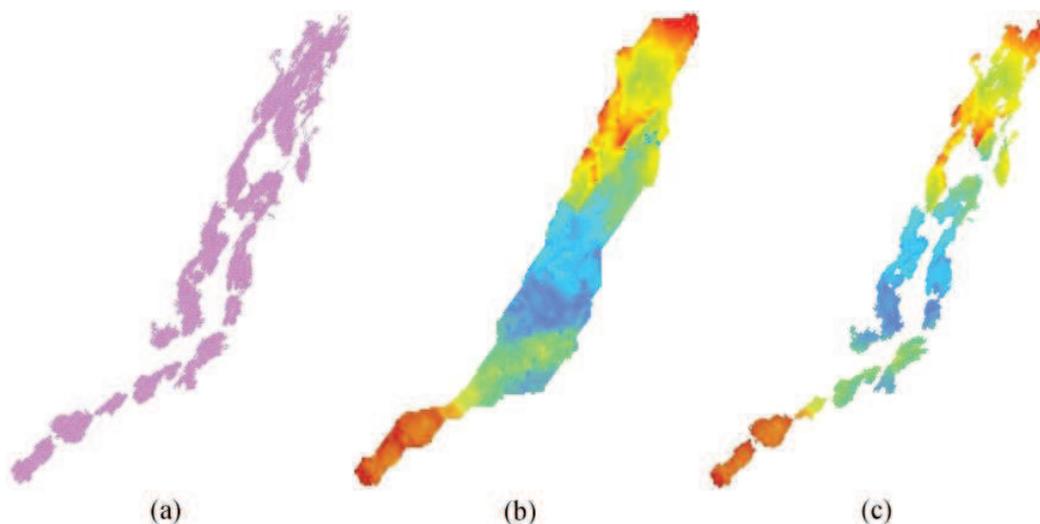


Figura 2. Conjunto de pontos de uma área geológica

Na Figura 2b pode ser observado que a informação apresentada possui grandes diferenças em relação à outra, pois várias áreas onde não havia informação foram preenchidas com a aplicação do algoritmo escolhido para o cálculo do valor dos nós do grid, assim como os parâmetros utilizados no mesmo. Na Figura 2c é apresentado um mapa no qual a informação fornecida está bem mais próxima daquela originalmente representada pela Figura 2a, sendo que esta proximidade se deve ao fato de que o algoritmo de predição e a configuração utilizada não inseriram grandes distorções no mapa gerado, ou seja, o tamanho das áreas vazias (áreas em branco) e das formas do mapa são semelhantes às do mapa de origem.

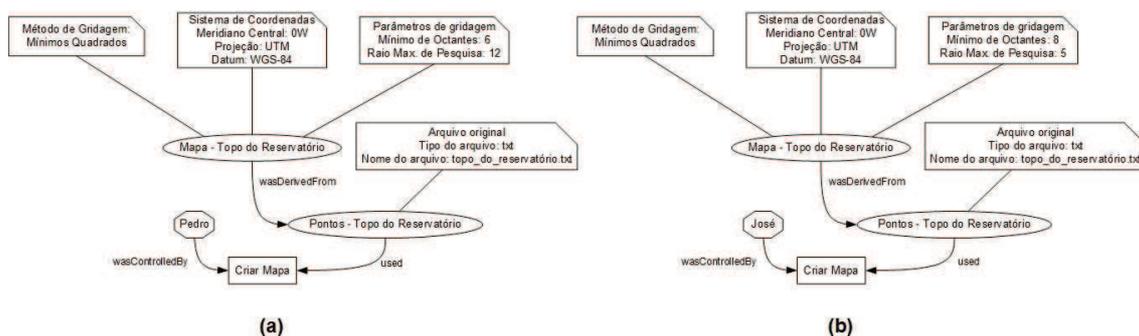


Figura 3. Grafos de proveniência relativos aos mapas representados nas Figuras 2b e 2c

As Figuras 3a e 3b exibem os grafos de proveniência gerados a partir das informações de sistema de coordenadas, algoritmos e dados originais utilizados no processo de criação dos mapas representados nas Figuras 2b e 2c, respectivamente. Sem esta informação de proveniência, é impossível realizar a avaliação do quão bom é o resultado final, por isso a importância de se conhecer a proveniência dos dados originais utilizados para se gerá-lo. Além disso, o recolhimento de informações de proveniência referente aos algoritmos utilizados pode ajudar a determinar se o método de predição é adequado para a situação geológica apresentada pelo conjunto de pontos originais.

5. Conclusões

Neste trabalho foi proposta uma arquitetura visando a oferecer suporte à implementação da coleta e recuperação de dados de proveniência em aplicações que façam uso das ideias apresentadas neste trabalho. A representação de dados de proveniência se baseou no modelo OPM, e estes são persistidos em arquivos utilizando-se a linguagem de marcação XML. Para a apresentação dos dados de proveniência, estes foram exportados no padrão DOT, gerando uma descrição do grafo que representa as informações de proveniência e pode ser processado por ferramentas específicas. Além disso, foi construída uma ferramenta como extensão (*plug-in*) para sistemas de caracterização desenvolvidos no Sigeo3. Para a geração da imagem foi utilizada uma implementação, ou melhor, um invólucro escrito utilizando-se a linguagem Java para a geração do grafo na especificação DOT (GraphViz) e a chamada do binário responsável pela geração da imagem.

Como trabalhos futuros, vislumbramos a possibilidade de criar um mecanismo para gerar imagens somente de informações de interesse de um usuário, ou seja, o

usuário escolheria as informações desejadas para a imagem do grafo. Além disso, consideramos importante realizar uma avaliação de desempenho sobre o acesso aos dados na arquitetura proposta.

Referências

- Barga, R.; Digiampietri, L. A. (2008) Automatic Capture and Efficient Storage of eScience Experiment Provenance. *Concurrency and Computation. Practice & Experience*, v. 20, p. 419-429.
- Bryant, I. e Xu, P. (2011). Integrating exploration tools. *Offshore Engineer*, agosto.
- Cheney, J. et al. (2009). Provenance: a future history. *Proceedings of the 24th ACM SIGPLAN conference companion on Object oriented programming systems languages and applications*. Anais. ACM.
- Freire, J. et al. (2008). Provenance for Computational Tasks: A Survey. *Computing in Science and Engineering*, v. 10, n. 3, p. 11–21, Maio.
- Gaspar, W., Braga, R., & Campos, F. (2010). SWfPS: Proposição Um Sistema de Proveniência de Dados e Processos no Domínio de Workflows Científicos. *Anais do VI Simpósio Brasileiro de Sistemas de Informação (SBSI 2010)*.
- Houlding, S. W. (1995). *3D Geoscience Modeling: Computer Techniques for Geological Characterization*, New York: SPRINGER. p. 7-37.
- Ludäscher, B. et al. (2005). Scientific Workflow Management and the Kepler System. *Concurr. Comput.: Pract. Exper.* Anais.
- Jones, T. A.; Hamilton, D. E.; Johnson, C. R. (1987). *Contouring Geologic Surfaces With The Computer*, Ed. Springer.
- Mendes, L. F. C. (2011). SciProv: Uma arquitetura orientada a agentes para gerenciamento, busca e recuperação de artefatos científicos. 176f. Dissertação (Mestrado em Modelagem Computacional) – Universidade Federal de Juiz de Fora, Juiz de Fora.
- Moreau, L. et al. (2011). The Open Provenance Model core specification (v1.1). *Future Gener. Comput. Syst.*, v. 27, n. 6, p. 743–756, Junho.
- Oinn, T. et al. (2007). Taverna/myGrid: Aligning a Workflow System with the Life Sciences Community. In: TAYLOR, I. J. et al. (Eds.). *Workflows for e-Science*. Springer London. p. 300 –319.
- Thomas, J. E. (2001). *Fundamentos de Engenharia de Petróleo*. 3. ed. Rio de Janeiro: Interciência, p. 23-53.
- Simmhan, Y.; Plale, B. e Gannon, D. (2005). A survey of data provenance techniques: Technical Report TR-618.