

# Metadata - descrevendo recursos das agências do governo

22/01/2021

Autor: Nelson Henrique Zanete

## 1. Introdução

O problema comum na troca de dados e informações está normalmente relacionado à falta de um padrão, simples e igualmente compreendido quanto ao **significado** por todas as agências do governo. Ora, como podemos estabelecer e administrar um padrão comum, igualmente compreensível em tão grande diversidade de setores como escolas, saúde, polícia, agricultura, bibliotecas, saneamento, instituições de pesquisa e muitas outras? A resposta está na criação de comunidades de descrição de recursos. Comunidades de descrição de recursos requerem habilidades para descrever certos tipos de recursos. Para recursos bibliográficos, por exemplo, atributos descritivos incluindo "autor", "título" e "assunto" são comuns. Muitas comunidades estão adotando padrões de *Metadata* como alternativas para a descrição de recursos e troca de informações e alguns destes padrões já estão sendo aceitos por outras comunidades de peritos. *Metadata* é dados sobre dados. O termo refere a qualquer dado usado para ajudar a identificação, descrição e localização de recursos eletrônicos na rede. Visando promover o uso de *Metadata* e responder questões desta natureza, este trabalho está organizado como segue: A seção 1 apresenta, a título de motivação, as comunidades de descrição de recursos e o incentivo recebido pelas agências de governos de outros países; a seção 2 apresenta um modelo de representação de *Metadata* e suas influências de diferentes fontes e a seção 3 apresenta as considerações finais com algumas sugestões adicionais.

## 2. Comunidades de descrição de recursos

O *National Archives of Australia* é uma agência de manutenção do padrão de *Metadata* AGLS, (*Australian Government Locator Service*). O padrão AGLS de *Metadata* é um conjunto de 19 elementos descritivos que departamentos do governo e agências podem usar para melhorar a acessibilidade dos seus serviços e informações na Internet. **[NAT 2000]**.

Os Governos Australianos endossaram o uso de *Metadata* para facilitar a descoberta de informações e serviços por todas as esferas: municipal, estadual e federal. Para ser efetivo, toda agência deveria usar os mesmos elementos. Elas são estimuladas a usar o padrão AGLS para interagir prontamente com outras agências.

AGLS é simples, contudo, poderoso o bastante para descrever coleções de recursos produzidos pelo governo. AGLS está baseado no padrão *Dublin Core* (DC) de *Metadata* que é comumente usado em muitas indústrias e setores estatais.

Outra comunidade de descrição de recursos é a Rede de Educação da Austrália (EdNA - The Educational Network of Australia). É uma rede da comunidade de educação australiana (sistema de educação governamental, não governamental, vocação educacional, treinamento, educação da comunidade de adultos e ensino superior) e é organizada com base no currículo australiano e suas ferramentas são grátis a pedagogos. É fundada pelo corpo responsável pela provisão de educação na Austrália.

Elementos de *Metadata* são recomendados para uso na EdNA. Alguns setores têm identificado seus próprios conjuntos de elementos obrigatórios, em particular o setor das escolas.

A implementação de *Metadata* é como um sistema simples, aplicável por toda a rede de educação, embora alguns componentes sejam mais orientados a um setor que outro. (por exemplo, o elemento de *EDNA.Userlevel* é mais pertinente a escolas). O conjunto de *Metadata* EdNA é composto por nove elementos, além dos já suportados 15 elementos *Dublin Core*. São eles: *EDNA.Entered*, *EDNA.Approver*, *EDNA.Reassessment*, *EDNA.Userlevel*, *EDNA.Categories*, *EDNA.Conditions*, *EDNA.Indexing*, *EDNA.Review*, *EDNA.Version*.

Estes elementos são criados especificamente para a rede de educação da Austrália. O formato é semelhante aos elementos *Dublin Core*, mas estes elementos, não terão nenhum significado fora do ambiente da EdNA.

O *Dublin Core Metadata Initiative* [**DUB 2000**], um simples modelo de descrição de conteúdo para recursos eletrônicos é um conjunto de elementos de *Metadata* que pretende facilitar a descoberta de recursos eletrônicos. Originalmente concebido para descrição de autoria de recursos da *Web*, chamou a atenção de comunidades de descrição de recurso formais como museus, bibliotecas, agências de governo e organizações comerciais.

A principal característica do conjunto de elementos do Dublin Core é o consenso e a experiência coletiva de uma variedade de especialistas interdisciplinares e internacionais da área de descrição de recursos.

As características que o distinguem como um candidato proeminente para descrição de recursos eletrônicos, entram em várias categorias:

**Simplicidade:** é pretendido que o Dublin Core seja utilizável por leigos como também por especialistas de descrição de recursos. A maioria dos elementos tem uma semântica comumente compreendida.

**Semântica e interoperabilidade:** promovendo um conjunto de descritores de entendimento comum que ajudam a unificar outros conteúdos de dados padrões, incrementa a possibilidade de interoperabilidade através de disciplinas.

**Extensibilidade:** adicionalmente, o conjunto de elementos de Dublin Core, inclui flexibilidade suficiente e extensibilidade para codificar estrutura e semântica mais elaborada, inerente em padrões de descrição mais ricos.

**Modularidade de *Metadata* na rede:** a diversidade de necessidades de *Metadata* na rede requer uma infra-estrutura que apóia a coexistência de pacotes independentemente mantidos. O Consórcio W3C começou a implementar uma arquitetura para *Metadata* .

Muitos métodos diferentes podem ser usados para registrar ou transferir *Metadata Dublin Core*. Métodos comuns incluem HTML, XML, RDF e bancos de dados relacionais.

O conjunto completo dos 15 elementos é: *Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage e Rights* e são usados para tentar descrever o significado dos elementos de um recurso.

### 3. Modelo de representação de *Metadata*

O *Resource Description Framework* (RDF) é uma infra-estrutura, que possibilita a codificação, troca e reuso de *Metadata* estruturado. O RDF permite adicionar declarações, sinônimos e palavras que não estão presentes nos recursos, mas que são pertinentes a eles.

Como resultado de muitas comunidades trabalhando em conjunto e concordando com princípios básicos de representação de *Metadata* e transporte RDF, tem sofrido influências de várias fontes. A principal influência tem vindo da comunidade de padronização da *Web* (*Web Standardization Community*), da comunidade de biblioteca (*Library Community*), da comunidade de documentos estruturados (*Structured Document Community*) e também da comunidade de representação de conhecimento (*Knowledge Representation Community*).

## 2.1 Modelo RDF

O fundamento básico do RDF é um modelo para representar propriedades nomeadas e valores de propriedade. Propriedades de RDF podem ser pensadas como atributos de recursos que neste contexto correspondem ao tradicional par atributo-valor. O modelo básico consiste em três tipos de objetos:

**Recursos:** Todas as coisas sendo descritas pelas expressões RDF são chamadas de recursos. Um recurso pode ser uma página da *Web* inteira, tal como um documento HTML, um recurso pode ser parte de uma página *Web*, por exemplo, um elemento específico HTML ou XML dentro de um documento origem.

**Propriedades:** Uma propriedade é um aspecto específico, uma característica, um atributo, ou uma relação usada para descrever um recurso.

**Declarações:** Um recurso específico junto com uma propriedade nomeada mais o valor daquela propriedade para aquele recurso é uma declaração RDF. Uma declaração RDF pode ser desenhada usando diagramas de arcos e nodos, onde os nodos (desenho oval) representam os recursos, os arcos (setas) representam as propriedades nomeadas. Nodos que representam literais serão desenhados como retângulo.

A representação do modelo de dados é usada para avaliar equivalência de significado. Duas expressões RDF são equivalentes se, e somente se, suas representações de modelo de dados forem a mesma. Esta definição de equivalência permite algumas variações sintáticas nas expressões sem alterar o significado. Considere as duas sentenças abaixo (tabela 2.1):

Tabela 2.1 - Sentenças sintaticamente diferentes com o mesmo significado

|   |                               |
|---|-------------------------------|
| 1 | O autor do Documento 1 é João |
| 2 | João é autor do Documento 1   |

Para humanos, estas declarações levam o mesmo significado (quer dizer, o *João* é o autor de um documento particular). Para uma máquina, porém, estas são sentenças completamente diferentes. Considerando que os humanos são extremamente adeptos de extrair significado de diferentes construções sintáticas, máquinas permanecem grotescamente ineptas [MIL 98].

Usando o modelo recursos, propriedades e objetos, RDF tenta prover um método sem ambigüidades de expressar semântica em uma codificação legível por máquina.

As duas sentenças, acima descritas, no modelo proposto pelo RDF têm as seguintes partes (tabela 2.2):

Tabela 2.2 - Partes de uma declaração RDF

|             |             |
|-------------|-------------|
| Recurso     | Documento 1 |
| Propriedade | Autor       |
| Objeto      | João        |

Nesse contexto, o esquema do grafo rotulado seria o mesmo para as duas sentenças, tornando a codificação "legível por máquina; semanticamente equivalente.

Figura 2.1 - Representação de modelo de dados RDF



## 2.2 Esquema RDF

Um esquema define as condições que serão usadas em declarações de RDF e serão dados significados específicos a eles. **[W3C 99]**. O esquema é um lugar onde são documentadas definições e restrições de uso para propriedades. Para evitar confusões entre independentes e possivelmente conflitantes definições do mesmo termo, RDF usa a facilidade de *namespace XML*. **[W3C 98]**.

*Namespaces* simplesmente são um modo para amarrar um uso específico de uma palavra dentro de um contexto, para o esquema onde a definição

pretendida será encontrada. Em RDF, cada predicado usado em uma declaração deve ser identificado com exatamente um *namespace*, ou *schema*.

Em RDF, todos os vocabulários são expressos dentro de um único modelo bem definido. Isto permite gerenciar os vocabulários de maneira **descentralizada** pelas comunidades independentes de peritos. [W3C 2000].

### 3. Conclusão

Comunidades de descrição de recursos têm-se preocupado com a localização, acesso e troca de informações em um contexto específico. Para esse grupo de pessoas, a linguagem XML tem prometido bons resultados. A proposta de *metadata* RDF não pretende substituir as tecnologias existentes de troca de informações, no entanto, sugere a adição de mais um recurso a elas. Com o significado e tipo das propriedades colocados juntos com o par atributo-valor de cada recurso, consultas, validações, verificações de integridade e decisões internas de processos de softwares se tornam mais simples. Mecanismos de validação poderiam evitar erros no uso das marcações; editores interativos poderiam sugerir valores válidos para propriedades, indexadores de texto poderiam reduzir a sobrecarga de informações, esquemas poderiam ser traduzidos para idiomas diferentes, participação e respaldo das agências que dominam aquela área de conhecimento, bem como o privilégio de manutenção e autoria do padrão, poderiam aproximá-las incentivando a colaboração, além de muitas outras possibilidades.

O custo de se adotar uma tecnologia desse tipo não está relacionado, na grande maioria, aos recursos de hardware e software. Pelo contrário, estão relacionados a uma metodologia de trabalho, pois é preciso aderir a uma estrutura conceitual para poder aplicá-las e usá-las.

Embora haja poucas evidências que comprovem a plena utilização de *metadata* pelas empresas, acreditamos que uma pequena mudança na cultura e na metodologia utilizada possa viabilizar esse projeto, trazendo para os esforços deste governo uma importante infra-estrutura para comunicação e troca de informações na rede.

### 4. Referências

1. DUBLIN CORE METADATA INITIATIVE. About the dublin core metadata initiative. (October 18, 2000). Disponível em: <http://dublincore.org/about/>. Acesso em: 23 set. 2002.

2. MILLER, E. An introduction to the resource description framework. (November 5, 1998). Disponível em: <http://www.dlib.org/dlib/may98/miller/05miller.html>. Acesso em: 23 set. 2002.
3. NATIONAL ARCHIVES OF AUSTRALIA. Australian government locator service (AGLS) introduction. (March 30, 2000). Disponível em: [http://www.naa.gov.au/recordkeeping/gov\\_online/agls/summary.html](http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html). Acesso em: 25 set. 2000.
4. W3C (WORLD WIDE WEB CONSORTIUM). (January 19, 1998). Disponível em: <http://www.w3.org/TR/1998/NOTE-xml-names-0119.html>. Acesso em: 23 set. 2002.
5. W3C (WORLD WIDE WEB CONSORTIUM). Resource Description Framework (RDF) model and syntax specification. (February 22, 1999). Disponível em: <http://www.w3.org/TR/REC-rdf-syntax/>. Acesso em: 23 set. 2002.
6. W3C (WORLD WIDE WEB CONSORTIUM). Resource Description Framework (RDF) schema specification 1.0 (March 27, 2000). Disponível em: <http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>. Acesso em: 23 set. 2002.