

Metadata for Information Management and Retrieval: understandings metadata and its use. David Haynes

JOYCE SIQUEIRA

Introdução

1. Resenha
2. Capítulo 2. Definindo, descrevendo e expressando metadados.
3. Capítulo 3. Modelagem de dados.

Capítulo 2. Definindo, descrevendo e expressando metadados

Metadados

1. Metadados são informações que são adicionadas aos dados e que têm como objetivo tornar mais fácil a sua organização.
2. Um item de um metadado pode informar do que se trata aquele dado numa linguagem inteligível para um computador.
3. Os metadados tem a função de facilitar o entendimento dos relacionamentos e evidenciar a utilidade das informações dos dados.

As linguagens de marcação de documentos *Standard Generalized Markup Language - SGML* e *eXtensible Markup Language - XML* são amplamente utilizadas para expressar padrões de metadados.

SGML

1. A linguagem de marcação **Standard Generalized Markup Language (SGML)** é um padrão internacional, não proprietário, utilizado para troca de dados e que pode ser utilizada por diferentes sistemas informatizados.
2. Um dos objetivos do SGML é garantir que documentos codificados de acordo com suas regras possam ser transportados sem perda de informação.



SGML

Uma linguagem de marcação precisaria de três atributos:

- Representação comum de dados - para que diferentes sistemas e aplicativos possam processar texto na mesma representação.
- A marcação deve ser extensível - para suportar todos os diferentes tipos de informações que devem ser trocadas. Existe uma variedade potencialmente infinita de tipos de documentos que podem ser gerados.
- Os tipos de documentos precisam de regras - regras formais para documentos de um tipo específico, que podem ser usadas para testar sua conformidade com o tipo e, portanto, como eles são processados.

1986
Electronic
Publishing
Challenges

SGML

```
graph TD; SGML[SGML] --> HTML[HTML]; SGML --> XML[XML];
```

1991
Web
Presentation
Challenges

HTML

1996
Data
Representation
Challenges

XML

HTML (HyperText Markup Language)

```
<!DOCTYPE html>
<html>
<head>
<title>Page Title</title>
</head>
<body>

<h1>This is a Heading</h1>
<p>This is a paragraph.</p>

</body>
</html>
```

XML

1. Trata-se de um subconjunto do SGML.
2. Oferece a capacidade de representar dados de forma simples, flexível e legível por humanos.
3. É um padrão aberto.
4. O XML pode ser usado como base para o intercâmbio de dados ou documentos entre pessoas, computadores e aplicativos.
5. Ele vai além do HTML, pois fornece a expressão de um contexto semântico para os dados, além de ser o componente semântico que dá ao XML a capacidade de trocar dados de maneira significativa, e essa é uma das razões para sua ampla aceitação.

HTML mostra
como
apresentar

```
<h1>Severino Severovitch</h1>
<h2>bill@norte.com.br</h2>
<p>
  <b>11</b>
  <i>9999 4321</i>
</p>
```

XML mostra
o que
significa

```
<nome>Severino Severovitch</nome>
<email>bill@norte.com.br</email>
<telefone>
  <ddd>11</ddd>
  <numero>9999 4321</numero>
</telefone>
```

```
<breakfast_menu>
  <food>
    <name>Belgian Waffles</name>
    <price>$5.95</price>
    <description>Two of our famous Belgian Waffles with plenty of real maple syrup</description>
    <calories>650</calories> </food>
  <food>
    <name>Strawberry Belgian Waffles</name>
    <price>$7.95</price>
    <description>Light Belgian waffles covered with strawberries and whipped cream</description>
    <calories>900</calories> </food>
  <food>
    <name>Berry-Berry Belgian Waffles</name>
    <price>$8.95</price>
    <description>Light Belgian waffles covered with an assortment of fresh berries and whipped cream</description>
    <calories>900</calories> </food>
  <food>
    <name>French Toast</name>
    <price>$4.50</price>
    <description>Thick slices made from our homemade sourdough bread</description>
    <calories>600</calories> </food>
  <food>
    <name>Homestyle Breakfast</name>
    <price>$6.95</price>
    <description>Two eggs, bacon or sausage, toast, and our ever-popular hash browns</description>
    <calories>950</calories> </food>
</breakfast_menu>
```

Document Type Definitions (DTD)

1. Conjunto de regras para o uso de XML para representar documentos de um tipo específico.
2. Fornece uma forma de expressão de metadados nas linguagens de marcação, pois se refere ao vocabulário e às regras usadas para descrever os metadados.
3. Define os elementos (ou campos) de um documento. Isso significa que documentos semelhantes podem ser definidos pelo mesmo DTD.



```
<ELEMENT HOTEL (NAME, ADDRESS, CATEGORY, ROOM+)>
  <ELEMENT NAME(#PCDATA)>
  <ELEMENT ADDRESS(STREET, CITY, STATE, ZIP)>
    <ELEMENT STREET(#PCDATA)>
    <ELEMENT CITY(#PCDATA)>
    <ELEMENT STATE(#PCDATA)>
    <ELEMENT ZIP(#PCDATA)>
  <ELEMENT CATEGORY(#PCDATA)>
  <ELEMENT ROOM(TYPE, PRICE)>
    <ELEMENT TYPE(#PCDATA)>
    <ELEMENT PRICE(AMOUNT, CURRENCY)>
  <ELEMENT AMOUNT(#PCDATA)>
  <ELEMENT CURRENCY(#PCDATA)>
```

```
<?xml version="1.0"?>
<HOTEL>
  <NAME>Cityhotel</NAME>
  <ADDRESS>
    <STREET>54 Chambers Street</STREET>
    <CITY>New York</CITY>
    <STATE>New York</STATE>
    <ZIP>10007</ZIP>
  </ADDRESS>
  <CATEGORY>4</CATEGORY>
  <ROOM>
    <TYPE>single</TYPE>
    <PRICE>
      <AMOUNT>140</AMOUNT>
      <CURRENCY>USD</CURRENCY>
    </PRICE>
  </ROOM>
  <ROOM>
    <TYPE>double</TYPE>
    <PRICE>
      <AMOUNT>180</AMOUNT>
      <CURRENCY>USD</CURRENCY>
    </PRICE>
  </ROOM>
</HOTEL>
```

XML Schema

1. **XML Schema** é uma linguagem baseada no formato XML para definição de regras de validação em documentos no formato XML.
2. Foi a primeira linguagem de esquema para XML a obter o *status* de recomendação por parte do W3C - World Wide Web Consortium, que consiste em consórcio internacional com centenas de membros com a finalidade de estabelecer padrões para a criação e a interpretação de conteúdos para a Web.
3. Esta linguagem é uma alternativa ao DTD, cuja sintaxe não é baseada no formato XML.

XSDL (XML Schema Definition Language)

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  targetNamespace="http://xml.netbeans.org/examples/targetNS"
  xmlns:tns="http://xml.netbeans.org/examples/targetNS"
  xmlns="http://xml.netbeans.org/examples/targetNS"
  elementFormDefault="qualified">
  <xsd:element name="Customer">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="Name">
          <xsd:complexType>
            <xsd:attribute name="First" type="xsd:string" />
            <xsd:attribute name="Last" type="xsd:string" />
          </xsd:complexType>
        </xsd:element>
        <xsd:element name="Address" />
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```


Namespace

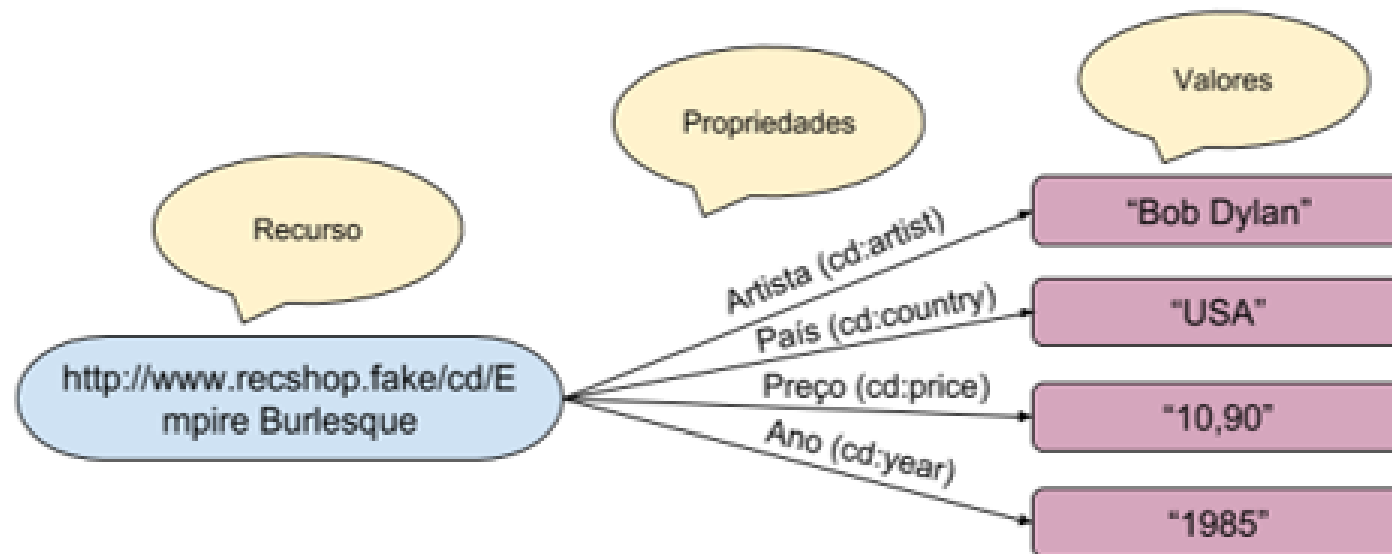
1. Utilizado para localizar definições da Internet para esquemas de metadados, garantindo maior consistência da terminologia usada para definir elementos de metadados e fornece uma maneira de compartilhar elementos.
2. O namespace que fornece a referência original aos elementos principais de Dublin é o seguinte: <http://purl.org/dc/elements/1.1/>

URI:	http://purl.org/dc/terms/abstract
Label:	Abstract
Definition:	A summary of the resource.
Type of Term:	Property
Refines:	http://purl.org/dc/elements/1.1/description
Refines:	http://purl.org/dc/terms/description
Version:	http://dublincore.org/usage/terms/history/#abstract-003
Term Name: accessRights	
URI:	http://purl.org/dc/terms/accessRights
Label:	Access Rights
Definition:	Information about who can access the resource or an indication of its security status.
Comment:	Access Rights may include information regarding access or restrictions based on privacy, security, or other policies.
Type of Term:	Property
Refines:	http://purl.org/dc/elements/1.1/rights
Refines:	http://purl.org/dc/terms/rights
Has Range:	http://purl.org/dc/terms/RightsStatement
Version:	http://dublincore.org/usage/terms/history/#accessRights-002

Capítulo 3. Modelagem de dados

RDF

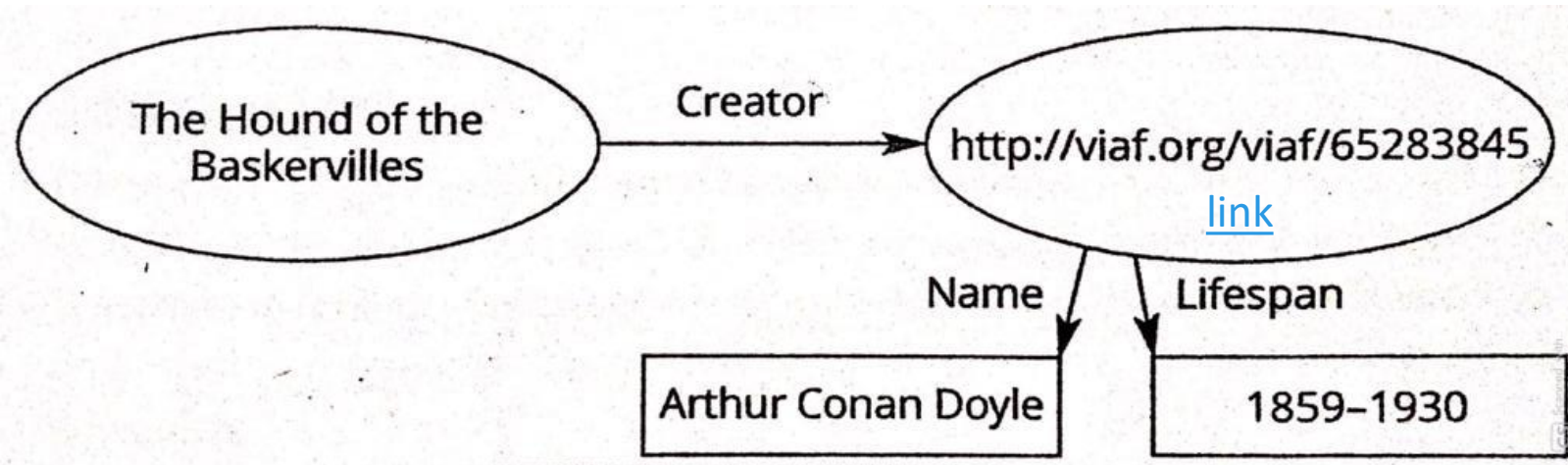
1. Estrutura escrita em XML.
2. Trata-se de um padrão para representar informações sobre dados na Web, destinado a representar metadados, que forneçam uma estrutura comum para expressar informações, sem perda de conteúdo, de forma interoperável (W3C RECOMMENDATION, 2004).
3. Baseia-se na concepção de que os dados possuem propriedades, que possuem valores e que os recursos podem ser descritos a partir de declarações (W3SCHOOLS.COM, 2018).



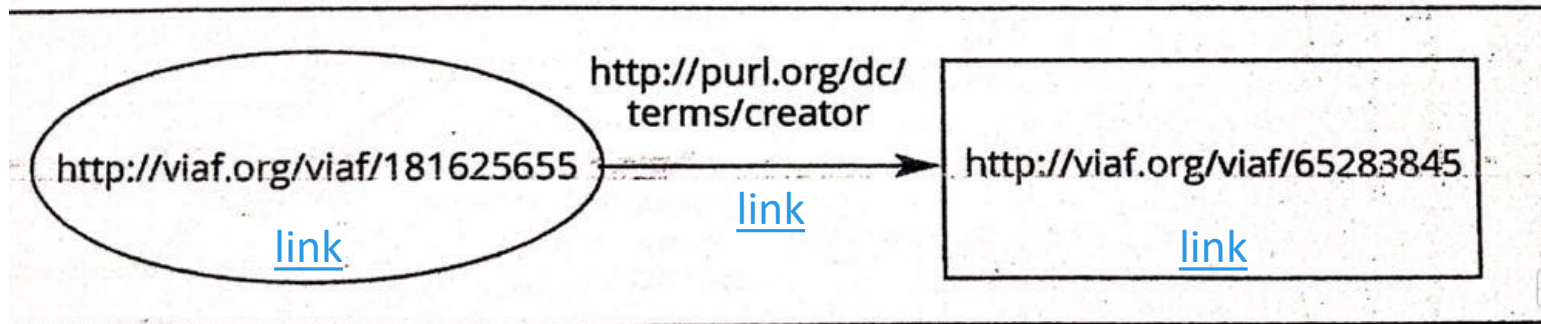
```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:cd="http://www.recshop.fake/cd#">
<rdf:Description rdf:about="http://www.recshop.fake/cd/Empire Burlesque">
  <cd:artist>Bob Dylan</cd:artist>
  <cd:company>Columbia</cd:company>
  <cd:price>10.90</cd:price>
  <cd:year>1985</cd:year>
</rdf:Description>
</rdf:RDF>

```

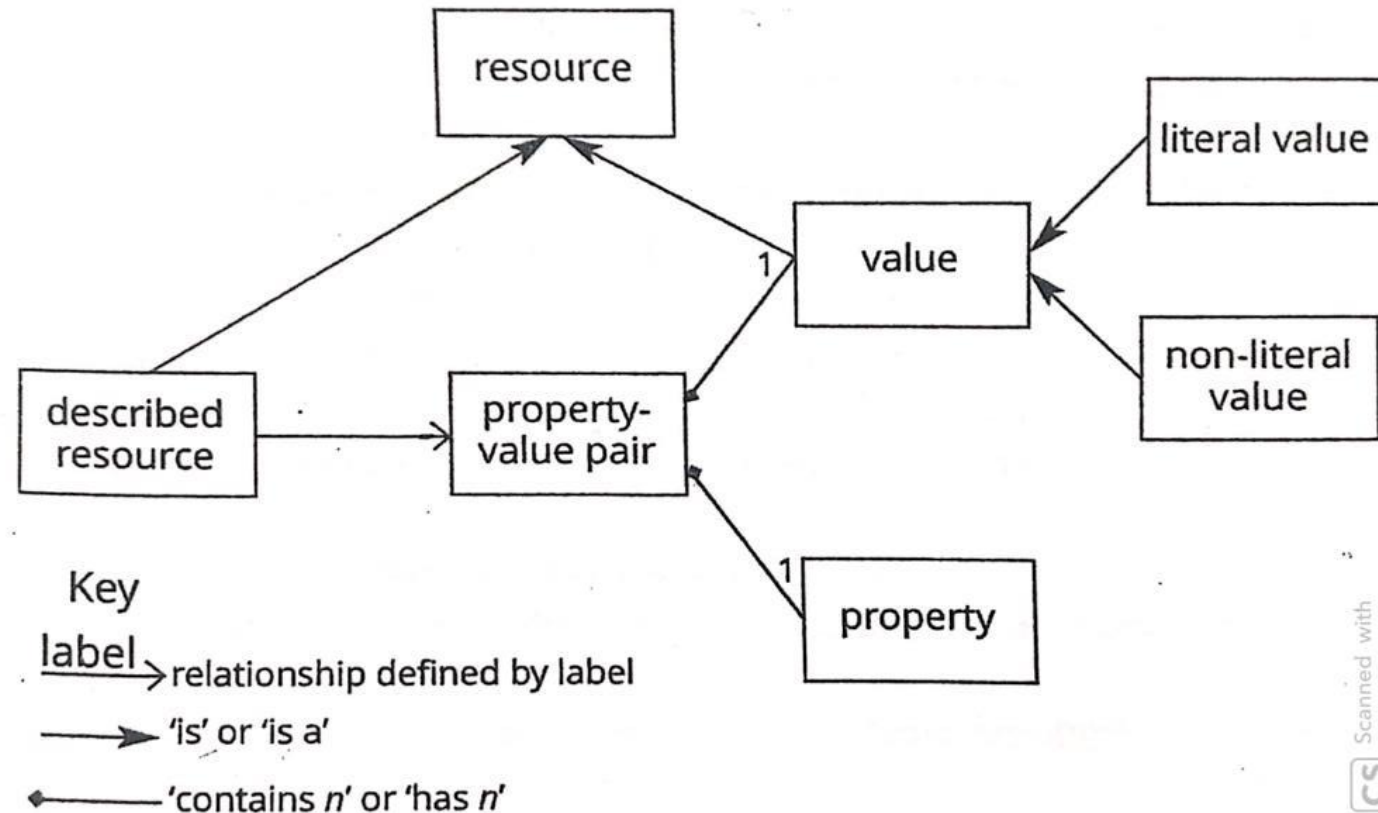


The Hound of the Baskervilles has author Arthur Conan Doyle
http://viaf.org/viaf/181625655 has dc:creator http://viaf.org/viaf/65283845
http://viaf.org/viaf/181625655 http://purl.org/dc/terms/creator
http://viaf.org/viaf/65283845



Dublin Core

1. É um padrão de metadados amplamente utilizado para descrever recursos on-line.
2. Dublin Core Metadata Initiative (DCMI) esta representado na figura abaixo, mostrando o relacionamento entre uma descrição de recurso e estabelecendo o fato de que cada par propriedade-valor contém uma propriedade e um valor. Por exemplo, a propriedade 'dc: creator' pode ter o valor 'Arthur Conan Doyle'.

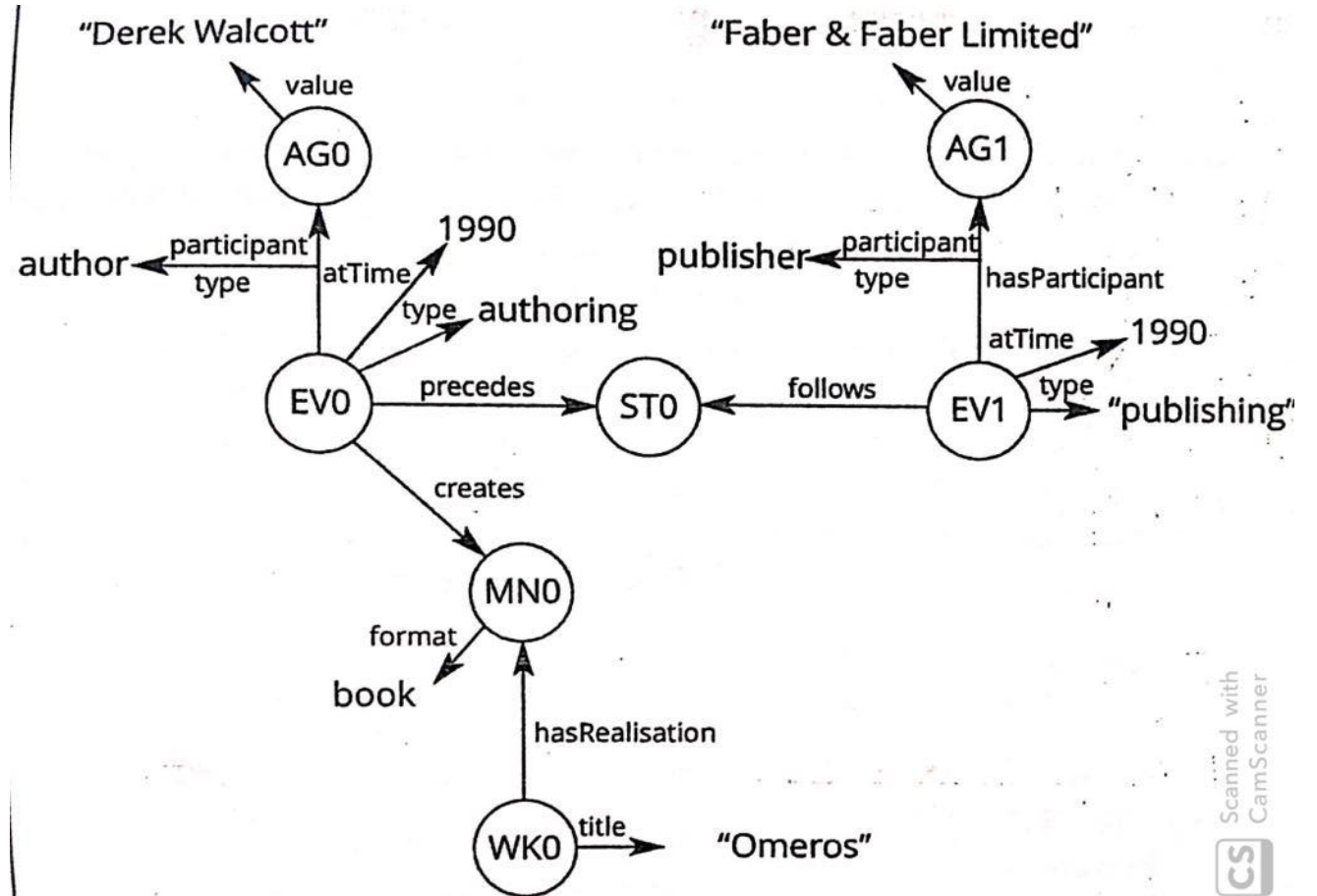


O Library Reference Model (LRM) e o desenvolvimento do RDA

1. RDA (Resource Description and Access) é um padrão de catalogação baseado no Library Reference Model, que substituiu os Requisitos Funcionais para Registros Bibliográficos (FRBR) e o Dados da Autoridade de Requisitos (FRAD)
2. O LRM é baseado na modelagem de relacionamento entre entidades e abrange dados. Foi projetado para suportar cinco tarefas genéricas do usuário: encontrar, identificar, selecionar, obter e explorar.

ABC Ontologia e Web Semântica

1. É um modelo básico de ontologia que fornece a base ideal para o desenvolvimento de domínio, função ou comunidade.
2. A Ontologia ABC foi desenvolvida para facilitar a interoperabilidade entre ontologias de metadados de diferentes domínios. Seu objetivo é 'modelar objetos físicos, digitais e analógicos mantidos em bibliotecas, arquivos e museus e na Internet'.



OAIS (Open Archival Information System)

O modelo OAIS foi desenvolvido para prover um modelo de preservação e acesso a dados.

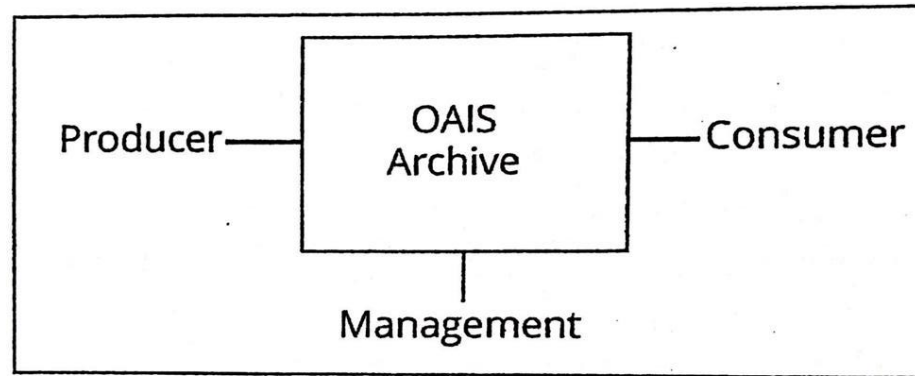


Figure 3.9 OAIS simple model

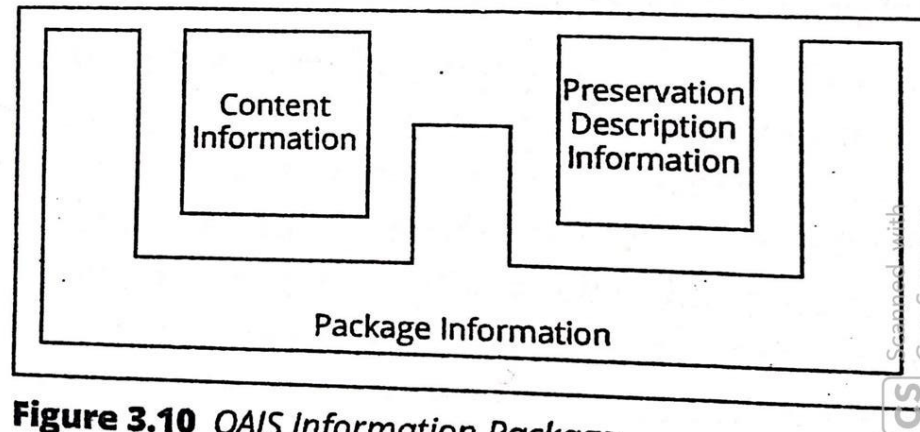


Figure 3.10 OAIS Information Package

Referências Bibliográficas

HAYNES, David. Metadata for Information Management and Retrieval: understandings metadata and its use. London: Facet Publishing. 2018. 2nd edition. 267p. Disponível na BCE. Capítulo 02 e 03.

W3CSchool. HTML. Disponível: <https://www.w3schools.com/html/default.asp>. Acesso em 16 set 2019.

W3CSchool. XML. Disponível: <https://www.w3schools.com/xml/default.asp>. Acesso em 16 set 2019.

Metadata for Information Management and Retrieval: understandings metadata and its use. David Haynes

JOYCE SIQUEIRA