

5 Conclusões

Sistemas de informação, desenvolvidos para diversos setores econômicos, necessitam com maior frequência de capacidade de rastreabilidade dos dados. Para habilitar tal capacidade, é necessário modelar a proveniência dos dados. Proveniência permite testar conformidade com a legislação, repetição de experimentos, controle de qualidade, entre outros. Habilita também a identificação de participantes (determinantes ou aderentes) como pessoas, organizações, agentes de software entre outros e permite associá-los a atividades, eventos ou processos. Pode ser utilizada para estabelecer níveis de confiança para as transformações dos dados.

Nesta dissertação propusemos um modelo genérico de proveniência criado com base no alinhamento de recortes de ontologias de alto nível, padrões internacionais e propostas de padrões que tratam direta ou indiretamente de conceitos relacionados à proveniência, entre outras fontes. Em seguida, descrevemos a origem das decisões de projeto, a metaproveniência, que permitiram a construção de um modelo genérico, sugerido como padrão de projeto para proveniência. Analisamos por fim uma aplicação prática do modelo, precedida por aplicações conceituais, que completam esta avaliação.

Avançamos cientes de um *slogan* muito característico na engenharia: “Essencialmente, todos os modelos estão errados, mas alguns são úteis.” (Box & Draper, 1987). De fato, cada vez que aprofundamos o estudo do modelo de proveniência proposto por este trabalho, tornamos a identificar pontos passíveis de correção; mesmo assim, as aplicações do modelo nos permitiram ratificar sua utilidade: senão como um padrão de projeto – porque ainda carece de avaliação mais ampla – ao menos considerando as conclusões apresentadas aqui. Começamos comentando cada objetivo delineado no início deste documento (seção 1.4).

O1. Identificamos os seguintes caminhos que alavancam o reuso e aumentam as possibilidades de integração: o amadurecimento de uma metodologia de construção de modelos com base na identificação de padrões, inspirada na engenharia de ontologias, inicialmente adaptada para o desenvolvimento de nosso modelo; a identificação de um mecanismo ainda pouco divulgado e utilizado, que é a organização de repositórios, catálogos e bibliotecas de modelos de referência e padrões de projeto, abertos publicamente

para consultas. Esse mecanismo permite o mapeamento e avaliação de conceitos (classes e propriedades) de forma estruturada, que ainda é pouco explorado ou até desconhecido para muitos domínios, facilitando a pesquisa e apoiando a identificação de invariantes e fragmentos.

O modelo conceitual de proveniência proposto neste trabalho é quase que em sua totalidade uma visão de classes e relacionamentos de outros modelos. Precisa ser refinado para que seja possível um veredicto se é ou não possível um modelo puro, onde todas as classes e propriedades são importadas. No modelo proposto neste trabalho, alguns pequenos ajustes de propriedades e classes foram necessários para sua consolidação, por isso retificamos que é um modelo em construção. Ao longo de seu desenvolvimento, as anotações das decisões de projeto do modelo foram fundamentais para registrar a proveniência do próprio modelo e, neste exercício, atentar para quão importante é, de fato, a existência da proveniência para a repetição de experimentos: cria uma base segura para a evolução de outros trabalhos.

A completude ainda é um desafio perseguido pelo modelo de proveniência genérico proposto, porque exige que cada entidade do modelo seja a representação conceitual para, pelo menos, uma instância do domínio de aplicação, e que todas as instâncias do domínio de aplicação também possuam um mapeamento para algum respectivo conceito geral no modelo de proveniência. Quando tal cenário não for satisfeito, extensões do modelo devem ser alcançáveis para satisfazer, então, a completude.

O modelo de proveniência adotou e adaptou a metodologia de construção de padrões de projeto conceitual de ontologias e, por isso, também consultou padrões de projeto de ontologia na busca por potencializar o seu reuso. O núcleo do modelo de proveniência apresenta conceitos também presentes no padrão D&S da ontologia DOLCE, o que facilita extensões análogas a este último. Ratificaremos mais detalhadamente adiante o princípio de reuso, resumindo o resultado obtido.

O2. A consulta a diferentes fontes, entre elas as ontologias de alto nível, permitiu:

- A confirmação de invariantes algumas vezes produzindo alinhamentos;
- O reconhecimento de que mesmo conceitos não unânimes podem ser úteis durante o processo de construção do modelo, já que existem sobreposições semânticas que afloram apenas quando fragmentos (conjunto de conceitos e seus relacionamentos) são avaliados;

- Observar que o trabalho desenvolvido, especialmente em relação às ontologias de alto nível, exige um grande esforço (manual) do projetista para a pesquisa e identificação de tais invariantes;
- Observar que os resultados dos alinhamentos de conceitos quase nunca adotam critérios semelhantes na sua apresentação e organização, o que dificulta a análise. Um caminho para uniformizar o alinhamento de conceitos é proposto por (Scharffe & Bruijn, 2005), como uma linguagem, e pode ser resumido pela Tabela 25.

Tabela 25: Linguagem para alinhamento entre ontologias (Scharffe & Bruijn, 2005)

Elemento da Linguagem	Tipo do Mapeamento
ClassMapping	Entre duas classes
AttributeMapping	Entre dois atributos
RelationMapping	Entre duas relações
ClassAttributeMapping	Entre uma classe e um atributo
ClassRelationMapping	Entre uma classe e uma relação
ClassInstanceMapping	Entre uma classe e uma instância
IndividualMapping	Entre duas instâncias

O3. A ISO 21127:2006 é a origem de mais de 80% - classes e propriedades - do modelo de proveniência (mínimo), que possui expansões por importação de classes e propriedades de outros projetos de renome, por exemplo, o FRBRoo, que está em processo de harmonização desde 2003 com o projeto que originou a ISO. O modelo de proveniência também importa duas das classes da Matriz de Perguntas de Weiderhold (1993) e uma classe da ontologia de topo de Sowa (1999). Incorpora ainda o núcleo de um importante padrão de projeto conceitual (D&S) da ontologia DOLCE. No estudo dos padrões internacionais, identificamos também que não há uma norma internacional dedicada à proveniência, mas a proveniência tem papel chave em padrões internacionais para a qualidade da informação e para a preservação digital. Por fim, devido à sua construção genérica, outras expansões podem ser projetadas a partir das fontes de construção ou ainda de outras diferentes delas favorecendo a interoperabilidade com padrões de metadados já estabelecidos e reconhecidos internacionalmente.

O4. Dos três projetos (FRBRoo, Harmony e INDECS) estudados, os dois primeiros apontaram exercícios similares de harmonização de seus modelos com o padrão ISO 21127:2006, internacionalmente reconhecido. No terceiro e último, a ISO foi fonte de referência. Todos, sem exceção, apresentaram

cobertura para conceitos de proveniência, em especial para a noção de evento, fundamental para registrar a história.

O5. A aplicação do modelo conceitual de proveniência para a construção de um modelo físico nos permitiu observar, a partir do foco nos eventos (*milestones*), um mapeamento completo para os conceitos da ferramenta Trac utilizada no domínio de gerenciamento de configuração de software. De forma complementar, exploramos aplicações conceituais do modelo de proveniência (*desktop* semântico e centro de informações) que permitiram identificar e compreender componentes de captura e filtro de proveniência, além de acomodar mapeamentos de conceitos-chave destes domínios em elementos do modelo de proveniência. Por fim, no estudo da generalização de *Design Rationale*, observamos que os conceitos de proveniência cobrem inicialmente os principais elementos do modelo Kuaba.

Com relação a trabalhos futuros, colocamos que há um potencial a ser explorado na direção das extensões aplicadas ao padrão de projeto D&S. O modelo conceitual de proveniência possui a propriedade especializada⁶⁷ “P14.1 in the role of” e classes CID5 Means, CID2 Participant e E5 Event - que podem representar o núcleo do padrão de projeto D&S, respectivamente *Functional Role*, *Course of Events*, *Endurant* e *Perdurant*. Por isso, sugerimos que se avalie de que maneira seria possível oferecer cobertura à captura da proveniência dentro de planos (*planning*). A expansão DISP deve ser estudada mais aprofundadamente porque a princípio, oferece a representação das noções *Source* e *Product*, respectivamente um participante que deve estar presente no início e ao final de uma atividade/evento/processo, mas que não tem controle sobre o que acontece. Gangemi (2005) apresenta uma extensão ao D&S para contemplar *planning*. Nessa direção, talvez um caminho seja aplicar o uso de Cálculo de Eventos (*Event Calculus*) a planos (*planning*) que contemplem a captura da proveniência, teoricamente, como destaca (Aydin et al., 2007). Indo mais além, a representação do conhecimento relativo à proveniência poderia ser explorada de forma semelhante ao projeto Inference Web⁶⁸.

⁶⁷ Especializada de “P14 carried out by (performed)” importada do modelo CIDOC CRM para o modelo mínimo, que tem classe-domínio E7 Activity e classe-imagem CID2 Participant.

⁶⁸ <http://www.inference-web.org/>

Pasin et al. (2007) assume como base as classes e propriedades da ISO 21127:2006 para modelar documentos, eventos e idéias e as estende, adicionando classes e propriedades de outras ontologias, para contemplar incertezas, pontos de vista e informações contraditórias. Um trabalho futuro pode explorar extensões similares para o modelo conceitual de proveniência. Em uma direção complementar, Prat & Madnick (2008) apresentam uma abordagem de proveniência para medir quão “acreditáveis” são os dados (*data believability*).

Sugerimos ainda que interseções com outros padrões de metadados sejam avaliadas. Por exemplo, a descrição de dados não estruturados proposta pelo modelo DRM⁶⁹ (*Data Reference Model*) ou ainda, o *framework* IDEF3⁷⁰.

O progresso da pesquisa precisa desafiar a estabilidade dos padrões? Possivelmente. Para alavancar pesquisa, novos padrões (modelos) surgirão e ganharão importância frente aos antigos padrões estáveis. Neste trabalho, exploramos a possibilidade de um novo padrão para proveniência, genérico, que facilita a interoperabilidade. Em nossa visão, a estratégia que alavanca *ordem* (estabilidade) e *progresso* é a modelagem conceitual baseada em padrões. Portanto, em primeiro lugar, a comunidade acadêmica e o mercado precisam compreender o desafio. Como entendê-lo? Quem precisa superá-lo? Quais ferramentas já existem? Por que faltam? Qual o investimento necessário? Responder a essas perguntas é compreender a origem do desafio: é conhecer as necessidades de manter proveniência no domínio de aplicação em questão.

⁶⁹ http://www.ocio.usda.gov/e_arch/doc/DRM_2_0_Final.pdf

⁷⁰ <http://www.idef.com/IDEF3.html>