




www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

## Armazienamento de Dados




Bernadette Farias Lóscio  
bfl@cin.ufpe.br



www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

- Agenda
  - Dados x metadados
  - Fontes de dados: de onde os dados podem ser extraídos?
  - Dados abertos: em qual formato publicar?
  - A importância dos metadados





www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE




DADOS por todos os lados








3




www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

- Dados
  - Fatos registrados, e que têm um significado implícito, sobre fenômenos do mundo real
  - Tipicamente representam valores (números, caracteres) de variáveis (qualitativas ou quantitativas)
  - Utilizados para transmitir, armazenar e deduzir informações

4



www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE




- metaDados
  - facilitam o entendimento dos relacionamentos e a utilidade das informações dos dados

título


Central do Brasil  
Walter Sales  
Fernanda Montenegro

diretor

atriz\_principal

5




www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

**dado** Valor sem um significado explícito

**informação** Significado associado ou deduzido de um conjunto de dados e de associações entre eles

**conhecimento** Informação adicional extraída dos dados ou do especialista do domínio da aplicação


6


 Armazienamento – Conceptos básicos  
 Bernadette Lóscio/CIn-UFPE



dado 8.848

A altura do monte Everest é 8.848m

informação


conhecimento


O Monte Everest é a maior montanha sobre a superfície da Terra e mede 8.848m





7


 Armazienamento – Conceptos básicos  
 Bernadette Lóscio/CIn-UFPE



Processar **informação** para gerar **conhecimento**


**conhecimento**


Processar **dados** para gerar **informação**

**informação**


**dado**

*Interesse comum de organizações públicas e privadas*








8




www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

## Quem são as fontes de dados?




9




www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

- Dados
  - Podem estar armazenados em
    - Fontes de dados privadas - disponíveis nas organizações
    - Fontes de dados públicas - disponíveis na Web
  - Fontes de dados: bancos de dados, documentos, planilhas, relatórios, formulários

10







www.od4d.org

## Armazienamento – Conceptos básicos

Bernadette Lóscio/CIn-UFPE

- Dados
  - Podem ter diferentes formatos
    - Dados estruturados (ex. bancos de dados relacionais)
    - Dados semi-estruturados (ex. documentos xml)
    - Dados não estruturados (ex. documentos texto)

11




www.od4d.org


## Armazienamento – Conceptos básicos

Bernadette Lóscio/CIn-UFPE

- Dados estruturados
  - Dados provenientes dos sistemas transacionais
  - Dados armazenados em bancos de dados relacionais
  - Possuem uma estrutura fixa e bem definida (esquema do banco de dados)
    - Esquema pré-definido
    - Todos os dados de acordo com o esquema

12





 Armazienamento – Conceptos básicos  
 Bernadette Lóscio/CIn-UFPE

[www.od4d.org](http://www.od4d.org)


Tabela Autores      Tabela Gêneros

<i>cod</i>	<i>titulo</i>	<i>cod_autor</i>	<i>ano</i>	<i>cod_genero</i>
L01	Mar Morto	autor_1	1936	genero_1
L02	A Estrada do Mar	autor_1	1938	genero_2
L03	O mundo da Paz	autor_1	1951	genero_3
L04	O Quinze	autor_2	1930	genero_1
L05	Tieta do Agreste	autor_1	1977	genero_1

Tabela relacional - Livros








13



 Armazienamento – Conceptos básicos  
 Bernadette Lóscio/CIn-UFPE  
 Dados

[www.od4d.org](http://www.od4d.org)

- Dados semi-estruturados
  - Ausência de uma estrutura regular, ou a estrutura é capaz de evoluir de forma imprevisível
  - Dados podem ser incompletos
  - Estrutura irregular (dados heterogêneos)
  - Tipos são apenas indicativos
  - A estrutura pode ser implícita
  - Dados na Web

14






www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

```
<?xml version="1.0" encoding="utf-8"?>
<livraria>
  <livro id="L01" ano="1936">
    <autor> Jorge Amado </autor>
    <titulo>Mar Morto</titulo>
  </livro>
  <livro id="L04" ano="1930">
    <autor>
      <nome>Rachel</nome >
      <sobrenome>de Queiroz</sobrenome >
    </autor>
    <titulo>O Quinze</titulo>
    <genero> Romance </genero>
  </livro>
</livraria>
```

15


www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

- **Dados não-estruturados**
  - **Ausência** de estrutura
  - Dados que ainda não foram “tratados” ou modelados
  - Dados armazenados em arquivos ou documentos







16













www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE




- Dados abertos: em qual formato publicar?
  - O formato deve ser processável por máquina
    - Deve permitir: o **uso/reuso**, **análise dos dados**, **diferentes visualizações**, **integração com outros dados**




www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

- Formatos de dados abertos
  - JSON
  - CSV
  - XML
  - RDF









www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE




- Formatos de dados abertos
  - JSON
    - É um formato de arquivo bem fácil de ser interpretado por qualquer linguagem de programação
    - Costuma ser mais fácil para os computadores processarem JSON do que outras




www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

- Formatos de dados abertos
  - CSV (comma separated vallues)
    - Formato compacto e, portanto, adequado para transferir grandes conjuntos de dados com uma mesma estrutura
    - É necessário oferecer documentação para explicitar a estrutura dos dados (o significado dos campos)









www.od4d.org

## Armazienamento – Conceptos básicos

Bernadette Lóscio/CIn-UFPE

- Formatos de dados abertos
  - Planilhas
    - Dados podem ser usados de forma imediata, adicionando-se apenas a descrição correta das colunas
    - Para disponibilizar uma planilha, pode ser necessário mudar o formato do arquivo, de forma que ela seja acessada por qualquer tipo de software






www.od4d.org


## Armazienamento – Conceptos básicos

Bernadette Lóscio/CIn-UFPE

- XML – Extensible Markup Language
  - Linguagem de marcação extensível
  - Metalinguagem para criação de outras linguagens
  - XML é o padrão para **troca** de dados na Web
  - Adequada para representação de dados semi-estruturados



22







www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

Exemplo de documento XML

```
<?xml version="1.0" encoding="utf-8"?>
<livraria>
  <livro id="L01" ano="1936">
    <autor> Jorge Amado </autor>
    <titulo>Mar Morto</titulo>
  </livro>
  <livro id="L04" ano="1930">
    <autor>
      <nome>Rachel</nome >
      <sobrenome>de Queiroz</sobrenome >
    </autor>
    <titulo>O Quinze</titulo>
    <genero> Romance </genero>
  </livro>
</livraria>
```

23

www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

- Composição de um documento XML - Elementos
  - Os elementos são os blocos principais da composição de um documento XML
  - Cada elemento possui um nome e um conteúdo
  - O conteúdo de um elemento é delimitado por marcações especiais, conhecidas como marcação (*tag*) de início e marcação (*tag*) de fim

CEPAL

IDRC CRDI Canada

W3C WORLD WIDE WEB Escritório Brasil




www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

As marcações de início e de fim são obrigatórias!




```
<livro>  
  <autor>Elmasri</autor>  
  <titulo>Fundamentals of Database Systems</titulo>  
  <preco>R$90.99</preco>  
</livro>
```




www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

- Lembre-se:
  - XML não define elementos
  - XML oferece uma sintaxe comum para armazenar informações de acordo com uma determinada estrutura
  - XML separa conteúdo de apresentação
  - As marcações de um documento XML oferecem significado ao seu conteúdo









www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

- Composição de um documento XML - O conteúdo dos Elementos
  - Cada elemento pode conter
    - caracteres
    - outros elementos
    - caracteres e outros elementos




www.od4d.org

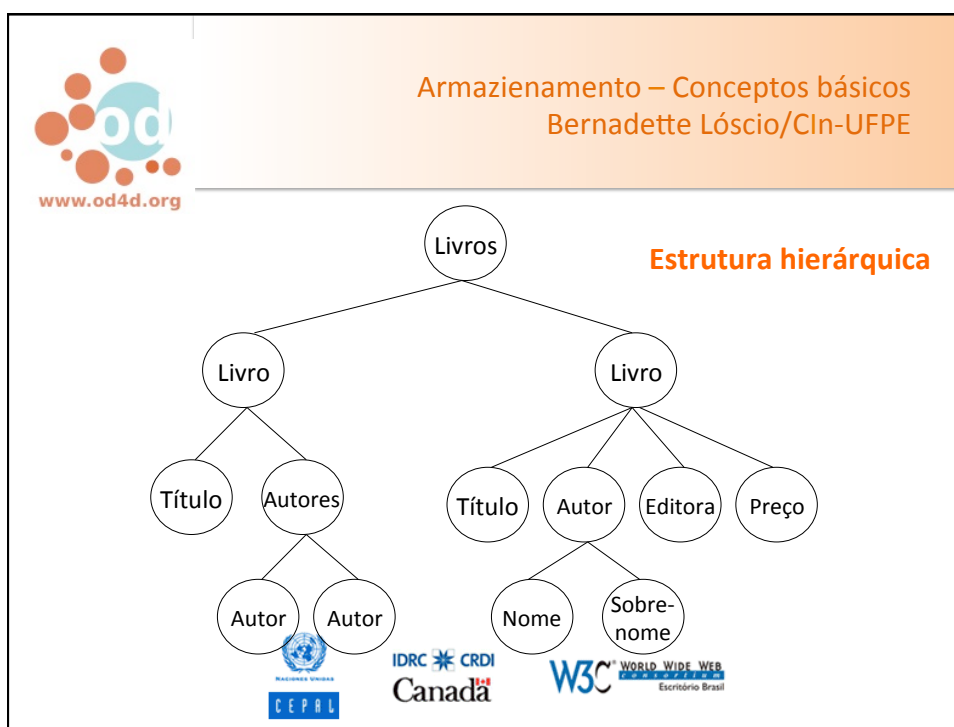
Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE  
Elementos Aninhados

Um elemento pode conter outros elementos

```

<livros>
  <livro>
    <titulo> Fund.... </titulo>
    <autores>
      <autor> Navathe </autor>
      <autor> Elmasri </autor>
    </autores>
  </livro>
</livros>
  
```



www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

**RDF - Resource Description Framework**

- RDF é o padrão para **publicação** de dados na Web
- Modelo de dados baseado em grafos
- Recursos são descritos por triplas

sujeito predicado objeto

*Paris é a capital da França*

CEPAL IDRC \* CRDI Canada W3C WORLD WIDE WEB Escritório Brasil

www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

### Grafos RDF

```

    graph TD
      ZeliaGattai[Zelia Gattai] -- married --> JorgeAmado[Jorge Amado]
      JorgeAmado -- livedIn --> Salvador[Salvador]
      JorgeAmado -- wrote --> MarMorto[Mar Morto]
      JorgeAmado -- wrote --> TietaDoAgreste[Tieta do Agreste]
      Salvador -- isIn --> Bahia[Bahia]
      Bahia -- partOf --> Brazil[Brazil]
      Ceara[Ceará] -- partOf --> Brazil
  
```

31

CEPAL IDRC CRDI Canada W3C WORLD WIDE WEB Escritório Brasil

www.od4d.org

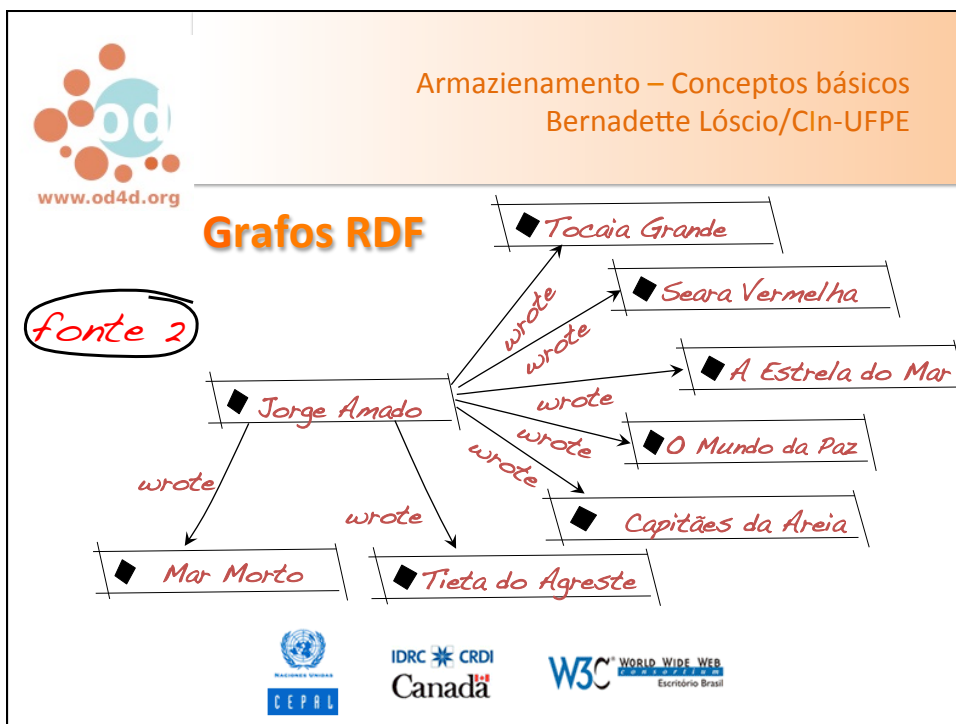
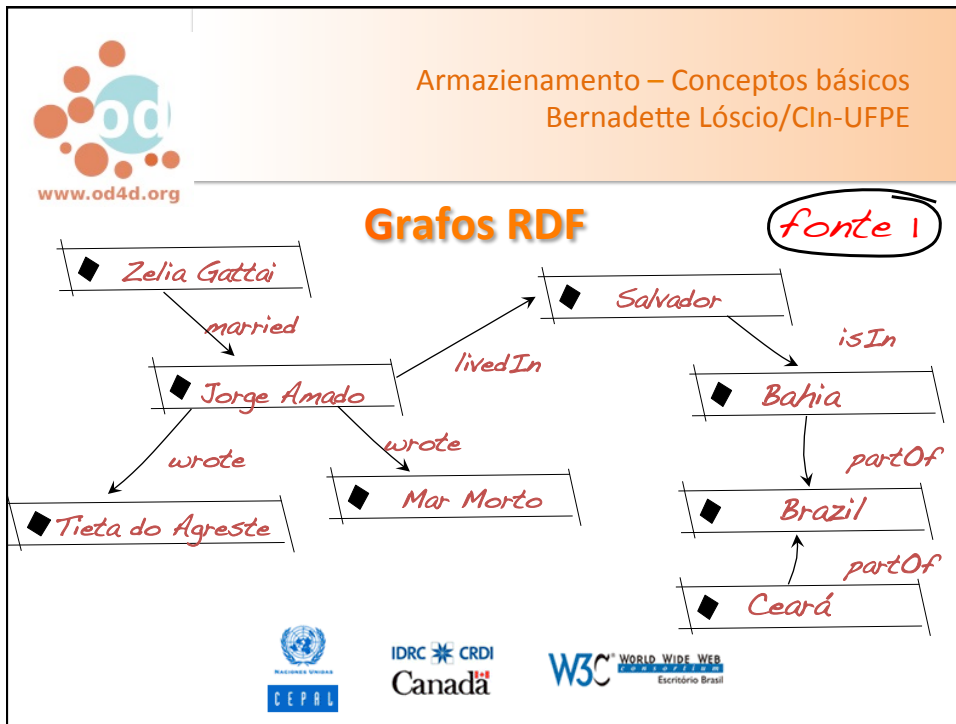
Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

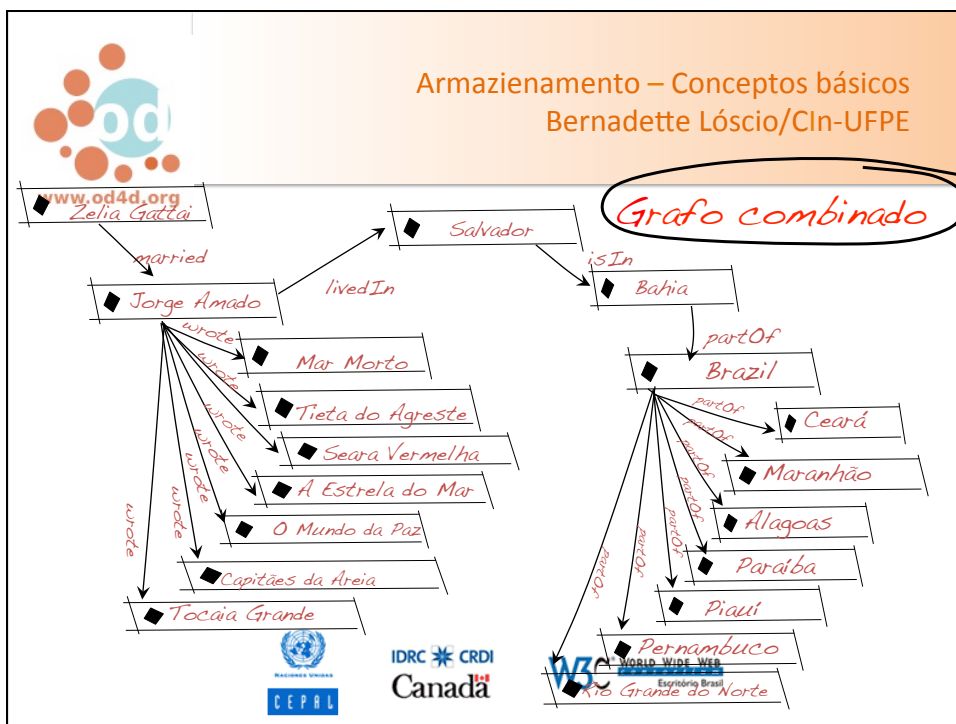
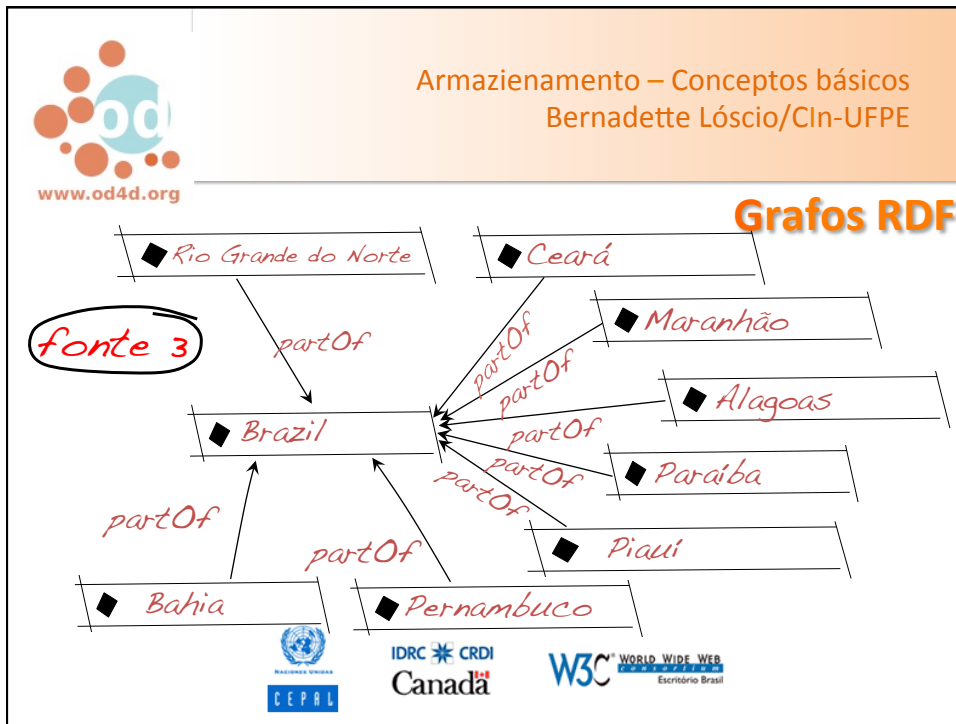
<i>sujeito</i>	<i>predicado</i>	<i>objeto</i>
Jorge Amado	Wrote	Mar Morto
Zelia Gattai	Married	Jorge Amado
Jorge Amado	Lived in	Salvador
Salvador	Is in	Bahia
Bahia	Part of	Brazil
Ceará	Part of	Brazil
Jorge Amado	Wrote	Tieta do Agreste


32

CEPAL IDRC CRDI Canada W3C WORLD WIDE WEB Escritório Brasil













www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE




- A importância dos metadados



www.od4d.org

Armazienamento – Conceptos básicos  
Bernadette Lóscio/CIn-UFPE

# Conclusões



38