

3

Modelo de Proveniência

Neste capítulo, apresentamos as preliminares à construção do modelo conceitual genérico para proveniência (seção 3.1). Então, discutimos a estratégia (seção 3.2) e tática (seção 3.3) do projeto do modelo de proveniência. Apresentamos a proveniência da proveniência do modelo (seção 3.4) e descrevemos o modelo de proveniência (seção 3.5). Por fim, fazemos as considerações finais (seção 3.6).

3.1.

Preliminares

Considere as diferentes interfaces de recuperação de informação que estão presentes em nosso dia a dia no ambiente pessoal, corporativo ou acadêmico, através do uso de sistemas Web ou não. Cada uma delas oferece um nível de detalhe diferente apresentando maior ou menor variedade e qualidade de metadados disponíveis para a busca.

Ainda assim, as respostas retornadas expõem apenas um pequeno grupo de todos os metadados de fato existentes. Muitas vezes, é através de algum conhecimento já adquirido que estabelecemos as relações que poderiam estar explícitas pelos metadados. Então, quando localizamos os objetos de interesse, manualmente criamos os metadados implícitos, relacionado os objetos entre si. Esses metadados representam a proveniência como anotação realizada após a criação do objeto.

Sinclair (2006c) argumenta que há vezes em que as relações entre os objetos já existem, mas não há registro de quando, onde, como ou porque se estabeleceram. Assim, somos forçados a criar hipóteses para visualizar o contexto histórico da relação a partir apenas de valores dos objetos retornados em uma busca. Acrescenta ainda que, varrendo os resultados da busca, conjecturamos relacionamentos mais profundos entre os itens retornados. Por isso, enfatiza que os mecanismos de consulta e visualização deveriam ser capazes de representar os relacionamentos entre os itens retornados. Por fim, sugere que o caminho para essa representação seria alcançado através do uso de modelos de metadados mais ricos que pudessem capturar o histórico completo dos objetos representados. Mas alerta que esse modelo deve superar

as diversidades das representações de materiais e mídias, bem como sua heterogeneidade.

Bearman et al. (1999) afirma que, para que os metadados sejam intercambiáveis, os relacionamentos implícitos devem ser explicitados. De fato, virtualmente nada pode ser compreendido ou interpretado sem sua relação com o contexto. Portanto, é necessário relacionar as entidades, sejam elas eventos, documentos digitais ou físicos, ou abstrações.

Para compreender como os relacionamentos serão estruturados, apresentamos as estratégias de projeto possíveis (seção 3.2), destacando a adotada em nossa modelagem.

3.2. Estratégia de Projeto (Proveniência no Centro)

As estratégias de projeto de modelagem, intrínsecas ao estudo de proveniência, podem ser compreendidas por duas formas (Globe, 2003):

- Derivação: é um caminho dentro de um fluxo (*workflow*), um script ou consulta; são as ligações (*links*) entre objetos que usualmente estão representadas pelas arestas de um grafo direcionado. A derivação é uma explicação de onde, quem, ou como alguma coisa foi produzida ou derivada.
- Anotações: são dados representadas como anotações dos objetos de informação, de forma estruturada, semi-estruturada ou em forma de texto livre. Os relacionamentos entre itens também podem se estabelecer a partir das anotações. Podem ser explicações relativas ao item, do porquê, quando, onde, quem, qual ou como.

A derivação então dá origem à estratégia centrada em evento (*event-centric*) ou processo (*process-centric*) onde as mudanças estão no centro. Por outro lado, a anotação está vinculada à estratégia centrada no dado (*data-centric*), algumas vezes também chamada de centrada no documento ou centrada no recurso (*document-centric* ou *resource-centric*).

Os projetos Dublin Core, OAIS e FRBR, que cobrem conceitos de proveniência (seções 2.1.2.1, 2.4.2.1 e 2.4.3.1 respectivamente), são exemplos de estratégia centrada no recurso. Lagoze (2001) explica que, das perspectivas técnicas e econômicas, por exemplo, o conjunto elementar do DC permite que valores sejam associados a um conjunto finito de propriedades. Nesta estratégia

o documento ou recurso é promovido à classe principal (*first-class*) enquanto que entidades como agente, evento, ação entre outras são apenas complementares.

Por outro lado, Lagoze (2001) esclarece que a necessidade de representações mais ricas, sugere uma abordagem mais expressiva dos modelos e vocabulários, onde deve ser possível definir e delimitar múltiplas entidades.

Essa necessidade é atendida pela estratégia centrada no evento, como adotam os projetos CIDOC CRM, FRBRoo, INDECS, HARMONY, (respectivamente seções 2.4.2.2, 2.4.3.1, 2.4.3.2 e 2.4.3.3). Hunter (2001) ressalta, por exemplo, que a adoção de uma visão centrada em eventos permite a modelagem de relacionamentos entre várias manifestações de uma criação. Adiciona que a estratégia habilita também a associação de propriedades entre eventos e agentes envolvidos no ciclo de vida de um recurso. Comenta ainda que essa estratégia de modelagem de metadados permite múltiplas visões para a recuperação dos dados.

Em nosso projeto, adotamos uma estratégia composta, que definimos como centrada na proveniência (*event-centric + event-relations + related resources*) porque permite a representação de agentes, ações, recursos, contexto entre outros, transformando conhecimentos implícitos em explícitos, favorecendo a interoperabilidade semântica (Yanosy, 2006).

Exploramos a tática (seção 3.3) que complementa a estratégia adotada, a partir da discussão de padrões de projeto de ontologia conceitual.

3.3. Tática de Projeto (PPCO)

Há problemas de modelagem que se repetem com frequência nos projetos de diferentes aplicações. Quando isto ocorre, a solução adotada deve preferencialmente ser descrita de forma genérica, adequada para cobrir várias classes de aplicações. Essa tática é conhecida como Padrão de Projeto (*Design Pattern*).

Na área de informática, o termo apropriado é Padrão de Projeto de Software (*Software Design Pattern*), mas seu uso abreviado ao termo original também é bastante comum. Gamma et al. (1995) destaca que os padrões de projeto são um método geral e abstrato de solucionar questões de modelagem, que tem se mostrado consistentemente eficaz na resolução de problemas da engenharia de software.

Aranguren (2005) apresenta os padrões de projeto de ontologia ODP (*Ontology Design Pattern*) como soluções de modelagens abstratas para problemas conhecidos da engenharia de ontologias. Alguns padrões de projeto de ontologia são propostos pelo Grupo de Engenharia e Padrões de Ontologia do W3C, abreviadamente conhecido como OEP³¹ (*Ontology Engineering and Patterns Task Force*).

Os padrões de projeto conceitual de ontologias PPCO (tradução da respectiva abreviação em inglês CODEP: *Conceptual Ontology Design Patterns*) tem uma aplicação análoga aos padrões de projeto de software (Gangemi, 2006).

Entretanto, Aranguren et al. (2007) chama atenção para a diferença entre CODEPs e ODPs: o conceito de CODEP (Gangemi, 2005) parece próximo ao conceito ODP, porém difere em nível de expressividade. CODEPs representam padrões conceituais e abstratos em linguagens que não necessariamente permitem inferência, enquanto que ODPs são representados através de linguagens (por exemplo, OWL) sustentadas por um formalismo (com Lógica de Descrição). Esta dissertação adotou uma tática de projeto inspirada na metodologia de construção de padrões de projeto conceitual de ontologias para o desenvolvimento do modelo conceitual genérico para proveniência (seção 3.5).

Adicionalmente a construção do CODEP, Gangemi (2005) sugere sua publicação (codificação e anotações) em repositórios e a disponibilização de um serviço de consulta, por exemplo, uma biblioteca ou um catálogo acessível via Web. Identificamos apenas um catálogo³² *on line* que permite pesquisas estruturadas nos padrões disponibilizados, enquanto que os demais repositórios³³ de padrões disponibilizam acesso, mas não consultas estruturadas.

Por hora, interessa-nos destacar um padrão que exemplifica a reificação³⁴ de relacionamentos indexados pelo tempo. Algumas classes de um padrão de ontologia representam reificações semelhantes à destacada pelo CODEP da

³¹<http://www.w3.org/2001/sw/BestPractices/OEP/>

³² <http://wiki.loa-cnr.it/index.php/LoaWiki:CPRpository>

³³ <http://odps.sourceforge.net/odp/html/index.html>
<http://www.omwg.org/TR/d7/patterns-library/>

³⁴ <http://www.thefreedictionary.com/reification> define reificação como: *to regard or treat (an abstraction) as if it had concrete or material existence*

Figura 30. Esse padrão apresenta os principais elementos de nossa estratégia de projeto: evento (ou processo), a evidência de relacionamentos e objeto (ou recurso).

O CODEP *Time-Index-Participation* da Figura 30 apresenta a reificação do padrão *Participation* (Masolo et al., 2003). Essa reificação representa uma configuração para: um evento, as entidades que participam desse evento e o intervalo de tempo no qual o evento ocorre.

Apresentaremos a proveniência da proveniência (seção 3.4), explicando a origem do modelo proposto. Em seguida, descrevemos o modelo de proveniência (seção 3.5).

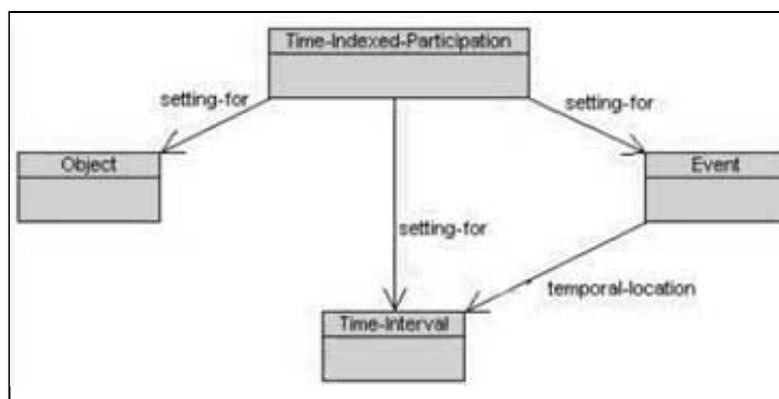


Figura 30: CODEP Participation reificado (Gangemi, 2005)

3.4. Metaproveniência

O esboço de um modelo de proveniência se manifestou a partir das noções abstratas de proveniência e do estudo de perguntas elementares ao estilo 5W1H³⁵. O segundo ponto importante da modelagem foi o foco na noção de evento, mas que inicialmente não havia se mostrado evidente pelas definições de dicionários. Na verdade, o conceito de evento estava implícito nas noções de história (*state of affairs*). A noção de evento foi então ratificada pela análise das definições de grupos tradicionais de pesquisa de metadados. Então,

³⁵ 5W1H origina-se das seis palavras em inglês: **What, When, Who, Why, Where** e **How**. É um mnemograma que se popularizou na linguagem empresarial porque ajuda a lembrar dos seis pontos principais de um plano de ação.

o estudo da proveniência baseada em eventos amadureceu com a seleção e identificação de normas e projetos que tinham como elemento central o conceito evento, e que ao mesmo tempo apresentavam uma cobertura para os conceitos abstratos de proveniência. Por fim, com o desafio de tornar o alinhamento de esquemas um problema tratável, projetamos um modelo de proveniência como um padrão conceitual através da busca de invariantes em todo o contexto estudado.

Após esse breve histórico da criação do nosso modelo de proveniência, ao longo desta seção destacaremos as partes do contexto de proveniência mais relevantes e que influenciaram mais significativamente a construção do modelo. Identificaremos os reusos possíveis e selecionados, assegurando que as classes importadas sejam apenas as necessárias e suficientes para a representação do modelo conceitual genérico de proveniência que nos referenciaremos também, abreviadamente, como modelo mínimo. O foco no modelo mínimo é um dos caminhos para assegurar a consistência da utilidade do modelo. As classes importadas estarão associadas inicialmente a conceitos abstratos, que são parte de nossa ferramenta cognitiva, mas que serão descartadas tão logo o modelo mínimo seja alcançado. Lembramos também que, neste capítulo, modelo de proveniência e modelo conceitual de proveniência são sinônimos para modelo conceitual genérico de proveniência.

Criamos uma generalização para conceito de proveniência a partir do termo proveniência do padrão Dublin Core (seção 2.1.2.1). A proveniência a que se refere nosso modelo (seção 3.5) é a de um conjunto de descrições de qualquer mudança de um objeto de informação, que registre sua história e que seja significativa para resguardar autenticidade, integridade e interpretação. Esse contexto inicial para proveniência é a justificativa principal para considerarmos o conceito de Evento associado à metaclassa *What*.

Dos aspectos de projetos de proveniência (seção 2.2) adotamos a estratégia centrada em eventos como derivação da estratégia orientada a processos. Com isso, ratificamos que o conceito abstrato *What* e a noção de evento são centrais ao modelo de proveniência. Não obstante, argumentamos que a proveniência em forma de anotação chamou atenção para a importância dos objetos de informação propriamente ditos, levando-nos a avaliar a modelagem de proveniência orientada ao dado. Descartamos esta última porque a centrada em eventos é mais expressiva, mas consideramos que de forma prática, o objeto de informação como produto é mais tangível.

Dentre as ontologias de alto nível (seção 2.3) estudadas, a ontologia que melhor representa o comprometimento com a integração a priori é a DOLCE (Masolo et al., 2003) porque foi a única que definiu um padrão de projeto de ontologia: D&S³⁶ (*Description & Situation*). Esse padrão é referência para outros reusos, por exemplo, as decomposições, “*Participation*”, “*Role->Task*”, entre outras, propostas na forma de CODePs por (Gangemi, 2005) e (Gangemi, 2006).

A Figura 31 exemplifica o padrão D&S, onde *Region*, *Endurant*, *Perdurant* e *Situation* são entidades primárias (“*first-class*”), portanto classes disjuntas. As classes *Parameter*, *Functional Role*, e *Course of Event* são do tipo *Description* e agregadas pela classe *S-Description* (contexto), também do tipo *Description*.

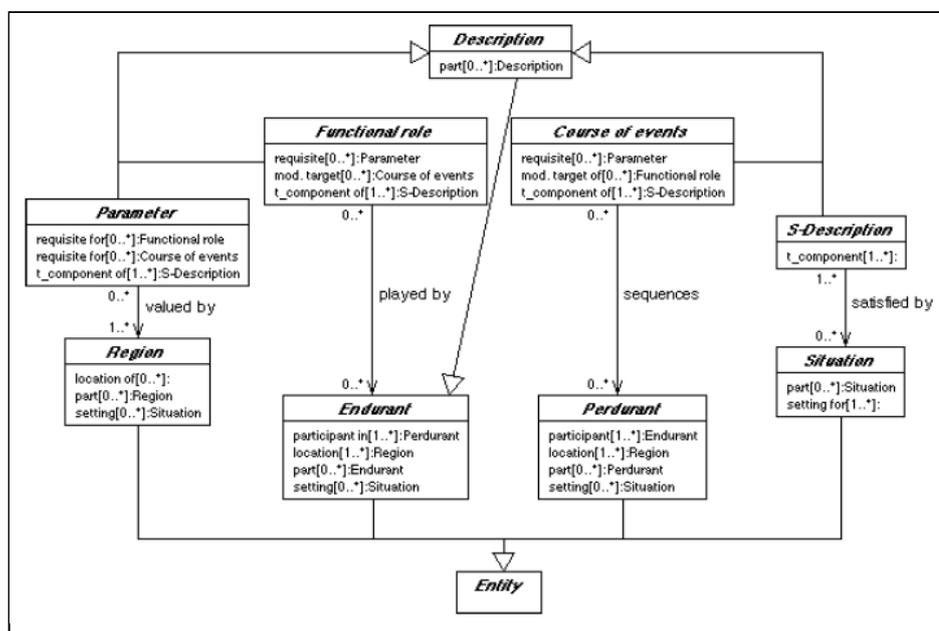


Figura 31: Padrão de Ontologia D&S proposto em (Masolo et al., 2003)

Uma forma mais simples de compreender o padrão D&S é sugerida por (Gangemi, 2006). A Figura 32 apresenta as principais classes e relacionamentos do padrão, e por simplificação didática, não representa todos os relacionamentos propostos pelo padrão D&S. Um contexto (instância de *S-Description*) define as classes reificadas, que por sua vez classificam as entidades primárias que são então uma configuração para uma situação (instância de *Situation*) que satisfaz o contexto.

³⁶ <http://wiki.loa-cnr.it/index.php/LoaWiki:DnS>

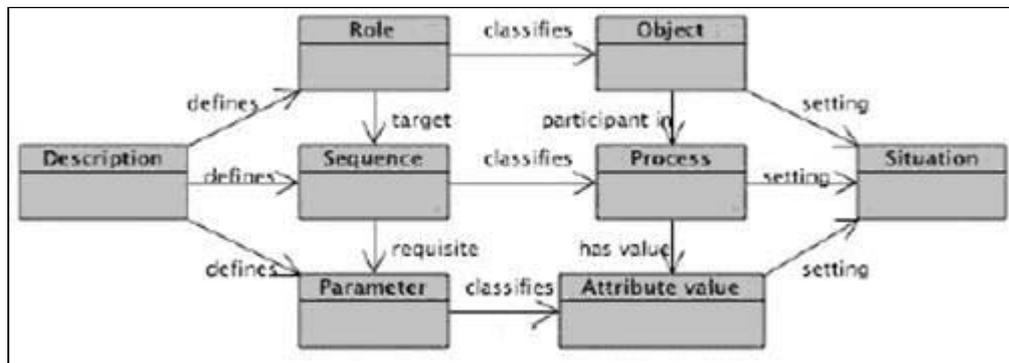


Figura 32: CODeP DnS³⁷ simplificado em (Gangemi, 2006)

Gangemi (2006) esclarece que esse padrão pode ser especializado para diferentes domínios de aplicação. Esse, como outros CODePs foram aplicados por mais de dois anos em projetos de sistemas de informação para a indústria pesqueira, seguradoras, integração de ontologias da biomedicina, controle de lavagem de dinheiro em instituições financeiras, acordos de níveis de serviço, aprendizado de ontologia biomolecular e contratos. Por fim, acrescenta ainda que CODePs também foram aplicados para gerenciamento de conteúdo digital.

Do D&S, não importamos a classe *Functional Role*, que estaria inicialmente, mapeada para o conceito abstrato *Which* porque a propriedade “P14.1 in the role of” importada da ISO 21127:2006 já representa esse conceito. Nesse padrão, outra classe nos chamou atenção: *Course of Events*. Essa classe foi uma candidata ponderada por nós em determinado momento da pesquisa para assumir o mapeamento para o conceito *How* porque suas instâncias são ordens esperadas. Optamos por não importar a classe *Course of Events* do padrão D&S, porque consideramos que ela restringe a possibilidade da representação para a pergunta Como (*How*). O conceito adotado é *Means*³⁸, importado da Matriz de Perguntas (*The Question Matrix*) de Wiederhold³⁹ (1993) para o modelo de proveniência.

³⁷ <http://www.loa-cnr.it/ontologies/ExtendedDnS.owl>

³⁸ Como (*How*) um resultado é obtido.

³⁹ Referimo-nos aqui a Chuck Wiederhold e não a Gio Wiederhold.

Buscando ancorar nosso modelo de proveniência em padrões internacionais (seção 2.4.2), mencionamos o padrão ISO/NP 8000-120 (*Data quality - Part 120: Master data: Provenance*), ainda não homologado, que ratifica a importância da proveniência para a qualidade dos dados, uma vez que esse padrão é parte da ISO 8000 em desenvolvimento com esse propósito.

Em seguida, descrevemos da ISO 14721:2003 (seção 2.4.2.1), equivalente a tradução brasileira da norma NBR 15472:2007 (seção 2.4.2.1), a parte dedicada à proveniência, apresentando os conceitos relacionados ao componente da informação de descrição de preservação. Essa norma ratifica a importância da modelagem de eventos, mas restringisse ao domínio arquivístico, limitando, portanto, as definições relacionadas à proveniência. Embora não tenhamos importado da norma brasileira nenhuma classe ou relacionamento, a sua homologação recente demonstra a importância deste padrão para a indústria nacional, bem como os esforços em adotá-lo para desenvolvimento de sistema arquivístico de caráter aberto, em diferentes instituições governamentais, como por exemplo, no Arquivo Nacional.

Concluindo o estudo de padrões internacionais, exploramos a ISO 21127:2006 (seção 2.4.2.2). A escolha desse padrão foi motivada por três razões. Primeiro, o padrão é a consolidação de dez anos de pesquisa na área de museologia (herança cultural) através do projeto CIDOC CRM. Segundo, porque o modelo desenvolvido é fortemente centrado em eventos e amplamente reconhecido pela comunidade internacional. E por fim, o modelo influenciou outras pesquisas (seção 2.4.3) que apresentam também coberturas para conceitos de proveniência e exploradas nesta dissertação.

De fato, a possibilidade do reuso e da extensão de uma ontologia reconhecida mundialmente, agrega ao modelo de proveniência maiores chances de interoperabilidade. Adicionalmente o sintoniza com uma infra-estrutura emergente que é a Web Semântica. O estudo aprofundado desse padrão ISO demonstra que o seu projeto de modelagem é adequado para a organização e descrição da proveniência porque modela eventos históricos e inerentemente dependentes do tempo, bem como objetos de informação relacionados por eles. Inicialmente desse modelo importaríamos as classes *Actor* e *Physical Thing*. Entretanto, estabelecemos como critério de seleção de classes a disjunção entre classes associadas à mesma metaclassa (conceito abstrato). De fato, não importamos as classes *Actor* e *Physical Thing* porque não são disjuntas. Para que o sejam, seria necessário resolver a interseção de instâncias: a classe *Biological Thing* - especialização indireta de *Physical Thing* – é a generalização

para *Person*, da mesma forma que *Actor* também o pode ser. O projetista deve ter atenção redobrada ao importar classes da ISO 21127 quando o modelo alvo tem como objetivo a disjunção. Com isso, optamos, portanto, pela importação da classe *Participant* da ontologia de topo proposta por (Sowa 1999) que é referência para a ontologia de alto nível SUMO (seção 2.3.3), como o conceito associado a *Who*. Essa decisão é interessante por duas razões: Sowa (1999) define subárvores abaixo do conceito *Participant* que chama de *Thematic Roles* e, foi ratificada pela presença do conceito *Participant* no modelo CIDOC CRM Core (seção 2.4.2.2). Descreveremos nos próximos parágrafos as metaproveniências relacionadas aos projetos estudados: FRBRoo, INDECS e Harmony.

Iniciamos pelo projeto FRBRoo (seção 2.4.3.1) que surgiu pela aproximação das comunidades de biblioteconomia e museologia que juntaram esforços para harmonizar ambos os modelos de forma a possibilitar uma integração mais efetiva entre essas áreas de conhecimento. O modelo de proveniência, enquanto proposta genérica, não deve atender exclusivamente ao domínio de biblioteconomia. Por isso, não iremos contemplar em nosso modelo mínimo os conceitos abstratos que são específicos para apenas alguns domínios. Entretanto ressaltamos que as classes *Work*, *Expression*, *Manifestation Product Type* e *Item* são consideradas elementares para o domínio de biblioteconomia e também importantes em outros domínios, por exemplo, e-commerce – projeto INDECS (seção 2.4.3.2). A representação concreta dos conceitos abstratos (*Work*, *Expression* e *Manifestation Product Type*) é *Item*. Todos esses últimos estão presentes no modelo FRBRoo.

Sugerimos que o modelo mínimo pode ser expandido a partir da importação das classes *Work*, *Expression*, *Manifestation Product Type* e *Item* respectivamente como especializações de *Conceptual Object*, *Information Object*, *Type* e *Information Carrier*. Todas estas quatro últimas são classes da ISO 21127:2006 – modelo CIDOC CRM (seção 2.4.2.2), mas apenas *Conceptual Object* e *Type* (especialização de *Conceptual Object*) foram importadas para o modelo mínimo porque representam respectivamente a noção de abstrações que são concebidas por participantes que têm intenção - em diferentes ontologias de alto nível (seção 2.3) são chamados de agentes – e, a noção de classificação. Com isso, a interoperabilidade com os modelos CIDOC CRM e FRBRoo seria assegurada, uma vez que a harmonização desses dois últimos modelos é a origem de nossa sugestão. A classe *Information Carrier*, que não faz parte do modelo mínimo, pode ser importada como especialização indireta de *Physical*

Thing sempre que algum domínio necessitar da noção de veículo que corresponde ao objeto físico propriamente dito. Podemos citar como exemplo de uma de suas instâncias (*Information Carrier*) a *Rosetta Stone*⁴⁰.

Com a noção de *Conceptual Object*, recuperamos os argumentos de (Sinclair et al., 2006b), que explicam que o CIDOC CRM Core é um subconjunto condensado de elementos de metadados da ISO 21127:2006 que captura os relacionamentos fundamentais que relacionam fenômenos (*things*), conceitos (*concepts*), pessoas (*people*), tempo (*time*) e lugar (*place*). Para essa captura, o CIDOC CRM Core oferece duas classes: *Relation* e *RelatedEvent*. A primeira representa instâncias de relacionamentos entre eventos e instâncias de “CRM Core” (entidade topo desse modelo). A segunda representa uma instância que é um auto-relacionamento entre eventos. Como a primeira é mais geral que a segunda, seria candidata a importação.

Ressaltamos que a avaliação feita durante o projeto do modelo, para reificar o conceito abstrato *Why*, foi ponderar as propriedades (relacionamentos) marcadas em negrito (Tabela 4 seção 2.4.2.2) repetidas (Tabela 16). As noções de *Belief* e *Desire* seriam cobertas respectivamente por “*was Influenced by*” e “*was motivated by*”. Analogamente, *Intention* cobriria “*was intended use of*”. Assim, o mapeamento para o conceito abstrato *Why* foi cogitado em determinado momento da pesquisa para ser dado não a partir de uma classe importada, mas sim através de propriedades importadas. De fato, não faria sentido um conceito abstrato apresentar propriedades.

A existência de propriedades (relacionamentos) qualificados que capturam a noção de razão – conceito abstrato *Why* – permitiu identificar a importância da representação de algumas relações, como uma classe reificada, nas expansões do modelo, analogamente ao modelo CIDOC CRM Core. Entretanto, esclarecemos que consideramos a noção de relação (*Relation*) também abstrata. Como alternativa, propomos o conceito – importado da Matriz de Perguntas (*The Matrix Question*) de Weiderhold (1993) – como a classe reificada adequada para o conceito abstrato *Why*. Conforme a estratégia adotada (seção 3.2), a explicitação da noção de relacionamento é fundamental. No modelo conceitual de proveniência esta noção é capturada pela classe *Reason*.

⁴⁰ A pedra Rosetta (*Rosetta Stone*) é o mais remoto “software” de linguagem de aprendizado do mundo, e tem seu original (*Information Carrier*) exposto no British Museum em Londres.

Tabela 16: Propriedades que capturariam parcialmente a noção de razão (*Reason*)

Id	Nome da Propriedade	Classe-Domínio	Classe-Imagem
P15	was influenced by (influenced)	E7 Activity	E1 CRM Entity
P17	- was motivated by (motivated)	E7 Activity	E1 CRM Entity
P19	was intended use of (was made for)	E7 Activity	E71 Man-Made Thing

Continuamos a descrever a metaproveniência, destacando que o projeto INDECS (seção 2.4.3.2) contribuiu para retificar os conceitos abstratos do projeto FRBR, apontando que a noção *Manifestation* é abstrata e não concreta como definida originalmente no modelo ER do FRBR. Adicionalmente, o projeto INDECS ratificou a importância da representação de relacionamentos, também definindo a classe *Relation* como a reificação da conexão (com atributos) entre duas ou mais entidades, conforme apresentado na Figura 33. No projeto INDECS a classe *Relation* é responsável por estabelecer relações entre concepções e percepções do mundo.

PUC-Rio - Certificação Digital Nº 0611888/CA

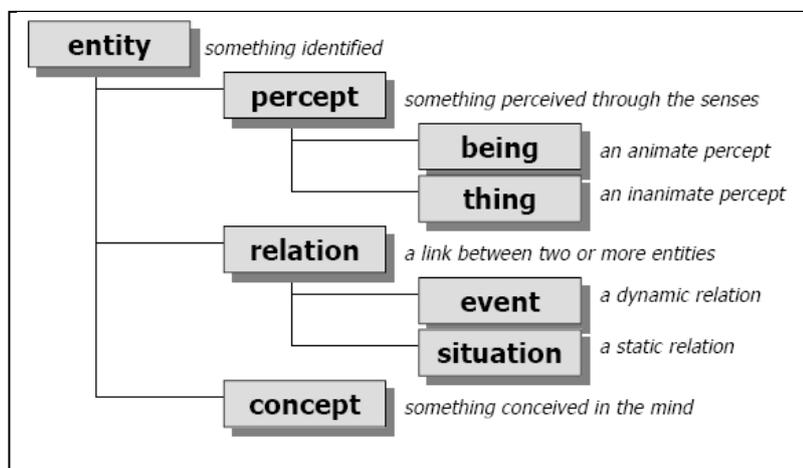


Figura 33: Entidades do modelo INDECS (Rust & Bide, 2000)⁴¹

⁴¹ As imagens desta fonte estão protegidas por Copyright. As utilizamos aqui amparados por http://en.wikipedia.org/wiki/Fair_use

Rust & Bide (2000) reconhecem que percepções (*percepts*) são seres inanimados (*things*) ou animados (*beings*), e relações (*relations*) como dinâmicas (*events*) ou estáticas (*situations*). Acrescentam que a classe evento desempenha um papel fundamental como cola do modelo: todos os metadados de relacionamentos são eventos por si, ou dependem de eventos para se estabelecer. Concluem que este mecanismo em transformar metadados em representações de eventos provê uma poderosa abordagem para a interoperabilidade.

Concluindo os projetos estudados, esclarecemos que do projeto Harmony também não importamos nenhuma classe para nosso modelo, mas frisamos que é um modelo significativo, ao ser, por si próprio, mais um exemplo de harmonização com a ISO 21127:2006, analogamente ao FRBRoo. O modelo ABC, produto do projeto Harmony, também importou classes do modelo FRBR. Mas a principal contribuição do estudo do projeto Harmony, já mencionada anteriormente, é expressa por (Lagoze & Hunter, 2001): os diversos modelos centrados em recursos são inadequados para expressar entidades como pessoas, lugares, idéias e especialmente transições temporais. O modelo ABC inclui as noções de Evento (*Event*) e Situação (*Situation*), que respectivamente capturam transições e propriedades existenciais, e ressalta ainda que são fundamentadas em modelagem de processos e extensões temporais para lógica de primeira ordem.

O projeto Harmony possui uma peculiaridade interessante: uma classe primária chamada *Abstraction*, noção não considerada como *first-class* pelo CIDOC CRM. Pelo contrário, *Persistent Item* tem como subclasse *Conceptual Object*, o que confunde um pouco os conceitos do que é Concreto ou Abstrato. As representações abstratas seriam então um “item persistente”, mas concretamente, apenas, através de uma instância da classe *Information Carrier*.

Uma possível analogia entre as classes primárias *Actuality* na ontologia de topo de Sowa (1999) e *Temporal Entity* no modelo da ISO 21127:2006 com *Endurant* e *Perdurant* do padrão D&S permite presumir que nosso modelo também contempla a estrutura núcleo definida nesse padrão. Assim como no padrão D&S, as classes *Functional Role* e *Course of Events* classificam respectivamente *Endurant* e *Perdurant*, o relacionamento qualificado “P14.1 in the role of” e a classe *Means* classificam *Participant* e *Event* que têm como classes raízes de seus respectivos ramos taxonômicos, as classes *Actuality* (Sowa, 1999) e *Temporal Entity* (ISO 21127 2006).

A Tabela 17 mapeia cada conceito abstrato de proveniência de interesse para as respectivas classes e especializações e identifica a origem da importação dos conceitos apresentados.

Para concluirmos o processo de construção do modelo, precisamos nos desfazer dos conceitos abstratos de proveniência que até o momento serviram ao papel de metaclasses (de classes primárias) para a realização dos mapeamentos. Considere, a partir de agora, que tenhamos chegado a um conjunto de classes importadas que seja satisfatoriamente genérico e interoperável, e que então podemos descartar as metaclasses (*Wh-questions*), utilizadas meramente como ferramenta cognitiva, sem prejuízo ao modelo e de forma transparente para o projetista.

Tabela 17: Mapeamento de conceitos abstratos de proveniência

Conceito Abstrato	Conceitos e Especializações	Origem do Conceito Importado
<i>What</i>	<i>Event</i>	CIDOC CRM
	- <i>Activity</i> - <i>Begin of Existence</i> - <i>End of Existence</i>	CIDOC CRM CIDOC CRM CIDOC CRM
<i>Who</i>	<i>Participant</i>	Sowa (1999)
<i>Where</i>	<i>Place</i>	CIDOC CRM
<i>When</i>	<i>Time-Span</i>	CIDOC CRM
	<i>Conceptual Object</i>	CIDOC CRM
<i>Why</i>	<i>Reason</i>	Matriz de Perguntas de Wiederhold (1993)
<i>How</i>	<i>Means</i>	Matriz de Perguntas de Wiederhold (1993)

Agora, é necessário identificar quais serão as nossas classes primárias e quais movimentações de classes devemos realizar para criar classes primárias de fato disjuntas. Primeiro, definimos a classe *Conceptual Object* como a generalização adequada para *Reason* e *Means*. Movemos as classes *Reason* e *Means* imediatamente para baixo de *Conceptual Object*, passando todos movidos, a serem especializações. O resultado obtido é apresentado na Tabela 18 que exhibe um detalhamento parcial da hierarquia de especialização, apenas com o propósito de manter o foco da discussão nas classes que nos interessam.

O conceito abstrato *Which - choice* segundo Wiederhold (1993) – não está presente associado a uma classe na Tabela 17, mas faz parte do modelo mínimo porque está representado pela propriedade qualificada “P14.1 in the role of” importada da ISO 21127:2006.

Tabela 18: Hierarquia parcial de especialização após descarte dos conceitos abstratos

Proveniência	Hierarquia	ID
ISO 21127:2006	<i>Event</i>	<i>E5</i>
ISO 21127:2006	- <i>Activity</i>	<i>E7</i>
ISO 21127:2006	- <i>Begin of Existence</i>	<i>E63</i>
ISO 21127:2006	- <i>End of Existence</i>	<i>E64</i>
Sowa (1999)	<i>Participant</i>	<i>CID2</i>
ISO 21127:2006	<i>Place</i>	<i>E53</i>
ISO 21127:2006	<i>Time-Span</i>	<i>E52</i>
ISO 21127:2006	<i>Conceptual Object</i>	<i>E28</i>
Wiederhold (1993)	- <i>Reason</i>	<i>CID4</i>
Wiederhold (1993)	- <i>Means</i>	<i>CID5</i>

Não apresentamos a classe E55 Type (ISO 21127:2006) na Tabela 18, para mantermos o foco da discussão nas classes mais relevantes, mas ratificamos que essa classe faz parte do modelo mínimo. Da mesma forma, não identificamos na hierarquia, porém importamos a classe E41 Appellation do modelo CIDOC CRM, como especialização de CID1 Provenance Entity, mantendo a coerência do corte feito na ontologia CIDOC CRM (ISO 21127:2006). As instâncias da classe E41 Appellation não são como instâncias de um banco de dados que são *surrogates* ou referências a entidades do mundo real. Cada instância dessa classe é apenas um *appellation* propriamente dito: o *appellation* “André” é nada mais que a palavra “André”, e não uma referência a uma pessoa, por isso, não carrega nenhuma semântica, o que explica a decisão de não colocá-lo como subclasse de E28 Conceptual Object.

A Tabela 18 inclui também uma coluna adicional que representa o identificador único da cada classe e utiliza a nomenclatura também importada da respectiva fonte. No caso das classes *Participant* (CID2), *Reason* (CID4) e *Means* (CID5), cujas fontes não oferecem um identificador único, optamos pela criação da seguinte nomenclatura para representá-lo: CID (*Class ID*) seguido imediatamente de um campo numérico incremental automático (CID1, CID2 e assim por diante). O identificador CID3 foi reservado para ser utilizado em

expansões ao modelo mínimo. Sugerimos ao projetista que avalie a reificação da propriedade “P14.1 in the role of” como classe CID3 Functional Role, o que compatibilizaria este conceito com a classe *Functional Role* do padrão D&S da DOLCE.

O resultado apresentado na Tabela 18 se não estivesse lastreado em padrões, fundamentando o comprometimento com a interoperabilidade e integração a priori, nem explicitasse quais são as respectivas origens dos conceitos, poderia ser visto então como um novo modelo. Da forma como está, seu verdadeiro propósito é o de uma “visão” que interrompe o ciclo vicioso de criação de novos modelos, ou ao menos restringe sua frequência.

Em resumo, até aqui descrevemos a origem do modelo de proveniência (mínimo) que reusa parcialmente as classes: da ontologia de topo de Sowa (1999), da Matriz de Perguntas de Wiederhold (1993) e da ISO 21127:2006. A hierarquia de especialização da Tabela 18 é a ponte (metaproveniência) para a consolidação das classes e relacionamentos do modelo conceitual genérico para proveniência (seção 3.5).

3.5. Modelo Conceitual

O modelo conceitual genérico para proveniência pode ser aplicado a diferentes domínios porque se baseia em classes que representam conceitos unificadores, de uso comum, idealmente não ambíguos, extraídos de padrões amplamente reconhecidos. A metaproveniência do modelo fundamentada pelo contexto para proveniência (capítulo 2) potencializa a sua interoperabilidade porque assegura a escolha de invariantes presentes ou alinhadas diretamente a classes de outros modelos. Com o objetivo de registrar o que aconteceu, o modelo habilita, através de suas expansões, a capacidade de rastreabilidade dos dados através dos relacionamentos entre evento e oferece, em sua versão minimal, a representação da proveniência. Adicionalmente, serve de base para um entendimento comum de como criar: uma visão para proveniência a partir de modelos de referência ou de outras fontes não ambíguas e uma metodologia de construção de modelos conceituais. Apresentamos as classes e propriedades do modelo mínimo (seção 3.5.1) e as expansões do modelo (seção 3.5.2).

3.5.1. Modelo Mínimo (3P)

Nesta seção, descrevemos o modelo conceitual genérico para proveniência, ou simplesmente o modelo mínimo, que esperamos poder evoluir para um padrão de projeto para proveniência – modelo conceitual 3p – se aplicado e validado para diversos outros domínios diferentes daqueles que avaliaremos como nossos casos concretos (capítulo 4).

Para as classes do modelo mínimo e suas expansões adotamos a nomenclatura padrão para identificação das classes das fontes de construção do modelo e, para as classes sem um identificador único além do seu nome, adotamos uma identificação própria do modelo de proveniência:

- As classes do modelo CIDOC CRM são identificadas pela letra “E” seguida de um número para o id da classe.
- As classes que não tem id na fonte de importação são identificadas por CID (*Class ID*) seguido de um número (CID1, CID2 e assim por diante).

Analogamente, as propriedades são identificadas da seguinte forma:

- As propriedades do modelo CIDOC CRM são identificadas pela letra “P”, seguida de um número.
- As outras propriedades são identificadas por PID seguido de um número.

Com exceção das classes *Reason* e *Means* importadas da Matriz de Perguntas de Wiederhold (1993) e, da classe *Participant* da ontologia de topo de Sowa (1999), as demais classes do modelo de proveniência (mínimo) são importadas da ISO 21127:2006 (seção 2.4.2.2).

Esta dissertação apresenta os termos em inglês em formato *itálico*. A partir de agora, como uma exceção a esta regra, ao referenciarmos uma classe, precederemos seu nome em formato regular – não itálico – sempre com seu respectivo identificador único. Com isso, ratificamos o formato das fontes de importação. As referências ISO 21127 (2006) e Doerr et. al. (2007) contém a descrição detalhada das classes e especializações não explicitadas no modelo mínimo e na expansão FRBR; estas referências de fato apresentam de forma completa as classes da ISO 21127:2006 e do modelo FRBRoo.

Definimos a classe topo da hierarquia de especialização do modelo de proveniência como *Provenance Entity* e identificador CID1. Assim, todas as

classes do modelo são especializações diretas ou indiretas desta classe. Ela possui as seguintes propriedades:

- Identificação por nome ou *appellation* (E41 Appellation), ou identificador único;
- Classificação por tipo (E55 Type);
- Descrição por texto livre.

A classe E52 Time-Span é um intervalo temporal que não se referencia a nenhum contexto geográfico ou cultural, diferentemente de instâncias de E5 Event que ocorrem (*take place at*) em instâncias de E53 Place ou, (*take place on or within*) em instâncias de E18 Physical Thing. Os eventos (E5 Event) têm duração definida por E52 Time-Span. As instâncias de E52 Time-Span algumas vezes são identificadas por instâncias de E49 Time Appellation, usualmente na forma de sua especialização de E50 Date. Outras vezes, as instâncias de E52 Time-Span podem ser discretas no tempo quando “P81 ongoing throughout” opera, ou incertas quando “P82 at some time within” opera. A classe-imagem dessas propriedades é a classe E61 Time Primitive que compreende instâncias de E59 Primitive Value. Esta última representa o tempo e, deve ser implementado com a apropriada validação, precisão e intervalo lógico para expressar variações de datas relevantes para o domínio de aplicação que adota o modelo.

PUC-Rio - Certificação Digital Nº 0611888/CA

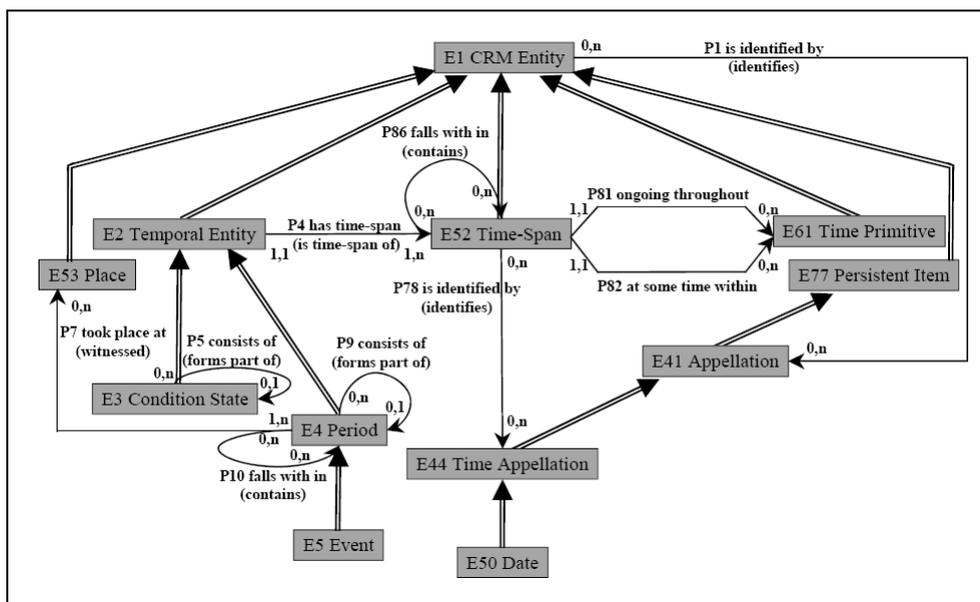


Figura 34: A classe E52 Time-Span, outras classes e relacionamentos da ISO 21127:2006

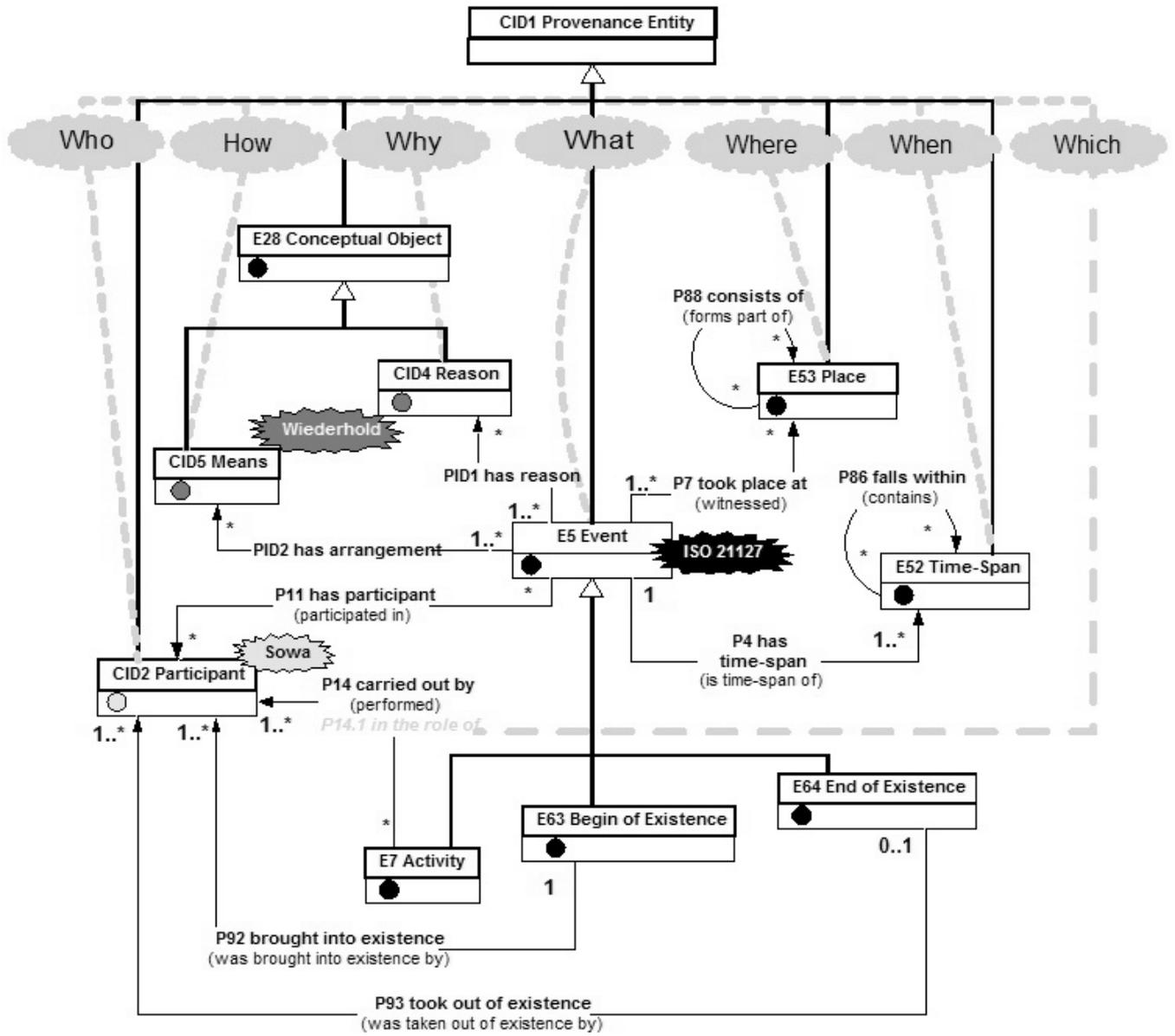
Acrescentamos que, para manter a compatibilidade com a ISO 21127:2006, importamos também a classe E59 Primitive Value, que é superclasse de E60 Number, E61 Time Primitive e E62 String, que representam respectivamente valores primitivos. Por isso, suas instâncias não são consideradas como elementos do domínio de discurso. A **Erro! Fonte de referência não encontrada.** original do modelo CIDOC CRM ilustra algumas dessas classes e relacionamentos (as linhas duplas significam tipo de, e as linhas simples são outros relacionamentos que têm seu significado descrito pelo nome da propriedade).

O modelo CIDOC CRM não se limita apenas a representação de eventos discretos, uma vez que, no domínio de museologia, podem existir objetos que tenham ou não o respectivo dado de quando foram criados que pode ser preciso ou aproximado. De forma análoga, por exemplo, um incêndio ou um furto respectivamente poderiam determinar a sua destruição ou mera incerteza sobre sua existência. No modelo conceitual de proveniência nos interessam apenas os eventos discretos, portanto, os que têm duração ou são identificados por uma determinada data.

Deixamos como nota que a opção de construção do modelo de proveniência quebrou a hierarquia original proposta pela ISO 21127:2006 para acomodar a noção de Abstrato como sendo disjunta da noção de Concreto, diferentemente da proposta na ISO 21127:2006 que acomoda ambas como um item persistente. Em nossa interpretação, as classes E28 Conceptual Object e CID2 Participant são disjuntas e, por isso, são classes primárias.

Em cada uma das próximas subseções descreveremos as classes primárias e suas respectivas especializações, adotando a nomenclatura da ISO 21127:2006. Apenas as classes presentes na Tabela 18 serão detalhadas. Para cada propriedade de cada classe, apresentamos seu identificador único, sua descrição, propriedade inversa entre parêntesis e a sua classe-imagem. A maioria dos exemplos é inspirada em exemplos da ISO 21127:2006. Para informações adicionais sobre outras especializações e propriedades veja ISO 21127 (2006) e Doerr et al. (2007).

Por fim, apresentamos na Figura 35 o diagrama UML referente ao modelo mínimo (seção 3.5.1) com objetivo de mantermos o foco nas principais classes e relacionamentos. Identificamos também o mapeamento para os conceitos abstratos que foram utilizados durante a construção do modelo, para facilitar a compreensão do texto daqui em diante.



PUC-Rio - Certificação Digital Nº 0611888/CA

Figura 35: Diagrama UML com principais classes e relacionamentos do modelo conceitual e ilustrado com conceitos abstratos (*Wh-questions*)

3.5.1.1. Classes

Classe importada da ISO 21127:2006 - E5 Event (*What*)

Id e nome da classe primária: E5 Event

Subclasse de: CID1 Provenance Entity

Superclasse de: E7 Activity, E63 Begin of Existence e E64 End of Existence

Nota de Escopo: esta classe compreende mudanças de estados independentemente de escala promovidos por arranjos de fenômenos culturais, físicos, legais ou tecnológicos. Tais mudanças afetam instâncias de CID2 Participant ou suas especializações. Um evento visto seu nível mais granular é uma mudança “instantânea” de estado. E5 Event pode ser analisado através de seus fenômenos componentes com espaço e intervalo de tempo.

Exemplos: a 2ª Guerra Mundial ou A Conferência de Yalta em 1945.

Propriedades importadas:

- P4 has time-span (is time-span of): E52 Time-Span
- P7 took place at (witnessed): E53 Place
- P11 has participant (participates in)⁴²: CID2 Participant
- PID1 has reason: CID4 Reason
- PID2 has arrangement: CID5 Means

Propriedades não importadas:

- P8 took place on or within (witnessed): E19 Physical Object
- P12 occurred in the presence of (was present at): CID2 Participant
- P114 is equal in time to: E5 Event
- P115 finishes (is finished by): E5 Event
- P116 starts (is started by): E5 Event
- P117 occurs during (includes): E5 Event
- P118 overlaps in time with (is overlapped in time by): E5 Event

Descrevemos os detalhes das especializações diretas importadas da ISO 21127:2006 da classe E5 Event. Porém, as subclasses destas classes não serão

⁴² A propriedade P14.1 original é had participant (participates in)

detalhadas aqui. Para maiores informações veja ISO 21127 (2006) e Doerr et al. (2007).

Id e nome: E7 Activity

Nota de Escopo: define ações intencionais desempenhadas por participantes (instâncias) CID2 Participant, que ocasionam mudanças de estado. Esta noção inclui ações complexas ou de longa duração.

Exemplos: a criação da parte 1 da tragédia “Fausto” por Goethe em 1806 ou a formação da Escola Bauhaus em 1919.

Propriedades importadas:

- P14 carried out by (performed): CID2 Participant
 - P14.1 in the role of: E55 Type

Propriedades não importadas:

- P15 was influenced by (influenced): E1 Provenance Entity
- P17 was motivated by (motivated): E1 Provenance Entity
- P19 was intended use of (was made for): E1 Provenance Entity
- P20 had specific purpose (was purpose of): E7 Activity
- P21 had general purpose (was purpose of): E55 Type
- P32 used general technique (was technique of): E55 Type
- P125 used object of type (was type of object used in): E55 Type

Destacamos duas especializações de E5 Event que simbolizam eventos marcantes de uma história (*state of affairs*).

Id e nome: E63 Begin of Existence

Nota de Escopo: representa eventos que simbolizam o início da existência de qualquer instância de CID2 Participant nascimento, criação, formação entre outros.

Exemplo: a criação da Torre Eiffel para a exposição universal em 1889 ou a re-criação do Panteão de Agripa (Roma) em 125 d.C.

Propriedades:

- P92 brought into existence (was brought into existence by): CID2 Participant

Id e nome: E64 End of Existence

Nota de Escopo: representa eventos que definem o fim da existência de qualquer instância de CID2 Participant morte, destruição, dissolução entre outros.

Exemplo: a destruição do Panteão de Agripa (Roma) por um incêndio em 80 d.C. ou a morte de Alexandre O Grande na Babilônia em 323 a.C.

Propriedades:

- P93 took out of existence (was taken out of existence by): CID2 Participant

Classe importada de Sowa (1999) - CID2 Participant (*Who*)

Id e nome da classe primária: CID2 Participant

Subclasse de: CID1 Provenance Entity

Superclasse de: nenhuma

Escopo: identifica um participante do evento. O participante pode determinar ou não o que acontece, pode estar presente apenas no início ou no fim sem exercer nenhum tipo de controle no respectivo evento do qual participa.

Exemplos: uma jaca que amassa o teto de um carro ou uma pessoa que observa cair, de outra forma: um ladrão que quebra o vidro de um carro ou uma câmera que o filma.

Propriedades: nenhuma

Descrevemos os detalhes das especializações diretas da classe E5 Event. Porém, as subclasses destas classes não serão detalhadas aqui. A classe E51 Contact Point não foi importada para o modelo de proveniência mínimo, mas pode ser alvo de futuras expansões. Para maiores informações veja ISO 21127 (2006) e Doerr et al. (2007).

As classes E39 Actor e E18 Physical Thing da ISO 21127:2006 não foram importadas para o modelo mínimo porque consideramos que estão cobertas pela generalização proposta CID2 Participant. Sugerimos ao projetista considerá-las como possíveis especializações diretas ou indiretas de CID2 Participant para o domínio de aplicação que esteja em construção. Destacamos que a classe E39 Actor possui uma propriedade importada de (Pasin et al., 2007) que permite a representação de conceitos (E28 Conceptual Object) concebidos.

Id e nome: E39 Actor

Nota de Escopo: compreende pessoas, indivíduos ou grupos, que tem o potencial de atuar intencionalmente em ou observar ações, nas quais podem assumir responsabilidades.

Exemplo: Marco Antonio Casanova (*person*) ou Representante de Alunos de Pós-Graduação (*legal body*) ou Comissão de Pós-Graduação (*group*)

Propriedades:

- P74 has current or former residence (is current or former residence of): E53 Place
- P75 possesses (is possessed by): E30 Right
- P76 has contact point (provides access to): E51 Contact Point
- P131 is identified by (identifies): E82 Actor Appellation
- PID5 conceives (is conceived by): E28 Conceptual Object

Sugerimos que a expansão permita também representar que um participante intencional (E39 Actor) tenha concebido um conceito através da propriedade “PID5 conceives (is conveyed by)”.

Id e nome: E18 Physical Thing

Nota de Escopo: inclui todos os itens físicos, com relativa forma estável, naturais ou artificiais.

Exemplo: A Mona Lisa ou Meu Notebook

Propriedades:

- P45 consists of (is incorporated in): E57 Material
- P46 is composed of (forms part of): E18 Physical Thing
- P49 has former or current keeper (is former or current keeper of): E39 Actor
- P50 has current keeper (is current keeper of): E39 Actor
- P51 has former or current owner (is former or current owner of): E39 Actor
- P52 has current owner (is current owner of): E39 Actor
- P53 has former or current location (is former or current location of): E53 Place
- P58 has section definition (defines section): E46 Section Definition
- P59 has section (is located on or within): E53 Place

Classe importada da ISO 21127:2006 - E53 Place (*Where*)

Id e nome da classe primária: E53 Place

Subclasse de: CID1 Provenance Entity

Superclasse de: nenhuma outra classe

Nota de Escopo: define a extensão no espaço no sentido físico: independente de fenômeno temporal (*temporal phenomena*) ou matéria (*matter*). Instâncias de E53 Place são usualmente determinadas pela referência a uma posição “imóvel” (prédio, cidade, montanha, rios, marcos geográficos). Um lugar (*Place*) pode ser determinado pela posição relativa a uma estrutura (*frame*) através da noção de “seção” (*section*), que pode assumir uma instância de E19 Physical Object como válida para determinação de E53 Place. Pode ainda ser identificada por uma ou mais instâncias da classe E44 Place Appellation, subclasse de E41 Appellation. Por fim, as instâncias de E53 Place podem ser também identificadas por coordenadas globais ou referências absolutas.

Exemplos: Coordenada geográfica do meridiano de Greenwich ou endereço do Observatório de Greenwich em Londres.

Propriedades importada e exibida⁴³ nas representações gráficas do modelo mínimo:

- P88 consists of (forms part of): E53 Place

Propriedades importadas e não detalhadas⁴⁴ nas representações do modelo mínimo:

- P87 is identified by (identifies): E44 Place Appellation
- P89 falls within (contains): E53 Place
- P121 overlaps with: E53 Place
- P122 borders with: E53 Place

Classe importada da ISO 21127:2006 - E52 Time-Span (*When*)

Id e nome da classe primária: E52 Time-Span

Subclasse de: CID1 Provenance Entity

Superclasse de: nenhuma outra classe

Nota de Escopo: compreende extensões temporais abstratas que tenham início, fim e duração. Instâncias da classe E52 Time-Span não tem nenhuma

⁴³ Com o objetivo de exemplificar uma propriedade que é um auto-relacionamento não qualificado.

⁴⁴ Com o propósito de manter o foco da discussão sobre as propriedades mais relevantes.

outra conotação semântica a não ser a de definir uma extensão temporal para instâncias de E5 Event válidas durante certo intervalo. Um E52 Time-Span pode ser identificado com uma ou mais instâncias de E49 Time Appellation. Uma instância de E52 Time-Span durante a qual uma biblioteca foi construída pode recair em outro intervalo que corresponda ao mandato de um político, embora, muito provavelmente, não exista uma relação contextual ou causal entre ambas.

Exemplos: A data da fixação ao teto do Panthéon e da demonstração do pêndulo de Foucault em Paris no ano de 1851, ou o intervalo que caracteriza a física de Galileu.

Propriedades importada e exibida⁴⁵ nas representações gráficas do modelo mínimo:

- P86 falls within (contains): E52 Time-Span

Propriedades importadas e não detalhadas⁴⁶ nas representações do modelo mínimo:

- P78 is identified by (identifies): E49 Time Appellation
- P79 beginning is qualified by: E62 String
- P80 end is qualified by: E62 String
- P81 ongoing throughout: E61 Time Primitive
- P82 at some time within: E61 Time Primitive

Classe importada da ISO 21127:2006 - E28 Conceptual Object (*Which, How e Why*)

Id e nome da classe primária: E28 Conceptual Object

Subclasse de: CID1 Provenance Entity

Superclasse de: CID4 Reason, CID5 Means e E55 Type

Nota de Escopo: compreende produtos não materiais do intelecto e informações produzidas por seres humanos com ou sem uso de aparatos tecnológicos que são objetos de discurso sobre identidade, circunstâncias de criação e implicações históricas. Caracteristicamente, instâncias desta classe são criadas, inventadas ou pensadas por algum E39 Actor ou classes que representem objetos de discurso que possuam racionalidade.

⁴⁵ Com o objetivo de exemplificar uma propriedade que é um auto-relacionamento não qualificado.

⁴⁶ Com o propósito de manter o foco da discussão sobre as propriedades mais relevantes.

Exemplos: a definição de *World Wide Web* de Tim Berners-Lee ou a obra de Michelangelo.

Propriedades: nenhuma

Agora, descrevemos os detalhes para as especializações diretas da classe E28 Conceptual Object.

Id e nome: CID4 Reason (*Why*)

Nota de Escopo: é qualquer motivação que descreva “o Porquê” (*Why*). Esta classe deve ser refinada pelo projetista de acordo com o domínio de aplicação.

Exemplo: a rede da informática parou por mais de 24h porque houve uma catástrofe na sala dos equipamentos de rede.

Propriedades: nenhuma

Id e nome: CID5 Means (*How*)

Nota de Escopo: é qualquer explicação que descreva “o Como” (*How*). Esta classe deve ser refinada pelo projetista de acordo com o domínio de aplicação. É necessária atenção para não confundir esta classe com suas variações, por exemplo, *how come* tem o mesmo significado que *why*.

Exemplo: uma sequência de eventos, um arranjo de objetos no espaço, um encadeamento de idéias, um *workflow* ou um processo.

Propriedades: nenhuma

A classe E55 Type não é apresentada nas representações gráficas (**Erro! Fonte de referência não encontrada.** e **Erro! Fonte de referência não encontrada.**) do modelo mínimo nem nas hierarquias de especializações (Tabela 17 e Tabela 18), propositalmente, com o único objetivo de focarmos a discussão nas classes mais relevantes.

Id e nome: E55 Type (*Which*)

Nota de Escopo: compreende conceitos arbitrários (universais) e prove um mecanismo para organizá-los em uma hierarquia.

Exemplo: terremoto, maremoto e furacão são tipos de catástrofe natural.

Propriedades:

- P127 has broader term (has narrower term): E55 Type
- P137 is exemplified by (exemplifies): CID1 Provenance Entity

3.5.1.2. Propriedades

Apresentamos na **Erro! Auto-referência de indicador não válida.** para cada classe descrita (seção 3.5.1.1), o identificador da propriedade, a descrição da propriedade e sua inversa entre parêntesis, a classe-domínio, a classe-imagem e a multiplicidade da propriedade (coluna Card.) e de sua inversa para as principais propriedades do modelo de proveniência. Para a lista completa de propriedades veja ISO 21127 (2006) e Doerr et al. (2007).

Tabela 19: Descrição das Propriedades do Modelo de Proveniência

Id	Descrição da propriedade	Classe-domínio	Classe-Imagem	Mult.
P7	took place at (witnessed)	E5 Event	E53 Place	0,n (1,n)
P11	has participant (participates in)	E5 Event	CID2 Participant	0,n (0,n)
P4	has time-span (is time-span of)	E5 Event	E52 Time-Span	1,n (1)
PID1	has reason	E5 Event	CID4 Reason	0,n
P14	carried out by (performed)	E7 Activity	CID2 Participant	1,n (0,n)
PID2	has arrangement	E5 Event	CID5 Means	0,n
P92	brought into existence (was brought into existence by)	E63 Begin of Existence	CID2 Participant	1,n (1)
P93	took out of existence (was taken out of existence by)	E64 End of Existence	CID2 Participant	(1,n) 0,1
P88	consists of (forms part of)	E53 Place	E53 Place	0,n (0,n)
P86	falls within (contains)	E52 Time-Span	E52 Time-Span	0,n (0,n)

3.5.2. Expansões

O projetista que expande o modelo mínimo deve se certificar de que a disjunção é preservada. Adicionalmente, recomendamos ao projetista que sempre consulte as fontes de importação – a ontologia de topo de Sowa (1999), a Matriz de Perguntas de Wiederhold (1993), a ontologia DOLCE, o modelo FRBRoo e a ISO 21127:2006 – durante o planejamento de expansões. Com isso, há maiores chances de potencializar a interoperabilidade do modelo estendido.

Sugerimos três expansões para o modelo mínimo. A primeira é a expansão de ordem parcial (PO) (seção 3.5.2.1). A segunda incorpora ao modelo mínimo a capacidade de rastreabilidade dos dados (seção 3.5.2.2). Por fim, sugerimos a expansão para herança cultural (seção 3.5.2.3) para domínios que precisem representar a noção de obra – F1 Work – e conceitos relacionados (seção 2.4.3.1) presentes no modelo FRBR.

3.5.2.1. PO (*Partial Order*)

Sugerimos a noção matemática de ordem parcial (*partial order*), abreviadamente PO, como uma extensão ao modelo mínimo. Uma ordem (total) é um caso particular de uma ordem parcial. Para os casos concretos (capítulo 4) expandiremos o modelo mínimo com a noção de ordem parcial. No modelo conceitual genérico de proveniência estamos interessados em ordens parciais de entidades temporais (E5 Event) porque nosso foco é a proveniência como um conjunto de descrições de qualquer mudança de um objeto de informação que registre sua história (eventos) e que seja significativa para resguardar autenticidade, integridade e interpretação. Com isso, especializamos CID5 Means com a classe CID6 PO Event que descreve ordens parciais de eventos.

Id e nome: CID6 PO Event

Subclasse de: CID5 Means

Nota de Escopo: representa uma ordem parcial de eventos entendida como uma explicação, ou seja, a representação sugerida por esta expansão, como resposta a pergunta Como (*How*).

Exemplo: uma sequência de eventos ou um projeto

Propriedades: PID4 arranges: E5 Event

A expansão PO permite também propor outras evoluções como a representação de ordens parciais de participantes (CID2 Participant) ou objetos conceituais (E28 Conceptual Object), respectivamente, por exemplo, um layout dos móveis em uma sala e uma ontologia de domínio. Neste caso, especializaríamos CID5 Means em CID8 PO Participant e CID9 PO Conceptual Object. Acrescentamos ainda, que a classe *Course of Events* – padrão D&S da DOLCE – poderia ser uma especialização (natural) futura da classe CID6 PO Event. Analogamente, sugerimos que a classe E29 Design or Procedure do modelo CIDOC CRM poderia ser uma especialização de CID5 Means.

Consultando a Figura 36 (cópia da Figura 1 da seção 1.1), identificamos ao topo: Fornecedor, Produtor, Transformador e Distribuidor que representam etapas da produção – arranjos de atividades e lotes - que podemos considerar como ordens parciais que descrevem “o Como (*How*)” da respectiva etapa.

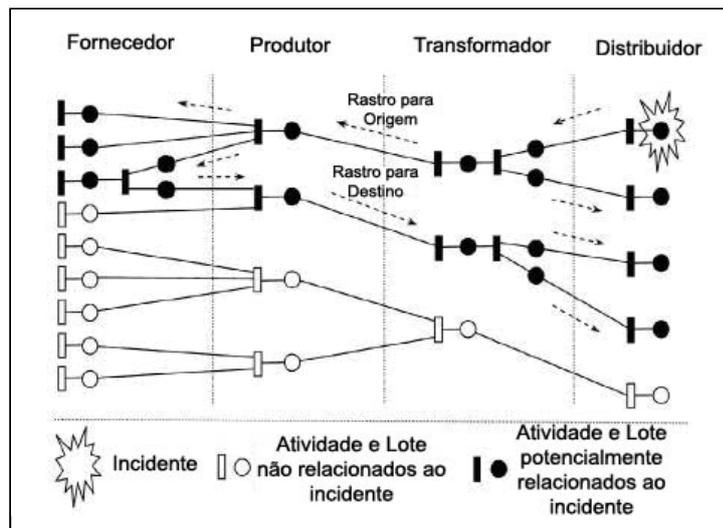


Figura 36: Rastreabilidade na Cadeia Produtiva baseado em (Bechini et al., 2005)

Acrescentamos que a noção de projeto pode ser representada como uma ordem parcial de eventos (E5 Event) ou uma ordem parcial de atividades (E7 Activity) com duração e relacionamentos causais (CID4 Reason) entre si. A noção ordem parcial de atividades com apenas um relacionamento dedutível -

ordem cronológica dos eventos a partir de respectivos *timestamps* - pode ser encontrada no Projeto *DrProject*⁴⁷ inspirado na ferramenta Trac (seção 4.5.3). Também é possível encontrar a noção de ordem parcial em cronogramas construídos a partir da ferramenta *Microsoft Project*⁴⁸. Um projeto pode ser interpretado como um DAG de atividades (E7 Activity).

Sowa (2001a) argumenta que os processos governados por leis (*law-governed process*) estão em algum ponto intermediário entre o caos e o determinismo absoluto. Por exemplo, a lei da gravidade impõe restrições suficientes para sabermos onde a água cai da ducha do banheiro, logo onde devemos ficar em pé para tomarmos banho, mas não é suficiente para prever a trajetória exata de cada gota de água (Sowa, 2001a). Com base nessa realidade, Sowa define que:

- Um evento é uma especialização de um processo discreto (tem início e fim conhecidos);
- Um processo discreto é um grafo acíclico direcionado cujos nós podem ser eventos ou estados. Portanto, um processo discreto consiste em instâncias de eventos ou estados;
- Um procedimento (*procedure*) é uma família – tipos de eventos ou tipos de estados – de processos discretos que podem ter ciclos. Apesar de permitidos, estes ciclos não promovem um retorno no tempo, mas apenas a alguma instância do mesmo tipo de evento ou tipo de estado.

A Figura 37 (Sowa, 2001a) ilustra três exemplos de notações gráficas diferentes para a representação de procedimentos: um fluxo (*Flow Chart*), onde os eventos são simbolizados pelos quadrados e as decisões por diamantes; uma máquina de estados finita (*Finite-State Machine*), onde os círculos simbolizam os estados e os arcos as transições entre estados; uma Rede Petri (*Petri Net*) – considerada uma fusão dos dois últimos procedimentos – onde os círculos são estados (*places*) e as barras entre arcos são eventos (*transitions*). (Sowa, 2001a)

⁴⁷ <https://www.drproject.org/>

⁴⁸ <http://www.microsoft.com/brasil/office/project/>

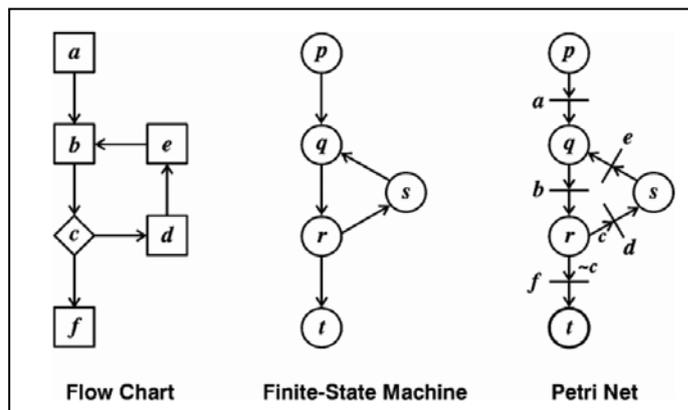


Figura 37: Três notações gráficas para especificação de procedimentos (Sowa, 2001a)

Com isso, podemos concluir que a noção de ordem parcial que adotamos para a classe CID6 PO Event contempla a representação de processos discretos onde os nós são eventos. Analogamente, o modelo mínimo com a expansão PO cobre a noção de projeto, uma vez que uma atividade (E7 Activity) é uma especialização de um evento (E5 Event).

3.5.2.2.

TR (*TR*aceability – *TR*ace and *TR*ack)

Rastreabilidade é a capacidade de realizar consultas para rastrear a origem (*trace*) e rastrear o destino (*track*) (Dorp, 2004). Rastrear dados é realizar consultas a proveniência (seção 1.1). O modelo mínimo não restringe como o projetista deve planejar a rastreabilidade e oferece apenas a propriedade, e não oferece sua inversa, que liga um evento (E5 Event) a uma motivação (CID4 Reason) ou analogamente, um evento a uma explicação (CID5 Means).

3.5.2.2.1.

RQR (*Reification of Qualified Relations*)

O primeiro relacionamento (propriedade) que nos interessa explorar é o auto-relacionamento qualificado entre eventos. Para criar essa noção nesta expansão, especializamos a classe CID4 Reason em CID7 Event Reason, que representa a reificação do relacionamento e que tem a função de atribuir uma justificativa, intenção, motivo ou causa à ligação entre eventos. Por isso, possibilita a captura de relacionamentos que não podem ser deduzidos, bem como a representação de eventos complexos. Comentamos que um

relacionamento dedutível seria, por exemplo, uma linha do tempo a partir dos *timestamps* dos eventos.

Id e nome: CID7 Event Reason

Subclasse de: CID4 Reason

Nota de Escopo: o auto-relacionamento entre eventos é explicitado através de atributos ou propriedades que representam respostas a pergunta Por que (*Why*), portanto, atribuindo mais significado ao dado (E5 Event). A propriedade “PID3 is reason to” é a inversa, não representada no modelo mínimo, à propriedade “PID1 has reason” que tem classe-domínio E5 Event e classe-imagem CID4 Reason.

Exemplo: a pintura do Juízo Final na Capela Sistina **sobrepôs** (CID7) a pintura de uma imagem da Virgem da Assunção de Perugin; o evento de assinatura do acordo de Crimeia **é parte do** (CID7) evento da Conferência de Yalta.

Propriedades: PID3 is reason to: E5 Event

A expansão do modelo mínimo a partir da adição de outros auto-relacionamentos – entre instâncias de CID2 Participant ou entre instâncias de E28 Conceptual Object – possibilitaria a representação de instâncias complexas. Por exemplo, no primeiro caso, poderíamos capturar a noção de que o filé *mignon* **é parte de** um boi, ou cada um de três diferentes quadros **é parte de** um tríptico. No segundo caso, seria possível imaginar que um campo de estudo **é comum a** uma área de pesquisa ou uma teoria e uma escola de pensamentos **são partes de** uma ciência.

Optamos por não representar no modelo mínimo tais auto-relacionamentos, delegando tal decisão para o projetista do domínio. Com isso, o modelo conceitual genérico para proveniência ratifica sua resiliência, podendo oferecer uma opção de recuperação do dado flexível à demanda do domínio que instancia o modelo.

Concluimos, retomando o exemplo do boi, que seria possível então a partir de uma consulta ao evento de abate do boi se chegar – através do auto-relacionamento entre CID2 Participant – até o filé *mignon*. Adicionalmente, uma consulta a eventos que se referem ao mesmo boi permitiria validar se o filé *mignon* é proveniente de um boi que foi vacinado contra a doença da vaca-louca. Este seria um dado que não está representado no banco de dados, porém pode

ser obtido através de uma consulta, fixando uma das facetas da proveniência: CID2 Participant (*Who*).

3.5.2.2.2. DISP (*Determinant, Immanent, Source Product*)

Adicionalmente ao auto-relacionamento, o rastro para origem ou para o destino de cada atividade (E7 Activity) – especializações de eventos (E5 Event) – pode interessar ao projetista que tem como objetivo a identificação de quais são as respectivas entradas (*Sources*) e saídas (*Products*). Além disso, o domínio onde será aplicada a expansão do modelo mínimo pode necessitar a representação distinta de participantes: os que controlam o curso de um evento (*Determinant*) e os que são expectadores meramente aderentes (*Immanent*) a ele.

A expansão DISP pode ser facilmente alcançada a partir da importação das classes *Determinant, Immanent, Source e Product* da ontologia de topo de Sowa (2001b). Estes conceitos dão origem a outras classes especializadas que Sowa (2001b) define como papéis temáticos (*Thematic Roles*). É interessante destacar aqui que na ontologia de topo de Sowa a classe *Role* está a muitos níveis acima da classe *Participant*, portanto, um participante é então uma especialização indireta de um papel. Acrescentamos ainda que o conceito de agente também aparece na hierarquia de especialização proposta por Sowa e pode ser generalizado para o conceito de participante.

Id e nome: CID10 Determinant

Subclasse de: CID2 Participant

Nota de Escopo: um determinante participa da direção de um processo, seja para iniciá-lo, alterar o seu fluxo ou encerrá-lo. (Sowa, 2001b)

Exemplo: um programador Java que codifica um componente

Propriedades: nenhuma

Id e nome: CID11 Immanent

Subclasse de: CID2 Participant

Nota de Escopo: um imanente (aderente) é um participante que está presente durante todo o processo, mas não controla ativamente o que acontece. (Sowa, 2001b)

Exemplo: a IDE (*Integrated Development Environment*) MyEclipse⁴⁹

Propriedades: nenhuma

Id e nome: CID12 Source

Subclasse de: CID2 Participant

Nota de Escopo: uma fonte (*Source*) deve estar presente no início do processo, mas não precisa estar presente durante o processo (Sowa, 2001b)

Exemplo: um documento de especificação de requisitos funcionais

Propriedades: nenhuma

Id e nome: CID13 Product

Subclasse de: CID2 Participant

Nota de Escopo: um produto (*Product*) deve estar presente ao final do processo, mas não precisa estar presente durante o processo (Sowa, 2001b)

Exemplo: a versão *release candidate* de um componente de software

Propriedades: nenhuma

Por fim, sugerimos a seguir a classe CID3 Functional Role como uma possível opção para a representação de papéis. Essa classe seria a reificação da propriedade “P14.1 in the role of” da ISO 21127:2006 que tem classe-imagem E55 Type.

Id e nome: CID3 Functional Role (*Which*)

Subclasse de: E28 Conceptual Object

Nota de Escopo: compreende responsabilidade causal ou legal quando definido para instâncias de CID2 Participant e restrições de papéis a serem desempenhados em atividades E7 Activity.

Exemplo: o papel de incendiário voador é desempenhado por um dragão (CID2), ou apenas um incendiário voador pode desempenhar a atividade de cuspir fogo (E7).

Propriedades:

- PID6 performs (carried out by): E7 Activity
- PID7 played by (plays): CID2 Participant

Neste caso, a classe tem dois pares de relacionamentos (com inversa) que a ligariam à classe CID2 Participant (ou alguma de suas especializações) e à classe E7 Activity. O papel poderia ser descrito de forma geral por instâncias de E55 Type ou mais restrita a partir de uma enumeração.

3.5.2.3.**CH (Cultural Heritage)**

Para as classes desta expansão adotamos a nomenclatura padrão para identificação das classes das fontes de construção, analogamente como fizemos na descrição do modelo mínimo. Acrescentamos que as classes do modelo FRBRoo, se importadas em expansões do modelo, devem ser identificadas pela letra “F” seguida de um número para o id da classe. Da mesma forma, as propriedades do modelo FRBRoo são identificadas pela letra “R” seguida de um número. Descrevemos então as classes desta expansão:

Id e nome: F1 Work

Subclasse de: E28 Conceptual Object

Nota de Escopo: inclui a soma de conceitos que afloram no curso de evolução de uma idéia original. Pode ser encontrado através de uma ou mais expressões que são dominadas pela mesma idéia. Uma obra (*Work*) pode ser produzida por mais de um E39 Actor simultaneamente ou através do tempo. Uma obra pode ter componentes que constituem o conceito geral. Um volume de uma trilogia representa apenas uma parte do conceito geral.

Exemplo: “Hamlet” de William Shakespeare ou “David” de Gian Lorenzo Bernini ou o padrão D&S da DOLCE.

Propriedades:

- R1 has constraining supertype (is constraining supertype of): E55 Type

⁴⁹ <http://www.myeclipseide.com>

- R2 has representative expression (is representative expression for): F2 Expression
- R57 is logical successor of (has successor): F1 Work
- R58 is derivative of (has derivative): F1 Work
- R65 is realised in (realises): F2 Expression

Id e nome: E73 Information Object

Subclasse de: E28 Conceptual Object

Superclasse de: F3 Expression

Nota de Escopo: inclui itens imateriais identificáveis como poemas, piadas, imagens, algoritmo, entre outros, que tenha uma estrutura reconhecível objetivamente. As instâncias de E73 Information Object não dependem de uma instância de E84 Information Carrier. Instâncias conceituais como tipos ou classes não são instâncias de E73 Information Object, bem como idéias sem uma forma de expressão reproduzível.

Exemplo: imagem do Pão de Açúcar ou o filme “A Cidade Perdida” de Andy Garcia.

Propriedades:

- P67 refers to (is referred to by): CID1 Provenance Entity
- P106 is composed of (forms part of): E73 Information Object
- P129 is about (is subject of): CID1 Provenance Entity

Id e nome: F3 Manifestation Product Type

Subclasse de: E55 Type

Nota de Escopo: inclui as definições de produtos de publicações. Uma instância de F3 Manifestation Product Type é a espécie da publicação (jornal, revista etc)

Exemplo: o formato ASCII ou binário do padrão D&S da DOLCE codificado na linguagem Java ou jornal, periódico, revista.

Propriedades:

- P2 should have type (should be type of): E55 Type
- P46 should be composed of (may form part of): F3 Manifestation Product Type
- P57 should have number of parts: E60 Number
- P104 subject to (applies to): E30 Right
- P105 right held by (right on): E39 Actor

Agora, descreveremos os detalhes para as especializações diretas da classe E73 Information Object. Essa classe possui outras especializações que não detalharemos aqui, porque estamos focando a atenção nas classes que são relevantes para nossa discussão. Para maiores informações veja ISO 21127 (2006) e Doerr et al. (2007).

Id e nome: F2 Expression

Subclasse de: E73 Information Object

Nota de Escopo: compreende a realização de trabalhos artísticos ou intelectuais de uma obra, na forma de itens imateriais identificáveis. Expressão é a linguagem na qual a obra (F1 Work) se expressa. Expressões não existem sem sua respectiva instância de E84 Information Carrier.

Exemplo: o padrão D&S da DOLCE na linguagem Java ou Prolog

Propriedades:

- R3 has representative manifestation product type (is representative manifestation product type for): F3 Manifestation Product Type
- R11 is composed of (forms part of): F2 Expression

As classes a seguir seriam as especializações indiretas de CID2 Participant que poderiam ser importadas para expansões do modelo de forma a mapear conceitos de domínios de biblioteconomia, e-commerce ou outro que demande a representação da noção de veículo de persistência. Sugerimos que ambos os conceitos estejam presentes nesta expansão na hierarquia de especialização descrita porque está alinhada ao resultado da harmonização do modelo FRBRoo com a ISO 21127:2006.

Id e nome: E84 Information Carrier'

Subclasse de: E22 Man-Made Object

Superclasse de: F5 Item

Nota de Escopo: inclui todas as instâncias de E22 Man-Made Object que são explicitamente projetadas para atuar como veículo de persistência física para instâncias da classe E73 Information Object. Isto permite que relacionamentos sejam estabelecidos entre E19 Physical Object e um respectivo conteúdo imaterial. Um E84 Information Carrier pode ou não conter informação. Algumas classes não descritas aqui, porém podem ser encontradas no ISO 21127 (2006) e Doerr et al. (2007).

Exemplo: A pedra Rosetta (*Rosetta Stone*)

Propriedades: nenhuma

Id e nome: F5 Item

Subclasse de: E84 Information Carrier

Nota de Escopo: compreende objetos físicos que carregam uma expressão de publicação produzida por algum processo industrial que envolve uma instância de F3 Manifestation Product Type.

Exemplo: o item impresso desta dissertação ou um DVD com todo o material pesquisado.

Propriedades:

- R10 is example of (has example): F3 Manifestation Product Type

Apresentamos nas Tabela 20 e Tabela 21 algumas propriedades das classes desta expansão. Para maiores informações veja ISO 21127 (2006) e Doerr et al. (2007).

Tabela 20: Descrição das Propriedades do Modelo de Proveniência (parte 1 de 2)

Id	Descrição da propriedade (Descrição da propriedade inversa)	Classe-domínio	Classe- Imagem
R1	has constraining supertype (is constraining supertype of)	F1 Work	E55 Type
R2	has representative expression (is representative expression for)	F1 Work	F2 Expression
R57	is logical successor of (has successor)	F1 Work	F1 Work
R58	is derivative of (has derivative)	F1 Work	F1 Work
R65	is realised in (realises)	F1 Work	F2 Expression
P2	should have type (should be type of)	F3 Manifestation Product Type	E55 Type
P46	should be composed of (may form part of)	F3 Manifestation Product Type	F3 Manifestatio n Product Type
P57	should have number of parts	F3 Manifestation Product Type	E60 Number
P104	subject to (applies to)	F3 Manifestation Product Type	E30 Right
P105	right held by (right on)	F3 Manifestation Product Type	E39 Actor
P67	refers to (is referred to by)	E73 Information Object	CID1 Provenance Entity
P106	is composed of (forms part of)	E73 Information Object	E73 Information Object

Tabela 21: Descrição das propriedades da expansão (parte 2 de 2)

Id	Descrição da propriedade (Descrição da propriedade inversa)	Classe-domínio	Classe- Imagem
P129	is about (is subject of)	E73 Information Object	CID1 Provenance Entity
R3	has representative manifestation product type (is representative manifestation product type for)	F2 Expression	F3 Manifestatio n Product Type
R11	is composed of (forms part of)	F2 Expression	F2 Expression

Exploraremos (capítulo 4) os casos concretos do modelo conceitual genérico para proveniência aplicando o modelo mínimo com a expansão PO.

3.6. Considerações Finais

A cada novo dia, governo, indústria e a academia contribuem com a introdução de novos modelos que se somam a uma infinidade de outros já construídos. Alguns deles são promovidos a padrões, que se listados aqui tornariam esta introdução longa e difícil. O interessante sobre padrões é que há inúmeros para serem escolhidos, mas poucas vezes cogitamos escolher algum porque acabamos por criar os nossos próprios. Com cuidado, se percebidos com maior profundidade, dentre os padrões existentes, é possível notar que há de fato fragmentos semelhantes, padrões de conteúdo (*content-patterns*), por vezes sintáticos, por outras, semânticos. Ainda assim, com disciplina, é possível destacar fragmentos, identificar e classificar invariantes e, por fim, interromper o ciclo vicioso da criação de padrões a partir de novos conceitos.

A despeito dos esforços de inúmeros grupos (seção 2.1.2) que estudam elementos comuns para representar metadados, é quase certo que, ao longo dos próximos anos, muitas, senão a maioria das comunidades, com suas distintas necessidades, inventarão modelos de metadados incompatíveis. Na busca por soluções, que fundamentalmente apresentam sobreposições de

objetivos. Se existe alguma metodologia geral que oriente o projetista de metadados, essa no mínimo não é adequadamente difundida nem reconhecida amplamente nos diversos domínios de conhecimento existentes.

Bearman et al. (1999) alerta que é de suma importância que modelos conceituais comuns para a troca de informação sejam identificados e adotados. Reconhecer a convergência de requisitos ajuda a minimizar as diferenças muitas vezes conflitantes das demandas em domínios distintos. Acrescenta que educadores já propuseram modelos para objetivos semelhantes aos da comunidade de comércio eletrônico.

Uma abordagem mais disciplinada durante o projeto de metadados poderia ser conquistada considerando o custo-benefício da introdução de novos conceitos. Uma metodologia para construção de modelos ampara-se na avaliação cuidadosa de propriedades, classes-domínio e classes-imagens onde operam, com o objetivo de identificar semelhanças de como os metadados estão relacionados e não apenas uma comparação entre tipologias diferentes. Este caminho não apenas habilita a equivalência semântica, mas também a integração de informações complementares.

Doerr et al. (2003) elucida que os esforços de harmonização entre diferentes modelos focam na análise de classes, propriedades e relacionamentos. Conclui que em alguns casos o nível de abstração de determinados conceitos é tão alto que não há como avaliar sua correteza. Também admite que conceitos como “evento” e “agente” podem apresentar semânticas distintas e, mesmo assim, apropriadas em cada contexto onde são adotados.

A conclusão em geral que se pode obter a partir do estudo de ontologias de alto nível (seção 2.3) é que o mundo é dual. Há possibilidades de interpretá-lo às vezes centralizado no recurso, e outras no evento ou processo (seção 2.4.3). Yang et al. (2003) argumenta que as pessoas interpretam (questionam) o mundo a partir da formulação de perguntas sobre essa dualidade. Tais perguntas podem ser respondidas na medida em que dados que descrevem as entidades questionadas, de alguma forma, foram preservados (seção 2.4.2.1) ao longo da história. É nosso objetivo, ressaltar que há caminhos que alavancam o reuso e aumentam as possibilidades de integração (seção 3.3).

Ao longo deste capítulo, apresentamos o nosso modelo de proveniência como uma abstração, não para um domínio específico, mas que procura ser intercambiável, ou seja, o modelo conceitual genérico para proveniência deve

ser capaz de oferecer um conjunto de metadados de uso comum a diferentes domínios para normalizar o intercâmbio dos dados entre aplicações.