

UM SERVIÇO PARA GERAÇÃO DE ESQUEMAS DE DADOS UTILIZANDO ARQUÉTIPOS

André Magno Costa de Araújo¹, Valéria Cesário Times¹, Marcus Urbano da Silva¹

Centro de Informática (CIN), Universidade Federal de Pernambuco, Recife, Brasil.

Resumo: Este trabalho apresenta um serviço capaz de criar dinamicamente esquemas de dados relacionais para armazenar o Registro Eletrônico em Saúde (RES) utilizando arquétipos. Baseado na arquitetura da openEHR, nós especificamos um conjunto de regras que mapeiam os atributos de dados, terminologias e restrições, e em seguida, cria as tabelas, campos e restrições de integridade referenciais do esquema de dados. Para validar o serviço aqui proposto, nós utilizamos arquétipos disponíveis no repositório da openEHR para extrair os seus elementos, e assim, gerar esquemas de dados relacionais para armazenar o RES. Como principais contribuições, destaca-se que, esquemas de dados são criados a partir de um padrão em saúde que uniformiza os atributos, terminologias e restrições. Além disso, a especificação do RES feita por um especialista do domínio por meio de arquétipos é utilizada no desenvolvimento de aplicações de saúde.

Palavras-chave: Software, Integração de Sistemas, Sistemas de Informação.

Abstract: This paper presents a relational data schemas creation service to store Electronic Health Record (EHR) using archetypes. Based on the openEHR architecture, we specify a set of rules which map data attributes, terminologies and constraints. Afterward, tables, fields and referential integrity restrictions of data schemas are dynamically generated. In order to validate the proposed service, we use archetypes from the openEHR repository to create the relational data schemas. As main contributions, we highlight that the data schemas are created based on a health standard and EHR specifications, made by a health professional, are used to develop health applications.

Keywords: Software, Systems Integration, Information Systems

Introdução

Os sistemas de Informação em Saúde (SIS) exercem hoje um importante papel na sociedade. Os SIS devem prover recursos de unicidade do Registro Eletrônico em Saúde (RES), de modo que o histórico clínico seja preservado e compartilhado com outros domínios da área da saúde¹. Ao longo dos últimos anos, é notório o esforço demandado por instituições governamentais, indústrias de software e a academia, na busca por soluções que padronizem o processo de desenvolvimento de aplicações em saúde, e que garantam a interoperabilidade do RES.

A criação de esquemas de dados para o armazenamento do RES é um tema relevante no ciclo de vida de um SIS e tem sido objeto de estudo de diversos trabalhos de pesquisa^{2,3}. Segundo a norma ISO/TS 18308⁴, o esquema de dados do RES deve prover alguns requisitos básicos, como: i) armazenar qualquer evento clínico relevante para os cuidados do paciente e ii) permitir o armazenamento de dados em tabelas de modo que as relações de dados entre as linhas e colunas sejam preservadas.

Diversos trabalhos de pesquisas têm apontado o uso do conceito de arquétipos da fundação openEHR, como uma alternativa para se construir aplicações em saúde que atendam as características de interoperabilidade e extensibilidade de esquemas de dados^{5,6}. A arquitetura dual proposta pela

openEHR realiza a separação dos atributos de dados que modelam o RES, das restrições e terminologias associadas aos dados clínicos, também conhecido como modelo de conhecimento. A especificação da modelagem dual proposta pela openEHR é realizada por meio de arquétipos. Na área da saúde, um arquétipo pode ser definido como uma expressão computacional utilizada para uniformizar os atributos de dados, terminologias e restrições do RES⁷.

A arquitetura proposta pela openEHR vem sendo debatida e amplamente divulgada nos últimos anos⁸. No entanto, nota-se que muitas são as dificuldades relatadas para se utilizar arquétipos em SIS⁹. As dificuldades relatadas se referem principalmente à falta de abordagens e ferramentas que construam esquemas de dados para o RES a partir de arquétipos^{7,9}. Nesse sentido, esse trabalho especifica um serviço capaz de criar dinamicamente esquemas de dados relacionais a partir da extração dos atributos de dados, terminologias e restrições de arquétipos. Para isso, nós especificamos um serviço chamado de *Representational State Transfer (REST) Application Programming Interface (API)* que a partir da leitura de um arquétipo retorna os elementos extraídos em formato *JavaScript Object Notation (JSON)*. A partir de requisições *GET* e *POST* na *API* desenvolvida, um segundo serviço consome os dados *JSON* e cria dinamicamente as tabelas, campos e restrições de integridade referenciais do esquema de dados proposto.

As demais seções desse artigo estão organizadas da seguinte forma. A Seção métodos descreve a metodologia de pesquisa utilizada no desenvolvimento deste trabalho, os conceitos básicos de arquétipos e os trabalhos correlatos identificados no estado da arte. A Seção Resultados descreve o serviço proposto neste artigo, enquanto que a Seção Discussão demonstra a criação de esquemas de dados a partir de arquétipos disponíveis no repositório da openEHR. Por fim, as considerações finais são descritas na Seção Conclusão.

Métodos

Para o desenvolvimento do serviço de criação de esquemas de dados proposto neste artigo, utilizou-se como metodologia a revisão de literatura e a pesquisa exploratória. Nesse sentido, investigou-se nos modelos computacionais fornecidos pela openEHR (i.e., *Demographic Information Model, Data Structures Information Model e openEHR Architecture*), os componentes necessários para a especificação do serviço aqui proposto; principalmente, como mapear dinamicamente arquétipos em esquema de dados relacionais. Para isso, buscou-se o entendimento das seguintes questões: i) como os arquétipos estão estruturados; ii) quais requisitos são necessários para a criação de um esquema de dados utilizando arquétipos; por fim, iii) a identificação das restrições e relacionamentos associados a cada elemento de um arquétipo. Além disso, esta seção descreve ainda os conceitos básicos utilizados para o desenvolvimento desse trabalho, e apresenta uma análise dos principais trabalhos correlatos.

Modelagem Dual e Arquétipos – A modelagem dual proposta pela fundação openEHR, objetiva a criação de um RES único, preservando o histórico e a evolução dos dados clínicos do paciente, podendo ser compartilhado e reutilizado por outros modelos de domínios da saúde¹. A arquitetura dual possibilita a separação entre as propriedades clínicas e demográficas do RES, dos padrões e das terminologias utilizados para dar significado semântico aos dados. O primeiro nível da modelagem dual contempla os componentes de linguagem de programação, linguagem de troca de informações e todos os outros componentes relacionados ao desenvolvimento de softwares, tais como: tecnologia de banco de dados e artefatos de softwares (e.g., diagramas da *Unified Modeling Language*, modelo conceitual e lógico de banco de dados). Um arquétipo pode ser definido como uma expressão computacional representada por restrições de domínio e baseada no modelo de referência⁷.

Em um arquétipo, a especificação dos atributos é realizada por meio de construtores de entrada de dados chamados de estruturas de dados genéricas. As estruturas de dados genéricas permitem

representar a heterogeneidade dos dados do RES por meio dos tipos, *ITEM_SINGLE*, *ITEM_LIST*, *ITEM_TREE* e *ITEM_TABLE*.

O tipo *ITEM_SINGLE* modela um único atributo de dado. Por exemplo, pode ser usado para especificar o sexo do paciente. O tipo *ITEM_LIST* agrupa um conjunto de atributos e é fisicamente representado na forma de uma lista. O endereço de um paciente com os itens, rua, número e CEP, exemplifica a modelagem por meio de um *ITEM_LIST*. O tipo *ITEM_TREE* especifica uma estrutura hierárquica de dados que é logicamente representada na forma de uma árvore. Ela pode ser usada, por exemplo, para modelar a avaliação física e neurológica de um paciente. Por fim, o tipo *ITEM_TABLE* modela os elementos de dados por meio de linhas e colunas, onde as linhas representam a definição do elemento, e as colunas o valor da informação. O resultado de um exame de análises clínicas contendo os campos dispostos em linhas, e os valores de referências em colunas, pode ser indicado como exemplo da *ITEM_TABLE*.

Cada atributo de uma estrutura de dados é caracterizado por um tipo de dado e pode ter ainda um conjunto de restrições de domínio e terminologias associadas. As terminologias dão significado semântico aos dados clínicos e podem ser representadas por meio de uma padronização de termos em saúde ou uma informação textual definida por um especialista do domínio.

Trabalhos Correlatos – A partir dos trabalhos investigados no estado da arte, apresentamos nesta seção uma análise sobre os principais estudos correlatos nas áreas de: i) desenvolvimento de aplicações em saúde baseadas em arquétipos e ii) mapeamento e persistência de arquétipos em banco de dados.

Späth e Grimson (2010)⁹ utilizaram a especificação fornecida pela openEHR para modelar e reestruturar o RES em um sistema legado. Inicialmente, o esquema do banco de dados proprietário foi reestruturado para que cada campo das tabelas utilizadas do banco de dados estivesse vinculado a um elemento do arquétipo. Os autores indicam que durante o processo de mapeamento dos arquétipos para o banco de dados, diversas dificuldades foram encontradas como, a falta de ferramentas e metodologias que apoiem o processo de modelagem de arquétipos em banco de dados.

De forma similar, Georg, Judith e Christoph (2013)², propõem uma abordagem para mapear arquétipos em uma base de dados legada e gerar templates para uma aplicação em saúde chamada de *ArchiMed*. Um conjunto de regras de mapeamento foram especificadas em um algoritmo, de modo a vincular cada elemento de dado do arquétipo com um campo de uma tabela no banco de dados. Os resultados apresentados indicam que a solução proposta não conseguiu mapear os arquétipos que continham estruturas de dados hierárquicas.

Outra alternativa para mapear arquétipos em repositório de dados é a criação de um mecanismo de persistência que serialize os elementos dos arquétipos em SGBD's. Nesse sentido, a fundação openEHR disponibiliza uma solução de persistência que serializa os as estruturas de dados em banco de dados relacional ou objeto relacional. A solução chamada de Node+Path¹⁰ utiliza o princípio da abordagem *Entity-attribute-Value (EAV)* para serializar todos os nós de dados em uma tabela de duas colunas. Assim, armazena-se o caminho do nó na primeira coluna, enquanto que o valor do nó serializado é armazenado na segunda coluna. No entanto, apesar da simplicidade de se armazenar estruturas hierarquizadas de dados, a recuperação das informações exige a criação de sentenças lógicas e complexas que podem comprometer o desempenho da aplicação.

Wang et al.(2015)¹² especificam um novo método de persistência de dados para apoiar o processo de mapeamento de arquétipos em banco de dados relacional. A solução consiste em um conjunto de regras de mapeamento que extraem os atributos de dados dos arquétipos e os armazena na forma de tabelas e colunas de dados. O método proposto estende também a linguagem *Archetype Query Language (AQL)*, adicionando os comandos de *Insert*, *Update* e *Delete* para suportar a manipulação de dados. No entanto, observando o esquema de dados gerado a partir do método proposto, percebe-se que as restrições e terminologias especificadas nos arquétipos não são foram consideradas. Além disso, trata-se de uma solução para um domínio de problema específico.

Embora os trabalhos aqui apresentados representem um avanço no desenvolvimento de aplicações em saúde baseadas na arquitetura da openEHR, há um consenso que a forma como os arquétipos são mapeados e persistidos em banco de dados necessita evoluir. Em virtude da heterogeneidade dos requisitos de dados encontrados em um domínio da saúde, identificou-se que pouca atenção tem sido dada na investigação de uma abordagem que crie dinamicamente esquemas de dados relacionais utilizando arquétipos.

A principal motivação do trabalho aqui proposto consiste no desenvolvimento de um serviço capaz de extrair dos arquétipos, os atributos de dados, as terminologias e as restrições, e assim, gerar dinamicamente esquemas de dados relacionais para o armazenamento do RES. Existem duas vantagens principais no uso do serviço proposto nesse artigo. Primeiro, o esquema de dados é criado a partir de um padrão em saúde que uniformiza os requisitos do RES. Segundo, auxilia projetistas de banco no mapeamento e criação de esquemas de dados utilizando arquétipos.

Resultados

Serviço de Extração dos Elementos de Arquétipos – O serviço proposto nesse artigo utiliza quatro passos principais para a geração de esquemas de dados relacionais utilizando arquétipos. A Figura 1 mostra a sequência de atividades executada pela abordagem proposta neste trabalho.

A primeira atividade consiste em identificar o tipo do arquétipo que será mapeado para o esquema de dados. De acordo com a definição da openEHR, cada tipo de arquétipo (i.e., *OBSERVATION*, *EVALUATION*, *INSTRUCTION*, *ACTION*, *SECTION*, *COMPOSITION*, *SINGLE*, *LIST*, *TREE*, *TABLE*, *ADMIN_ENTRY*, *CLUSTER*) representa um conceito do RES que pode ser criado para modelar um dado domínio da saúde.



Figura 1: Ciclo de atividades da REST API.

Identificado o tipo do arquétipo, o serviço inicia a execução de um conjunto de regras de validação que avaliam a estrutura e os metadados do arquivo XML. O objetivo da atividade 2 é certificar-se que se trata de um arquétipo válido e em conformidade com a especificação da openEHR. Após validar a estrutura do arquétipo, a REST API extrai os elementos (i.e., atributos de dados, restrições e terminologias) e inicia o processo de criação do esquema de dados.

Para a geração do esquema de dados, o serviço desenvolvido busca por quatro elementos básicos no arquétipo, que são: a) tipo do arquétipo, b) estrutura de dados, c) terminologias e d) restrições. A Figura 2 mostra um arquétipo extraído do repositório da openEHR e criado por meio da ferramenta *Archetype Editor*¹². Nela é possível observar como os elementos estão dispostos no esquema XML.

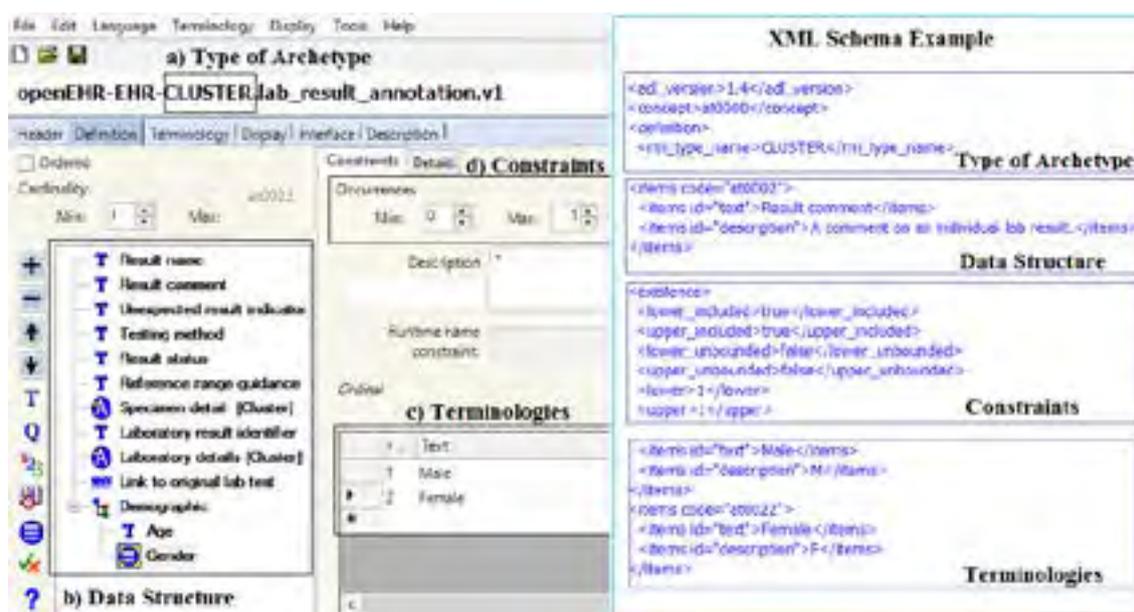


Figura 2: Estrutura e definição dos elementos de um arquétipo.

Os quatro elementos citados acima são de suma importância para a geração do esquema de dados. Primeiro, o tipo define o contexto em saúde que se está modelando o RES por meio de arquétipos. Segundo, as estruturas de dados (i.e., atributos) representam as características (i.e., campos) de cada arquétipo sobre o qual se deseja armazenar informações. Terceiro, na criação de um arquétipo, o especialista de domínio pode associar às estruturas de dados, um conjunto de terminologias ou valores pré-definidos. Por exemplo, na Figura 2, observa-se que para o atributo *Gender*, o especialista definiu previamente as terminologias, *Male* e *Female*. A utilização de terminologias e códigos pré-definidos é uma prática comum e facilitadora na área da saúde, principalmente porque ajudam na padronização de diagnósticos, medicamentos, exames, entre outros. Por esse motivo, todas as terminologias presentes em um arquétipo são mapeadas para o esquema de dados proposto. Por fim, as restrições especificadas sobre as estruturas de dados são importantes mecanismos de validação de dados em um arquétipo. Por exemplo, o campo de um determinado resultado de exame só poderá armazenar informações dentro de uma faixa de valores (e.g., de 10 a 20), ou a data da realização de um exame deve expressar o formato: dia, mês, ano e hora (e.g., 12/02/2016 14:20). Dada a sua importância, as restrições também são consideradas e adicionadas ao esquema de dados.

A Figura 3 mostra a arquitetura desenvolvida para que aplicações de saúde possam fazer uso do serviço de geração de esquemas de dados utilizando arquétipos.

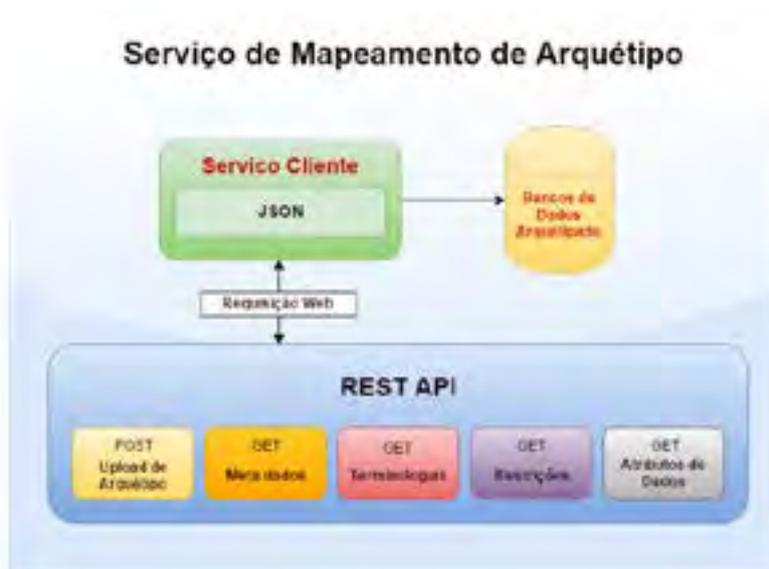


Figura 3: Arquitetura do Serviço de Mapeamento de Arquétipos

Cinco métodos principais realizam a tarefa de extrair dos arquétipos os elementos necessários para a criação de esquemas de dados, são eles: i) upload de arquétipos - por meio de uma requisição POST no endereço [http://polyehrap.azurewebsites.net/api/archetypes /upload/{organizationId}/{language}](http://polyehrap.azurewebsites.net/api/archetypes/upload/{organizationId}/{language}), esta funcionalidade realiza o upload de arquétipos XML para a extração dos seus elementos; ii) extração dos Metadados - esta funcionalidade retorna os metadados do arquétipo XML por meio de uma requisição GET no endereço <http://polyehrap.azurewebsites.net/api/archetypes/metadata/{archetypeId}>. Os metadados descrevem informações a respeito do tipo do arquétipo, objetivo, uso, autores e pode ser usado por exemplo, para realizar o controle de versões do arquétipo em um repositório de dados; iii) extração das estruturas de dados - as estruturas de dados especificadas nos arquétipos podem ser retornadas por meio de uma requisição GET no endereço <http://polyehrap.azurewebsites.net/api/archetypes/datastructure/{archetypeId}>. De acordo com especificação do arquétipo, esta funcionalidade pode retornar um array com um único elemento, uma lista de dados, uma estrutura de dados hierarquizada ou uma estrutura tabular com linhas e colunas; iv) extração das terminologias - esta funcionalidade retorna as terminologias e vocabulários em saúde especificados em cada atributo de dados de um arquétipo. Por meio de uma requisição GET no endereço [http://polyehrap.azurewebsites.net/ api/archetypes/terminologies/{archetypeId}](http://polyehrap.azurewebsites.net/api/archetypes/terminologies/{archetypeId}), uma lista contendo o identificador, a terminologia e a estrutura de dados é retornada para a aplicação solicitante. Por fim, v) extração das restrições - as restrições especificadas nos arquétipos validam e restringem o tipo, o formato e intervalo de dados em um atributo de dado. Por meio de uma requisição GET no endereço [http://polyehrap.azurewebsites.net/api/archetypes/constraints /{archetypeId}](http://polyehrap.azurewebsites.net/api/archetypes/constraints/{archetypeId}) é possível retornar, o identificador do atributo, juntamente com a lista de restrições encontradas no arquétipo.

Discussão

Criação de Esquemas de Dados – A partir da *REST API* desenvolvida, qualquer aplicação pode utilizar os métodos e consumir os elementos extraídos (i.e., padrão JSON) dos arquétipos para serem persistidos em bancos de dados legados. No entanto, nós desenvolvemos um segundo serviço que exemplifica e demonstra como um arquétipo pode ser mapeado dinamicamente para um esquema de dados relacional. Para essa demonstração, utiliza-se três arquétipos do repositório da openEHR (i.e., *Blood Pressure*, *Family history e Apgar*), e um arquétipo construído na ferramenta *Archetype Editor* que representa a consulta de um paciente (i.e., Figura 4).

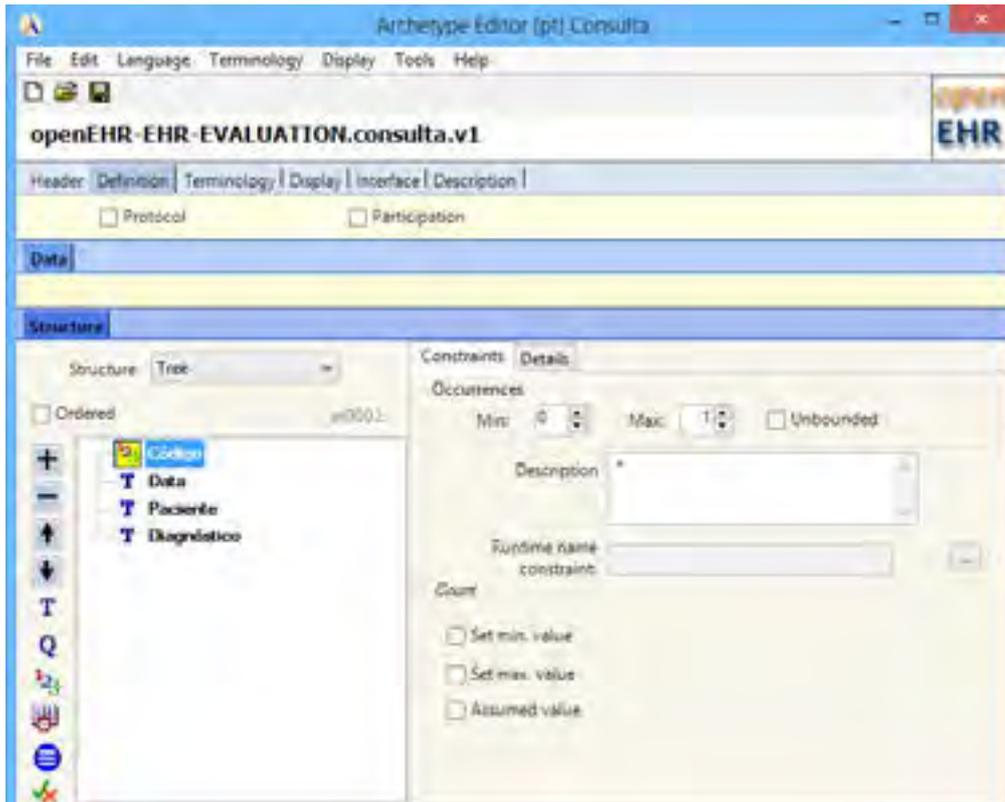


Figura 4: Exemplo do Arquétipo Consulta do Paciente

A Figura 5 mostra a interface e as principais funcionalidades do serviço desenvolvido para consumir os métodos da *REST API* e gerar o esquema de dados relacional. Por meio de uma requisição *POST* no método *upload ()*, nós enviamos os arquétipos para a extração dos seus elementos. Observe que a após a confirmação do método, o serviço obtém e mostra em formato *JSON*, os atributos de dados, terminologias e restrições dos arquétipos.



Figura 5: Serviço de Consumo da API e Criação do Esquema de Dados

Uma vez consumido os elementos dos arquétipos, nós utilizamos um conjunto de regras de mapeamento para construir dinamicamente o esquema de dados a partir dos elementos extraídos. Nesse caso, os atributos de dados são utilizados como campos, as terminologias são armazenadas e guardam a referência de cada atributo; por fim, as restrições restringem o conteúdo e o tipo de dado do atributo. A Figura 6 mostra o esquema de dados criado a partir dos quatro arquétipos utilizados nessa demonstração. Já a Figura 7 mostra um conjunto de registros inseridos na tabela *Blood Pressure* após a criação do esquema de dados.

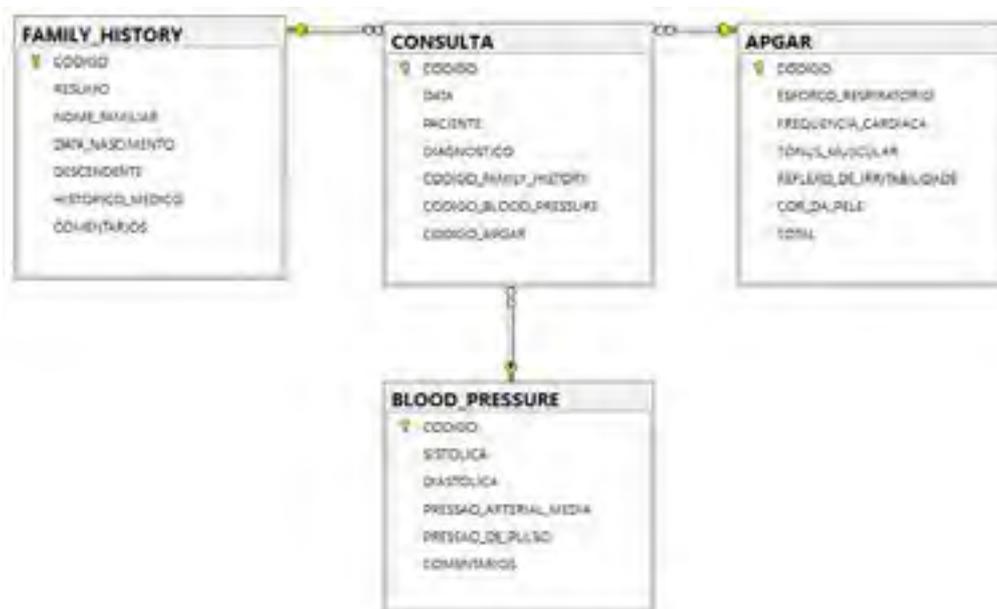


Figura 6: Esquema de dados criado utilizando Arquétipos

```
CREATE TABLE BLOOD_PRESSURE
(
    CODIGO INT PRIMARY KEY IDENTITY(1,1),
    SISTOLICA NUMERIC(10,3),
    DIASTOLICA NUMERIC(10,3),
    PRESSAO_ARTERIAL_MEDIA NUMERIC(10,3),
    PRESSAO_DE_PULSO NUMERIC(10,3),
    COMENTARIOS TEXT
);
SELECT FROM DBO.BLOOD_PRESSURE;
```

	CODIGO	SISTOLICA	DIASSTOLICA	PRESSAO_ARTERIAL_MEDIA	PRESSAO_DE_PULSO	COMENTARIOS
1	1	80.000	90.000	12.800	12.800	Paciente normal
2	2	80.000	100.000	14.800	14.800	
3	3	80.000	100.000	14.800	14.800	Sem observações

Figura 7: Exemplo da Tabela criada a partir do Arquétipo Blood Pressure.

Conclusão

Neste artigo nós apresentamos um serviço capaz de extrair e mapear dinamicamente os elementos de arquétipos openEHR em esquemas de dados relacionais. Inicialmente uma REST API extrai e disponibiliza em JSON, os atributos de dados, terminologias e restrições do arquétipo informado.

Depois, um segundo serviço consome os elementos extraídos, e por meio de um conjunto de regras de mapeamento, cria-se o esquema de dados relacional. Existem duas vantagens principais na criação de esquemas de dados a partir de arquétipos. Primeiro, utiliza-se um padrão em saúde que uniformiza os atributos de dados e terminologias em aplicações de saúde. Segundo, a especificação do RES feita por um especialista do domínio em um arquétipo é utilizada no desenvolvimento de SIS.

Agradecimentos

Este trabalho foi parcialmente financiado pela Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), sob as subvenções APQ-0173-1.03 / 15 e IBPG-0.809-1.03 / 13.

Referências

- [1] Marco E, Thomas A, Jorg R, Asuman D, Gokce L. A Survey and Analysis of Electronic Healthcare Record Standards. *ACM Computing Surveys*. 2005; 277–315.
- [2] D. Georg, C. Judith, R. Christoph. Towards plug-and-play integration of archetypes into legacy electronic health record systems: the ArchiMed experience. *BMC Medical Informatics and Decision Making*. 2013; 1-12.
- [3] André M C Araújo, Valéria C Times, Marcus U Silva. Uma Ferramenta de Modelagem baseada em Arquétipos, XIII Congresso Brasileiro de Informática em Saúde - CBIS 2012;
- [4] International Organization for Standardization: ISO/TS 18308 health informatics - requirements for an electronic health record architecture. Disponível em: http://www.iso.org/iso/iso_catalogue.htm. Acesso em jun. de 2016.
- [5] Michele M V F Santos, Luciana Tricai Cavalini. Modelagem da Ficha de Notificação de Violência no modelo openEHR. *J. Health Inform.* 2015; 7(3): 99-100.
- [6] Bernstein K, Bruun RM, Vingtoft S, Andersen SK, Nøhr C. Modelling and implementing electronic health records in Denmark. *International Journal of Medical Informatic*, 2005: 213-220.
- [7] D. Lloyd, T. Beale, S. Heard. openEHR Architecture: Architecture Overview. Disponível em: <http://www.openehr.org/releases/1.0.2/architecture/overview.pdf>
- [8] André M C Araújo, Valéria C Times, Sergio CB Soares. A Conceptual Data Model for Health Information Systems. *The 14th International Conference on Software Engineering Research and Practice*; 2016 Jul 25-28; LAS Vegas, USA.
- [9] Späth MB, Grimson J. Applying the archetype approach to the database of a biobank information management system. *International Journal of Medical Informatics*. 2010; 1-22.
- [10] Node + Path Persistence, disponível em <https://openehr.atlassian.net/wiki/pages/viewpage.action?pageId=6553626>. Acesso em jun. de 2016.
- [11] Wang L, Min L, Lu X, Duan H. “Archetype relational mapping - a practical openEHR persistence solution”; *BMC Medical Informatics and Decision Making*, pp. 1-18, 2015.
- [12] Archetype Editor. Disponível em <http://www.openehr.org/>. Acesso em jun. de 2016.

Contato

André Magno Costa de Araújo, Centro de Informática, UFPE, Av. Jornalista Aníbal Fernandes, s/n – Cidade Universitária (Campus Recife) 50.740-560 - Recife – PE, Fone: + 55 81 2126-8430, E-mail: amca@cin.ufpe.br

