

Padrões de Metadados

JOYCE SIQUEIRA

Introdução

Resenha: ZENG, Marcia Lei e QIN, Jian. Metadata Standards. In: ____.
Metadata. 2ª Edição. 555p. Chicago: Neal-Schuman/American Library Association. 2016. Capítulo 10, p. 401-484.

Propósito Geral dos Metadados

- Nas últimas décadas os metadados extrapolaram o universo bibliográfico, para também descrever objetos na web possibilitando acesso as informações.
- Dublin Core é considerada a primeira associação mental feita em relação aos metadados.
- Porém, Marc é o padrão de catalogação mais abrangente, criado de forma extensiva (antes da era digital) e sofreu várias transformações nos últimos anos, a fim de atender aos requisitos da era digital.

Dublin Core

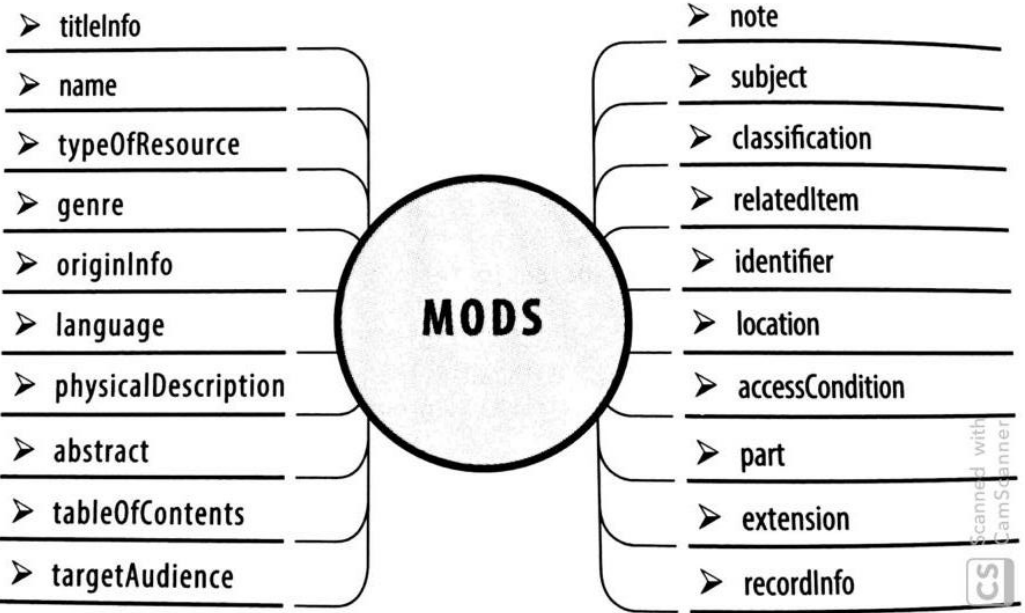
- Proposto para descrever os recursos para documentos eletrônicos em rede.
- Antes dele:
 - índices gerados automaticamente, usados pelos mecanismos de pesquisa, que geralmente continham poucas informações para serem úteis.
 - registros de catalogação, tais como MARC e Text Encoding Initiative (TEI), ambos praticamente impossíveis de serem adotados na descrição dos documentos eletrônicos disponíveis na Internet.
- O Dublin Metadata Workshop pretendeu transcender as fraquezas colocadas por esses dois, com desejo de fornecer uma estrutura simples e gerenciável para a descrição de recursos eletrônicos.

DC Element Name	Definition
1. Title	A name given to the resource.
2. Creator	An entity primarily responsible for making the resource.
3. Subject	The topic of the resource.
4. Description	An account of the resource.
5. Publisher	An entity responsible for making the resource available.
6. Contributor	An entity responsible for making contributions to the resource.
7. Date	A point or period of time associated with an event in the lifecycle of the resource.
8. Type	The nature or genre of the resource.
9. Format	The file format, physical medium, or dimensions of the resource.
10. Identifier	An unambiguous reference to the resource within a given context.
11. Source	A related resource from which the described resource is derived.
12. Language	A language of the resource.
13. Relation	A related resource.
14. Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
15. Rights	Information about rights held in and over the resource.

MODS e a Família MARC

- A grande quantidade de dados de catalogação nos formatos MARC é um vasto e valioso repositório de informações.
- No entanto, deve ser lido, processado, recuperado e usado.
- Vale ressaltar que os novos formatos não estão convertendo os campos de dados MARC em tags XML, mas que mais semântica está sendo adicionada para atender às novas necessidades de dados antigos.
- É por esse motivo que o *Object Description Schema* (MODS) foi projetado.

FIGURE 10-1-1
MODS top level elements



```

<mods xmlns="http://www.loc.gov/mods/">
  <titleInfo>
    <title>Germany by bike : 20 tours geared for discovery /</title>
  </titleInfo>
  <name type="personal">
    <namePart type="given">Nadine</namePart>
    <namePart type="family">Slavinski</namePart>
    <namePart type="date">1968-</namePart>
    <role><roleTerm type="text">creator</roleTerm></role>
  </name>
  <typeOfResource>text</typeOfResource>
  <originInfo>
    <place><placeTerm type="code" authority="marc">wau</placeTerm>
      <placeTerm type="text"> Seattle, Wash. :</placeTerm>
    </place>
    <publisher>Mountaineers,</publisher>
    <dateIssued>c1994</dateIssued>
    <issuance>monographic</issuance>
  </originInfo>
  <language>
    <languageTerm type="code" authority="iso639-2b">eng</languageTerm>
  </language>
  <physicalDescription>
    <extent>238 p. : ill., maps ; 22 cm.</extent>
  </physicalDescription>

```

Metadados para objetos culturais e recursos visuais

Metadados para objetos culturais e recursos visuais

- Coleções digitais e os recursos visuais têm sido o foco de muitas instituições, pois possibilitam o acesso a milhares de itens considerados tesouros culturais.
- Dois nomes influentes devem ser reconhecidos ao introduzir padrões: o *Getty Research Institute* e a *Visual Resources Association (VRA)*.
- Os padrões para estrutura de dados e semântica - principalmente as categorias para a descrição de obras de arte (CDWA) e as categorias principais da associação de recursos visuais (Núcleo VRA), formaram a base das práticas de metadados nas comunidades de museus e artes visuais.

Categories for the Description of Works of Art (CDWA)

- O CDWA foi editado por Baca e Harpring e publicado pela primeira vez em 2006.
- A CDWA identifica o vocabulário que torna as informações distribuídas em diversos sistemas mais compatíveis e acessíveis.
- CDWA utiliza-se dos tesouros desenvolvidos pela *The Getty Institute: Art & Architecture Thesaurus*, *Cultural Objects Name Authority (CONA)*, *Getty Thesaurus of Geographic Names (TGN)* e *Union List of Artist Names (ULAN)*.
- Essa cobertura abrangente requer uma estrutura complexa que resultou em 31 categorias amplas, com cerca de 540 categorias e subcategorias. Um pequeno subconjunto das categorias é considerado essencial, representando as informações mínimas necessárias.

CDWA Categories

OBJECT, ARCHITECTURE, OR GROUP

- Object/Work*
- Classification*
- Titles or Names*
- Creation*
- Groups
- Styles/Periods/Movements
- Measurements*
- Materials and Techniques*
- Inscriptions/Marks
- State
- Edition
- Facture
- Physical Description
- Condition/Arrangement
- Conservation/Treatment History
- Subject Matter*
- Context
- Descriptive Note
- Critical Responses
- Related Works
- Current Location*
- Copyright/Restrictions
- Ownership/Collecting History
- Exhibition/Loan History
- Cataloging History
- Related Visual Documentation
- Related Textual References*

AUTHORITIES/ VOCABULARY CONTROL

- Person/Corporate Body Authority*
- Place/Location Authority*
- Generic Concept Authority*
- Subject Authority*

Legend

* = Required

CCO – Cataloging Cultural Objects

- Embora o CCO seja um padrão de conteúdo que fornece diretrizes para selecionar, ordenar e formatar dados usados para preencher elementos em um registro de catálogo, ele especifica informações descritivas importantes, necessárias para criar um registro para um trabalho e imagem.
- Seus 116 elementos e subelementos são derivados dos metadados CDWA.
- Recomenda que o registro mínimo inclua a maioria dos itens (se não um conjunto de elementos principais) que inclui os elementos de metadados mais essenciais.
- O registro mínimo deve conter valores de dados para todos os elementos essenciais, sempre que possível.

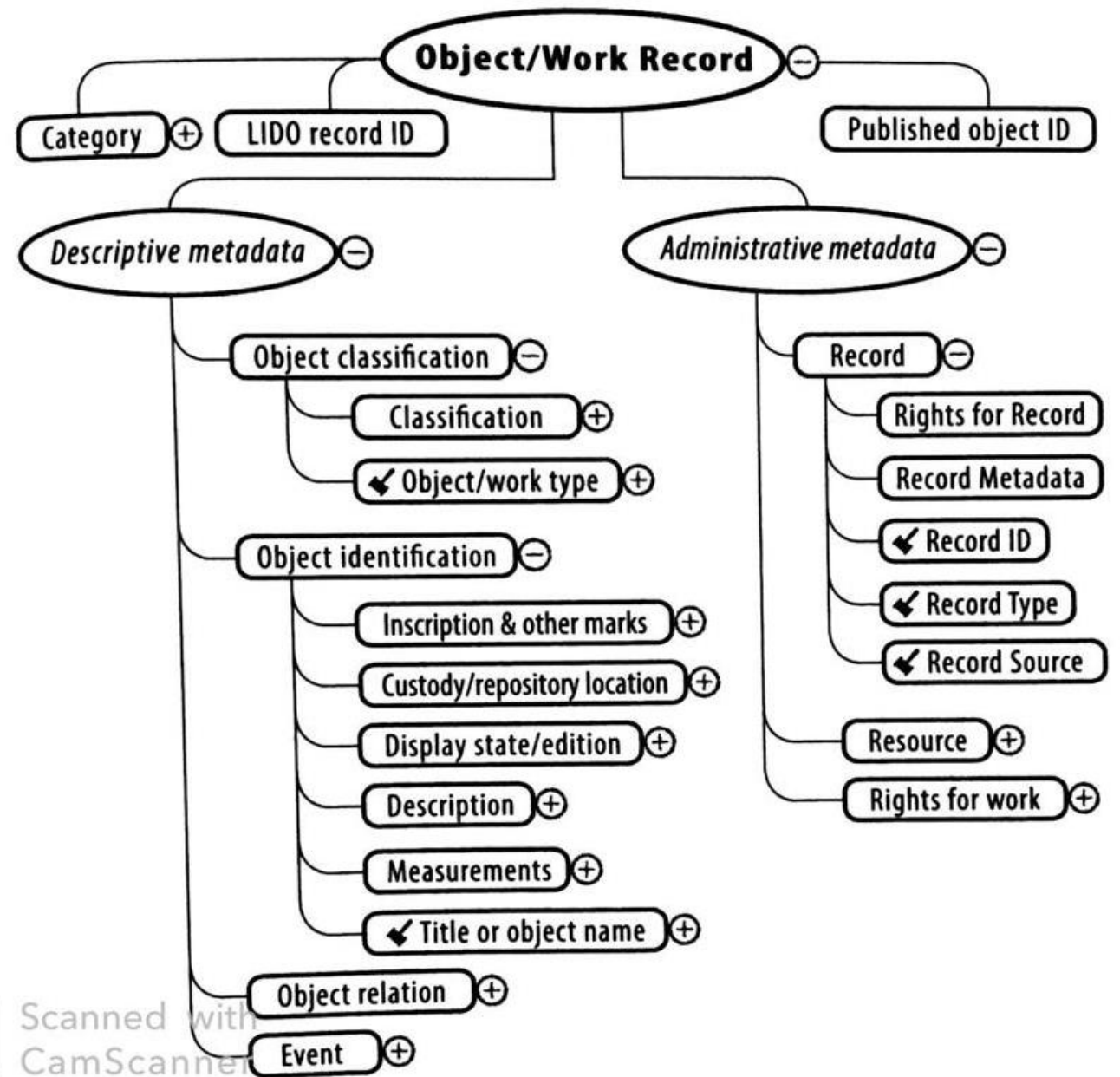
CCO chapter		Element	Y/N	Authority
1. Object Naming	Work Type	y	Y	Concept authority
	Title	y	N	
	Title Type	n	Controlled list	
	Language	n	Controlled list	
2. Creator Information	Resource	n	Y	Source authority
	Creator display	y	N	
	Controlled Creator	y	Y	Name authority
	Role	y	Y	Concept authority
	Creator Extent	n	Controlled list	
3. Physical Characteristics	Attribution Qualifier	n	Controlled list	
	Measurements display	y	[Sub-elements of some elements are controlled by controlled lists]	[Some sub-elements are linked to authorities]
	Materials and Techniques display	y		
	Edition display	n		
	State display	n		
	Inscriptions	n		
	Facture	n		
	Physical Description	n		
	Condition and Examination History	n		
Conservation and Treatment History	n			
4. Stylistic, Cultural and Chronological Information	Style	n	Y	Concept authority
	Culture	n	Y	Concept authority
	Display Date	y	N	
5. Location and Geography	Current Location display	y	[Some sub-elements are controlled by controlled lists]	[Some sub-elements are linked to authorities]
	Creation Location display	n		
	Discovery Location display	n		
	Former Location display	n		

CDWA Lite e LIDO

- Representa uma versão “mais leve” do CDWA, que contém os elementos principais a serem usados para descrever objetos culturais.
- Ele é entregue como um esquema XML e é baseado nas categorias, estrutura e diretrizes necessárias emitidas pela CDWA.
- O esquema CDWA Lite foi substituído pelo LIDO, o esquema *Lightweight Information Describing Objects*, em 2010.
- O LIDO pode ser usado para descrever todos os tipos de objetos com o suporte a um ambiente multilíngue.
- Ele define quatorze grupos de informações, obrigatórias.

CDWA Lite e LIDO

- De acordo com o documento 1.0, as informações em um registro LIDO são organizadas em dois grupos: descritivos e administrativos - com cada grupo contendo algumas subdivisões adicionais



VRA Core

- O VRA 2.0 foi lançado em 1997, expandindo o conceito para incluir objetos efêmeros e não artísticos e substituiu a palavra objeto pela palavra trabalho. Também enfatiza a diferença entre trabalho e imagem.
- O VRA Core Versão 3.0 combinou dois conjuntos de elementos separados para trabalhos e imagens da Versão 2.0 e se tornou um único conjunto de elementos.
- A diferença entre a versão 3.0 do VRA e o CDWA não é apenas o tamanho dos elementos (dezessete no total), mas também a estrutura, que é semelhante ao Dublin Core, caracterizada por sua estrutura plana mais "qualificadores".

VRA Core

- O lançamento do VRA Core 4.0, em 2007, trouxe mudanças dramáticas em termos de sintaxe.
- As alterações nesta nova versão foram feitas para tornar o VRA Core compatível com XML.
- Toda a estrutura aninhada do VRA Core 4.0 partiu do DC (que possui uma estrutura plana) e passou a ser consistente com o CDWA Lite.

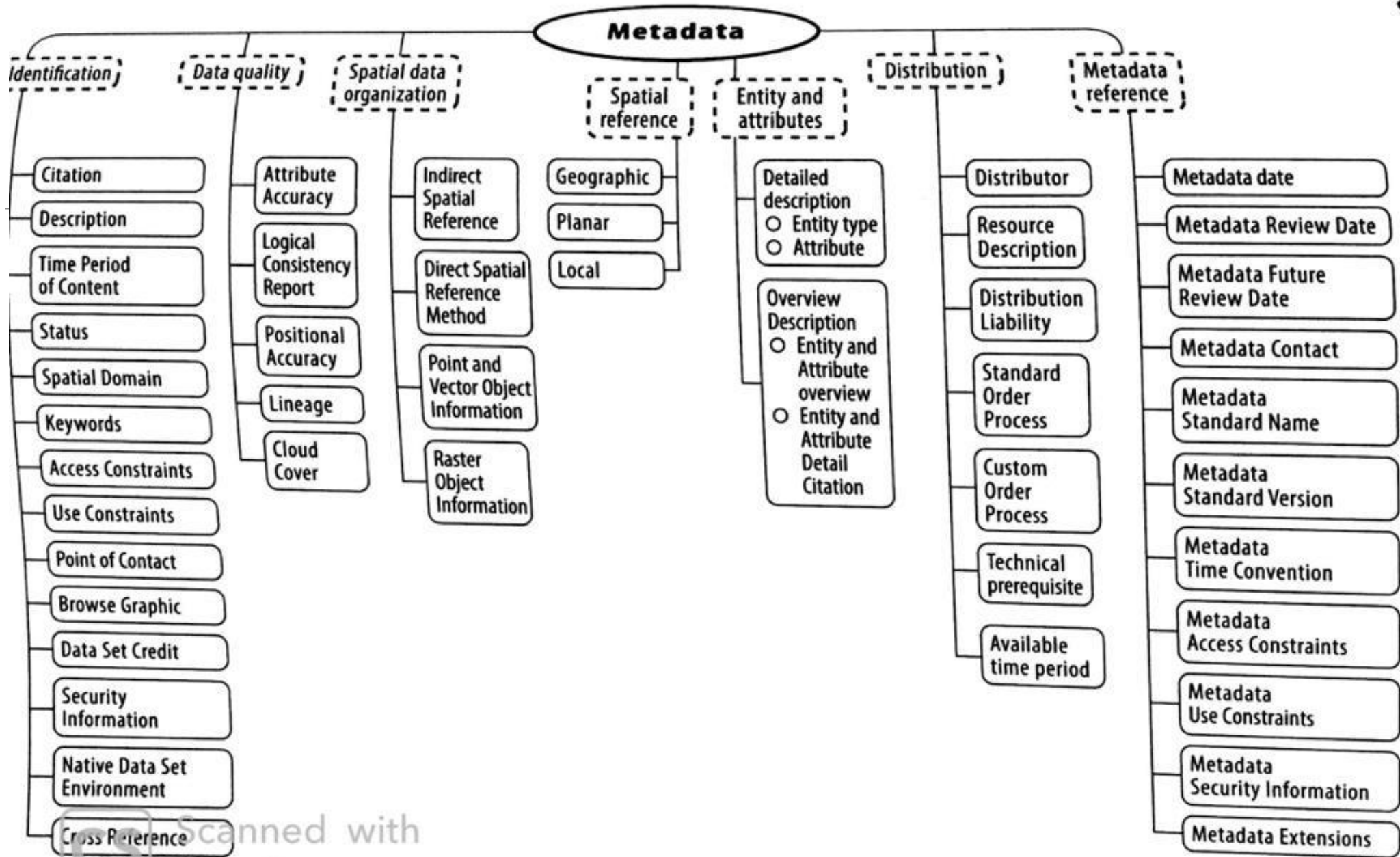
Metadado para Dados de Pesquisa

Metadado para Dados de Pesquisa

- Os metadados para dados de pesquisa são todas as informações, além dos próprios dados brutos, que um usuário precisaria fazer uso completo e preciso para análises científicas subsequentes.
- O verdadeiro desafio não é armazenar inúmeros petabytes (1 milhão de gigabytes), mas recuperar e analisar essas informações em tempo real.
- A criação de metadados, portanto, torna-se ainda mais crítica, a fim de permitir descoberta, compartilhamento, gerenciamento e análise de dados eficientes.

Metadados para dados geoespaciais

- Nos Estados Unidos da América, o Comitê Federal de Dados Geográficos (FGDC) desenvolveu o primeiro padrão de metadados, o Padrão de Conteúdo para o Digital Geoespacial Metadados (CSDGM), em 1994, divulgou algum compartilhamento entre agências e disseminou a segunda versão em 1998.
- O CSDGM consiste em 334 elementos, divididos em 7 categorias: identificação, qualidade dos dados, organização de dados espaciais, referência espacial, entidade e atributos, distribuição e informações de referência de metadados.
- Requer conhecimento aprofundado do tema para inserção correta dos dados.

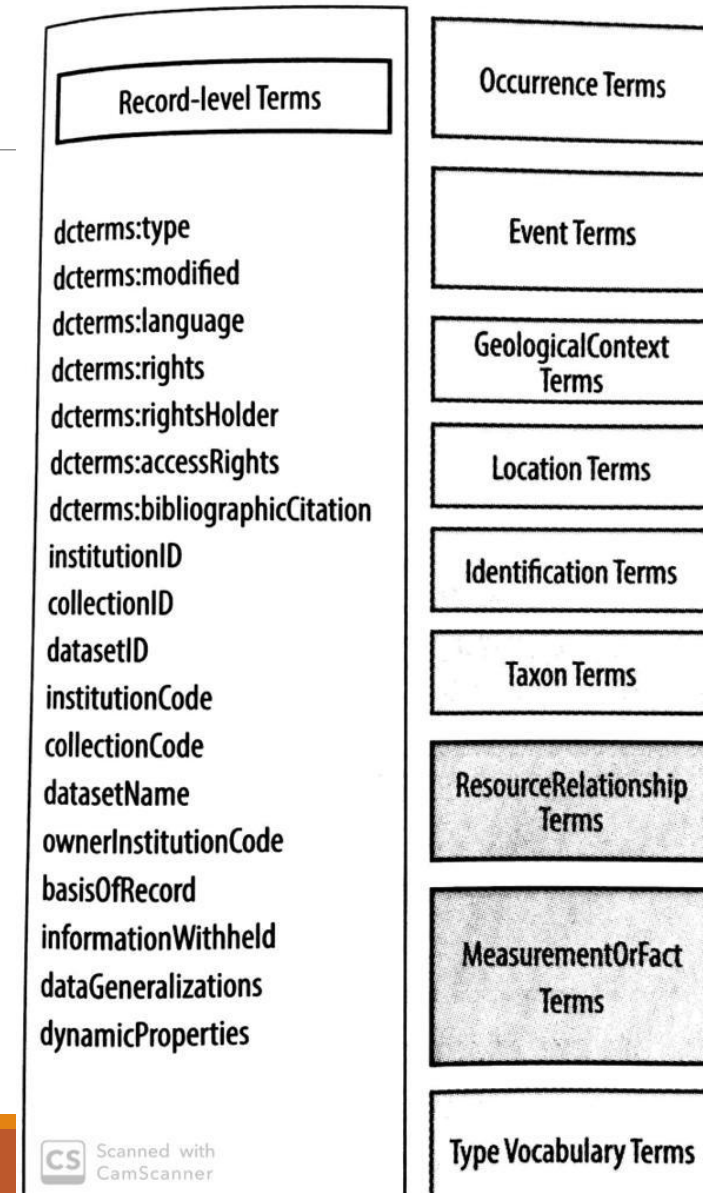


Metadado para dados sobre ecologia e biodiversidade

- O estudo do ambiente natural e dos seres vivos gerou espécimes, amostras e informações relacionadas, que podem ser alojadas em museus ou centros de pesquisa e documentadas.
- O objetivo dos registros de história natural de metadados, que geralmente são padrões para dados de biodiversidade e ecologia, é facilitar o acesso e o compartilhamento de dados "fornecendo um vocabulário básico bem definido em uma estrutura flexível para minimizar as barreiras à adoção e maximizar a reutilização" (Wieczorek et 2009).
- Existem dois padrões básicos de metadados recomendados para dados de biodiversidade, ambos desenvolvidos pelo Grupo de Trabalho de Bancos de Dados Taxonômicos (TDWG)

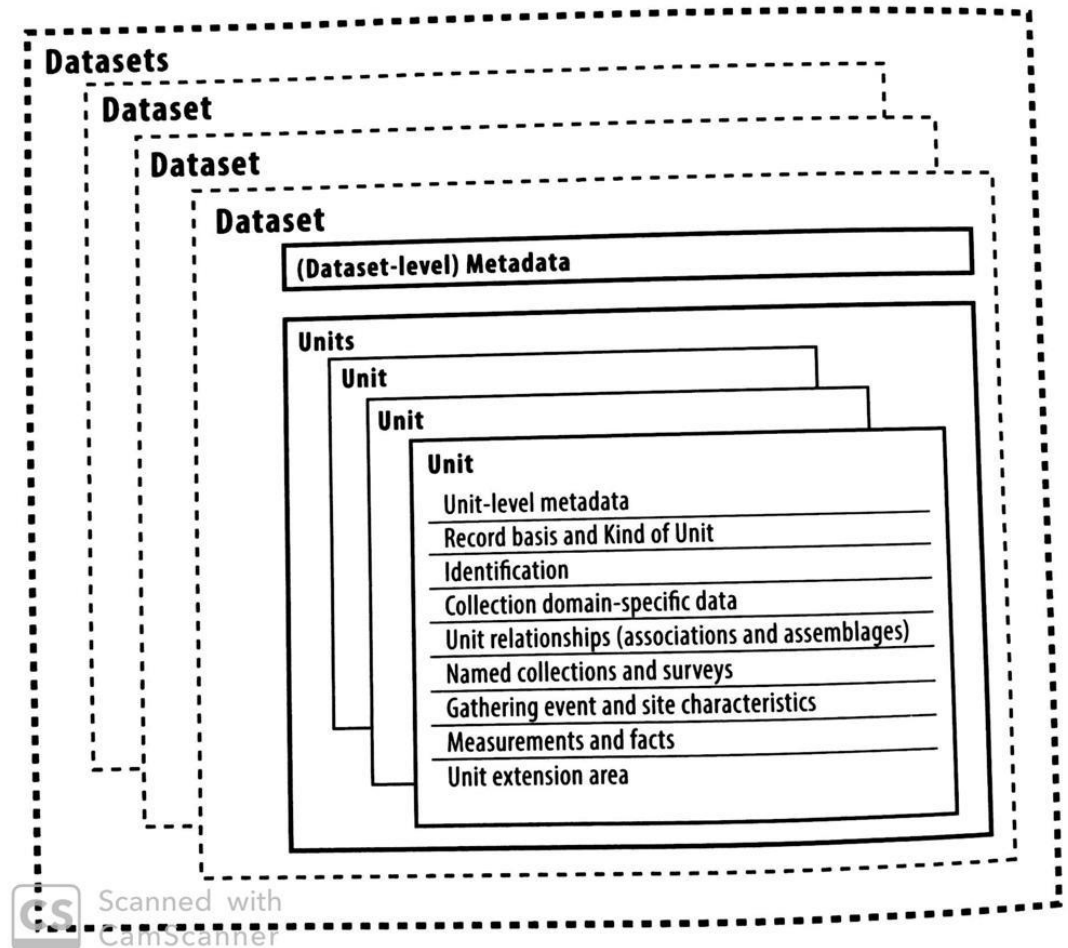
Metadado para dados sobre ecologia e biodiversidade

- O Darwin Core possui uma estrutura simples e plana que agrupa todos os elementos em várias categorias.



Metadado para dados sobre ecologia e biodiversidade

- Outro padrão no domínio da biodiversidade é o Acesso aos Dados (ABCD), que foi projetado para fornecer um conjunto padrão de elementos para o registro de informações sobre "espécimes vivos e preservados, juntamente com observações em campo de Coleções Biológicas que não produziram espécimes de comprovantes" (Thomson et al. 2006).



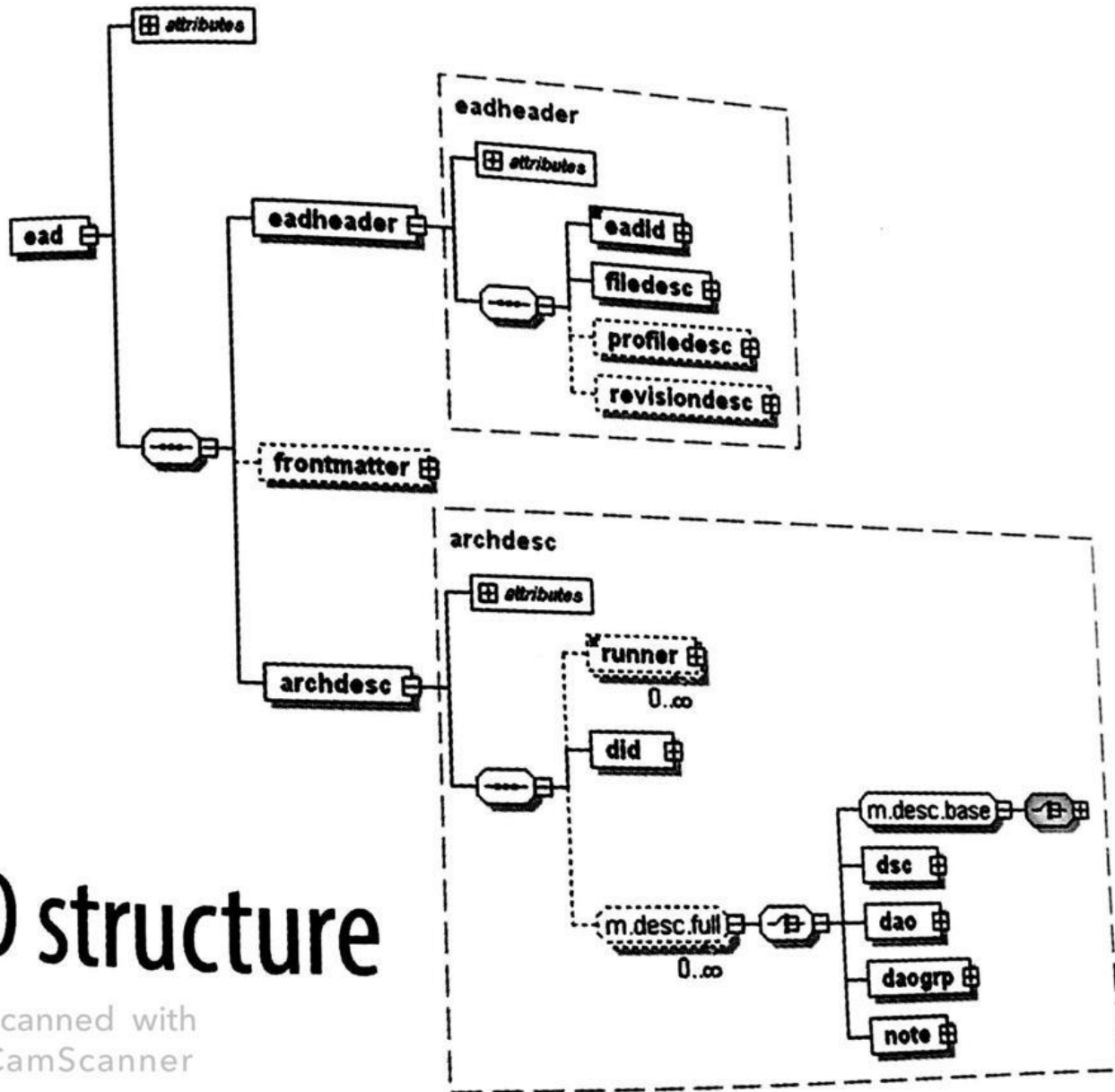
Metadados para dados de pesquisa em ciências sociais

- O projeto *Data Documentation Initiative* (DDI) foi iniciado em 1995, com o objetivo de estabelecer um padrão internacional para descrever dados de ciências sociais, comportamentais e econômicas.
- A versão mais recente do DDI (versão 3.2) é baseada no ciclo de vida da pesquisa conceitualizado por Charles Humphrey (2006), que delinea um processo de pesquisa desde os conceitos iniciais até a coleta de dados e o processamento até a distribuição, descoberta e análise de dados.
- A especificação DDI envolve linhas de desenvolvimento: DDI-Lifecycle e DDI-Codebook.
 - O DDI-Lifecycle OS foi projetado para "documentar e gerenciar dados em todo o ciclo de vida, da conceituação à publicação e análise de dados e além" (DDI Alliance 2009),
 - O DDI-Codebook é uma versão leve, o padrão direcionado para a documentação de dados de pesquisas. O DDI 3.2 foi projetado com uma abordagem modular.

Metadados para Arquivos

Metadados para Arquivos

- Os metadados de arquivos funcionam como auxiliares de busca no domínio do arquivo, priorizando grupos de materiais de arquivos.
- No discurso da comunidade de arquivos, um grupo de arquivos (ou grupo de registros) é um agregado de itens que compartilham uma proveniência comum.
- Um grupo de arquivamento pode conter vários números e tipos de itens. Fisicamente, esses itens podem ser armazenados em pastas e caixas e devem ser classificados por série.
- Encontrar recursos é fundamentalmente uma ferramenta para facilitar a descoberta de informações nessas coleções.



EAD structure

CS Scanned with CamScanner

Metadados para publicações e comunicados da imprensa

Metadados para publicações e comunicados da imprensa

- Com o rápido crescimento da publicação eletrônica, a indústria editorial e outros players, inclusive os editores de publicações próprias, começaram a implementar padrões de metadados para descrever produtos de maneira completa e precisa e para compartilhar com eficiência esses dados com parceiros comerciais e consumidores.
- ONIX e EPUB.

ONIX (ONline Information Exchange)

- ED EUR (editora) é um grupo internacional para o comércio eletrônico dos setores de livros e seriados, e o desenvolvedor da troca de informações on-line ONIX, que se trata de conjunto de padrões que o publicador usa para distribuir informações eletrônicas sobre seus produtos a livrarias atacadistas, varejistas eletrônicos e de varejo, outros editores e qualquer outra pessoa envolvida na compra e venda de livros.
- O ONIX é uma família de formatos XML para a comunicação de metadados, usando elementos de dados comuns, "comp. Posites" e listas de códigos.
- O padrão ONIX define uma lista de campos de dados sobre publicação e descreve como enviar esses dados em uma "mensagem ONIX".
- Essa família sempre crescente já lançou vários padrões usados internacionalmente e estão sendo desenvolvidas atividades adicionais, como a incorporação de termos de licenciamento no ONIX.
- A versão 3.0 da revisão 1 (3.0.1), lançada como *ONIX DOI Registration Formats*, é uma coleção de formatos de registro DOI, permitindo que editores e outros comuniquem os metadados exigidos por uma agência de registro para registrar a atribuição DOI.

EPUB

- EPUB é um padrão de e-book aberto iniciado pelo *International Digital Publishing Forum* (IDPF) para o padrão de formato de distribuição e intercâmbio para publicações e documentos digitais, lançado em 2007.
- O EPUB define meios de representar, empacotar e codificar conteúdo da Web estruturado e semanticamente aprimorado - incluindo HTML, CSS, SVG, imagens e outros recursos - para distribuição em um formato de arquivo único.
- Existem abordagens notáveis nesta especificação, por exemplo: além dos elementos gerais de metadados que descrevem a publicação, ele também fornece os meios para anexar registros bibliográficos completos a uma publicação usando o elemento 'link'.
- Permite que documentos XHTML anotem a marcação de documentos com metadados avançados. Isso corresponde às tendências da marcação semântica para fins efetivos de processamento e acessibilidade.

Metadado para objetos multimídia

MPEG

- Produzido pelo *International Press Communications Council (IPTC)*, que define padrões para metadados de notícias desde 1979 (IPTC 2010).
- Em princípio, a especificação de metadados do IPTC Photo não inclui metadados técnicos (por exemplo, capturados por câmeras digitais) porque estes são abordados por outros padrões existentes.
- O padrão também estendeu a seção Imagem para abranger os metadados a serem usados para trabalhos de arte ou imagem granular e adicional (pessoa, organização, local, objetos, além de para modelos humanos em uma imagem liberada pelo modelo).

ID3v2

- É um contêiner de metadados frequentemente usado para arquivos de áudio MP3, suportado ativamente por software como iTunes e Windows Media Player.
- Geralmente contém o nome do artista, Título da música, ano e gênero de cada arquivo de áudio.
- Não se limita inteiramente ao áudio musical; outros tipos de arquivos de áudio também são suportados.
- O ID3V2, tratado como um padrão informal, definiu um número significativo de elementos. Existem elementos com texto, cerca de trinta quadros adicionais para cobrir comentários, letras / texto e vários identificadores e códigos.

Metadados para preservação

OAIS

- O início do Marco de Referência OAIS foi concebido para apoiar o desenvolvimento de padrões para a preservação a longo prazo de informações digitais obtidas a partir de observações de ambientes terrestres e extraterrestres.
- O Painel 2 do Comitê Consultivo da NASA para Sistemas de Dados Espaciais (CCSDS) coordenou o desenvolvimento desses padrões.
- "Um OAIS é um arquivo, constituído por uma organização de pessoas e sistemas que aceitaram a responsabilidade de preservar informações e disponibilizá-las.
- Os resultados do OAIS incluem o estabelecimento de uma estrutura comum de termos e conceitos, a identificação das funções básicas de um sistema de informações de arquivo e as definições de um modelo de informação.

PREMIS

- Um grupo de trabalho internacional estabelecido pelo Grupo de Bibliotecas de Pesquisa (RLG) primeiro descreveu os tipos de informações que devem ser associadas a um objeto digital arquivado em 2001-2002.
- O novo projeto, A Metadata Framework para apoiar a preservação de objetos digitais propôs uma lista de elementos de metadados de protótipos. Com base nisso, o Grupo de Trabalho PREMIS, realizou uma pesquisa mundial das implementações de repositórios de preservação de 2003-2004 (OCLC / RLG 2004).
- O mais atual, o PREMIS 3.0, possui cinco tipos de entidades: Objeto, Ambiente, Evento, Agente e Declaração de Direitos. Deles há subentidades, por exemplo, a Entidade Intelectual é uma subentidade em Objeto.

Metadados para descrição de agentes

VCard

- A maioria dos tipos de propriedades do vCard é representada simplesmente pelo nome e valor do tipo de propriedade, conforme definido na especificação.
- Eles incluem: FN, NICKNAME, BDAY, GEO, TITLE, PAPEL, CATEGORIES, SOURCE, NOTE, PRODID, TEL, EMAIL e outros.
- Foi mapeado para o OWL em um documento do W3C Note, vCard Ontology - para descrever Pessoas e Organizações (Iannella e McKinney 2014).

```
<vcard:Individual rdf:about="http://example.com/me/johndoe">  
  <vCard:fn> John Doe </vCard:fn>  
  <vCard:bday> 1975-01-01 </vCard:bday>  
  <vCard:title> Metadata Librarian II</vCard:title>  
  <vCard:role> Coordinator </vCard:role>  
</vcard:Individual>
```


FOAF

- FOAF se tornou um dos vocabulários de metadados mais usados na nuvem *Linked Open Data*.
- Ele evoluiu gradualmente desde a sua criação em meados dos anos 2000 e tem sido amplamente aceito pelos membros da cultura da Internet, especialmente em relação ao desenvolvimento da Web Semântica.
- O desenvolvimento e manutenção do FOAF apresenta uma preocupação por sua viabilidade a longo prazo.
- Em 2011, para lidar com essa preocupação, a FOAF e a DCMI chegaram a um acordo de que os dois proprietários colaborarão.

FOAF

- Serviços de redes sociais na Internet, como Orkut, Facebook, Friendster, Myspace e o Google+ têm atraído a manutenção de relacionamentos existentes e modelos de descrição simples para permitir que os usuários criem seus perfis, descrevam seus dados sociais, profissionais e pessoais, façam upload de fotos e vídeos e definam a opção para quais grupos de pessoas poderão ver essas informações.

FOAF Core

- - [Agent](#)
 - [Person](#)
 - [name](#)
 - [title](#)
 - [img](#)
 - [depiction](#) ([depicts](#))
 - [familyName](#)
 - [givenName](#)
 - [knows](#)
 - [based_near](#)
 - [age](#)
 - [made](#) ([maker](#))
 - [primaryTopic](#) ([primaryTopicOf](#))
- - [Project](#)
 - [Organization](#)
 - [Group](#)
 - [member](#)
- - [Document](#)
 - [Image](#)

Social Web

- [nick](#)
- [mbox](#)
- [homepage](#)
- [weblog](#)
- [openid](#)
- [jabberID](#)
- [mbox_sha1sum](#)
- [interest](#)
- [topic_interest](#)
- [topic](#) ([page](#))
- [workplaceHomepage](#)
- [workInfoHomepage](#)
- [schoolHomepage](#)
- [publications](#)
- [currentProject](#)
- [pastProject](#)
- [account](#)
- [OnlineAccount](#)
- [accountName](#)
- [accountServiceHomepage](#)
- [PersonalProfileDocument](#)
- [tipjar](#)
- [sha1](#)
- [thumbnail](#)
- [logo](#)

Referência Bibliográfica

ZENG, Marcia Lei e QIN, Jian. Metadata Standards. In: _____. **Metadata**. 2ª Edição. 555p. Chicago: Neal-Schuman/American Library Association. 2016. Capítulo 10, p. 401-484.