

Metadados

Seminário de Conclusão de Curso

Eduardo Caetano Barbosa

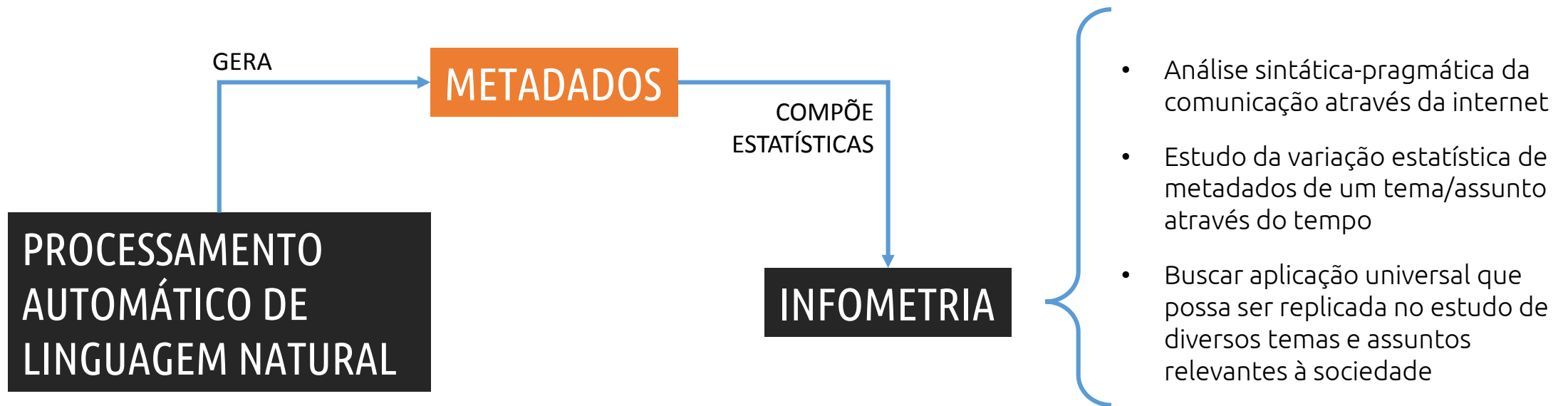
9 de junho de 2015

CBD 5140 | Produção e Recuperação de Informação: Teorias e Métodos

Professoras Giovana Deliberali Maimone e Nair Yumiko Kobashi

Porque Metadados?

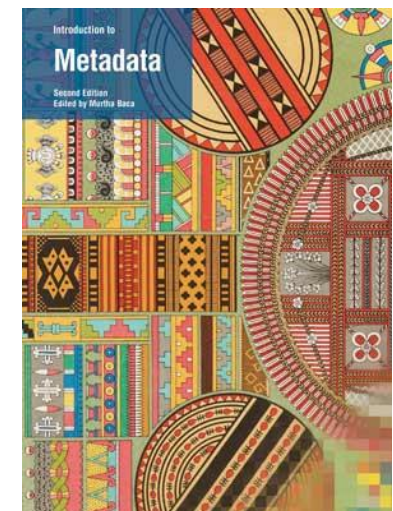
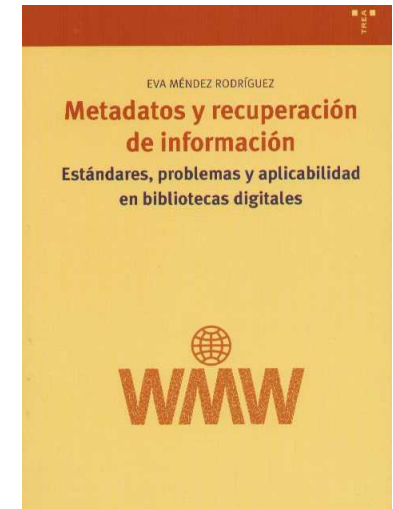
Um dos conceitos centrais a ser explorado como parte de projeto de pesquisa, que busca estudar métodos e práticas de Infometria gerada através do processamento automático de linguagem natural de conteúdos (objetos informacionais), disponíveis e acessíveis publicamente na internet (veículos de imprensa tradicionais, nova imprensa, blogs, redes sociais e acadêmicas)

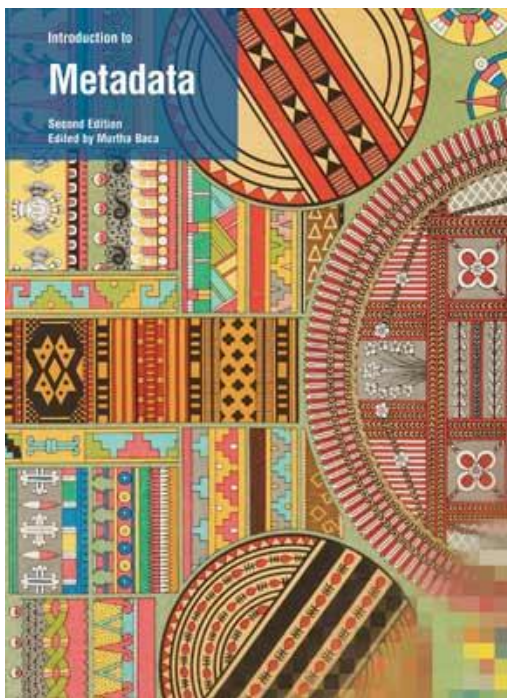


O seminário

Este seminário visa explorar a bibliografia do curso sobre o conceito de **Metadados**, previamente discutido em aula sob a ótica da autora Eva Méndez, com base no primeiro capítulo de seu livro *Metadatos y recuperación de información*.

Meu objeto de estudo, como contraponto, é a introdução sobre o conceito de Metadados feito por Anne Gilliland, no capítulo introdutório do livro *Introduction to Metadata: Pathways to Digital Information*, organizado por Murtha Baca, que faz parte da bibliografia indicada na disciplina, e referenciado também como fonte usada no livro de Eva Méndez.





Sobre o livro e autora

Introduction to Metadata: Pathways to Digital Information

Lançado em 2008, este livro é a segunda edição de uma revisão feita digitalmente em 2000, originalmente escrita em 1998. Autores que colaboraram anteriormente revisaram e atualizaram seus textos, entre eles Gilliland, que abre o livro.

Murtha Baca, a organizadora, encabeça o programa de acesso digital à História da Arte, no instituto Getty.

[Download do livro \(PDF\)](#)



Anne J Gilliland

- Professora e diretora da especialização em estudos de arquivo (Archival Studies Specialization), no departamento de Estudos da Informação UCLA;
- Afiliada ao UCLA's Center for Digital Humanities
- Diretora do Archival Education and Research Initiative (AERI)

Setting the Stage

- Metadados, literalmente “dados sobre dados” se tornou largamente usado, porém constantemente pouco especificado. O termo tem compreensão diversa entre as comunidades profissionais que desenham, criam, descrevem, preservam e usam sistemas e recursos de informação.
- Criamos e interagimos com metadados cada vez de forma mais digital
- Tradicionalmente a criação e uso de metadados foi de responsabilidade de profissionais da informação engajados na catalogação, classificação e indexação, mas suas aplicação nos meios digitais e a internet aumentam suas interpretações e uso para muito além do que apenas os profissionais da informação.
- Tem sido cada vez mais importante que, além dos profissionais de informação, criadores e utilizadores de conteúdo digital entendam os principais papéis dos diferentes tipos de metadados em assegurar acessibilidade, autoria, interoperação, granulação e preservação do patrimônio cultural e seus sistemas de registro.

Compreensão mais ampla de metadados é a soma total do que se pode dizer sobre qualquer Objeto de Informação (*Information Object*)

OBJETO DE INFORMAÇÃO

- Item ou grupo de itens digitais, não importando tipo ou formato, que pode ser dirigido ou manipulado como um objeto único por um computador (imagens digitais de obras de arte, PDFs, etc)
- Pode ser usado para “substitutos” digitais de documentos (DLO) e para descrever registros relacionados a objetos e/ou coleções (catálogo de registros, *finding aids*, etc)

3 características principais que devem (deveriam) refletir nos metadados:

Conteúdo - relativo ao que o objeto contém ou fala sobre. Intrínseco ao objeto de informação;

Contexto - indica os aspectos de quem, o quê, porquê, onde e como são associados ao objeto. Extrínsecos ao objeto

Estrutura - relativo ao conjunto formal de associações por dentro de, ou entre objetos individuais de informação. Podem ser intrínsecos, extrínsecos ou ambos.

Table 1. **A Typology of Data Standards**

Type	Examples	Esquemas transversais de estrutura
Data <i>structure</i> standards (metadata element sets, schemas). These are “categories” or “containers” of data that make up a record or other information object.	The set of MARC (Machine-Readable Cataloging format) fields, Encoded Archival Description (EAD), Dublin Core Metadata Element Set (DCMES), Categories for the Description of Works of Art (CDWA), VRA Core Categories	Esquemas baseados em padrões de valores
Data <i>value</i> standards (controlled vocabularies, thesauri, controlled lists). These are the terms, names, and other values that are used to populate data structure standards or metadata element sets.	Library of Congress Subject Headings (LCSH), Library of Congress Name Authority File (LCNAF), LC Thesaurus for Graphic Materials (TGM), Medical Subject Headings (MeSH), Art & Architecture Thesaurus (AAT), Union List of Artist Names (ULAN), Getty Thesaurus of Geographic Names (TGN), ICONCLASS	Padrões de conteúdo / preenchimento de campos
Data <i>content</i> standards (cataloging rules and codes). These are guidelines for the format and syntax of the data values that are used to populate metadata elements.	Anglo-American Cataloguing Rules (AACR), Resource Description and Access (RDA), International Standard Bibliographic Description (ISBD), Cataloging Cultural Objects (CCO), Describing Archives: A Content Standard (DACS)	Padrões que proporcionam interoperabilidade
Data <i>format/technical interchange</i> standards (metadata standards expressed in machine-readable form). This type of standard is often a manifestation of a particular data structure standard (type 1 above), encoded or marked up for machine processing.	MARC21, MARCXML, EAD XML DTD, METS, MODS, CDWA Lite XML schema, Simple Dublin Core XML schema, Qualified Dublin Core XML schema, VRA Core 4.0 XML schema	

Note: This table is based on the typology of data standards articulated by Karim Boughida, “CDWA Lite for Cataloging Cultural Objects (CCO): A New XML Schema for the Cultural Heritage Community,” in *Humanities, Computers and Cultural Heritage: Proceedings of the XVI International Conference of the Association for History and Computing*: 14–17 (September 2005) (Amsterdam: Royal Netherlands Academy of Arts and Sciences, 2005). Available at <http://www.knaw.nl/publicaties/pdf/20051064.pdf>.

Metadados de Biblioteca

Fornece acesso intelectual e físico para coleções, incluindo índices, abstracts e registros bibliográficos. Criados tanto por humanos como por máquina (MARC, AAT, ...)

Metadados de Arquivo e Museus

Tradicionalmente usa metadados de contexto: Identificar e preservar valores de evidência de registros e artefatos. Archival description inclui adesão de registros, facilitadores de busca (finding aids) e registro de catálogos (AMC, MARC21, EAD, ...)

Inicialmente acessíveis apenas especificamente no local físico, iniciativas de interoperabilidade desses arquivos começa a partir dos anos 90. Na comunidade dos museus, esta integração é um pouco mais lenta e restrita a grandes instituições com diversas unidades e acervos.

[MOAC](#): Museums and the Online Archive of California (sucesso limitado).

“... there is no single metadata standard that is adequate in describing all types of collections and materials”

Mapeamento de metadados em comunidades específicas de conhecimento, que almejam a Interoperabilidade .

Porém a preocupação com estrutura dos objetos de informação no desenvolvimento de metadados é menos evidente (mesmo que estrutura sempre tenha sido importante antes mesmo do meio digital).

Quanto mais um objeto de informação é estruturado, mais essa estrutura pode ser explorada para recuperação (buscas), manipulação e inter-relação com outros objetos de informação.

O papel da estrutura tem se tornado cada vez mais relevante com os avanços da capacidade e sofisticação do processamento de computadores.

“In the online world, multiple descriptive relationships between objects can be supported simultaneous, and some of this may more effectively support **new types of users and uses in a environment that is not mediated by a reference archivist**”

Em um ambiente em rede onde um usuário tem acesso não mediado a objetos de informação, metadata:

- Certifica a autenticidade e a 'completude' do conteúdo
- Determina e documenta o contexto do conteúdo
- Identifica e explora a relação estrutural que existe dentro e entre objetos de informação
- Abre uma gama de pontos de acesso intelectual para um crescente e diverso tipo de usuário
- Provê algumas das informações que um profissional da informação pode ter especificado usando métodos tradicionais, referências pessoais e settings de pesquisa.

Porém uma conceituação mais inclusiva de metadados é necessária para incorporar em sistemas de informação. Repositórios específicos criam metadados relacionados a:

- Administração
- Acessibilidade
- Preservação e uso de coleções

Exemplos:

- Registros de aquisição
- Catálogos de exposição
- Regras de licenciamento
- Metadados educacionais

Aplicados fora dos repositórios originais, metadados tem um escopo ainda mais abrangente de interpretações:

- Internet –metatags de html
- Catalogação de imagens digitais (informações do arquivo de imagem)
- Arquivista de ciências sociais (dados intrínsecos à uma pesquisa)
- Arquivista de áudio digital
- ...

Em todas as diversas leituras, metadados não só descrevem o objeto de informação, também documenta como aquele objeto se comporta, sua função e uso e relações com outros objetos e como foi e deve ser gerido ao longo do tempo.

Teorias e métodos diferem consideravelmente devido às diferenças específicas de campos profissionais e culturais

Metadados criados por usuários - Folksonomia

- (2008) Ganha relevância e discussão na web
- Blogs, Flickr, Youtube, Social Bookmarks
- usuários criam, compartilham e copiam metadados
- Criadas em comunidades de interesse de nicho, de forma a identificar, recuperar, categorizar e promover conteúdo da Web
- Baixo custo, porém baixíssimo controle de qualidade

Nota: Embora o termo 'folksonomia' tenha saído de cena de forma discreta, desdobramentos diretos destas ações estão presentes no constante uso de Hashtags pelas redes sociais. Estas em si não poderiam (não deveriam) ser consideradas metadados como conhecemos, pois estão envoltas à modismos e neologismos ligados a classificações efêmeras.

Table 2. **Different Types of Metadata and Their Functions**

Type	Definition	Examples
Administrative	Metadata used in managing and administering collections and information resources	<ul style="list-style-type: none"> • Acquisition information • Rights and reproduction tracking • Documentation of legal access requirements • Location information • Selection criteria for digitization
Descriptive	Metadata used to identify and describe collections and related information resources	<ul style="list-style-type: none"> • Cataloging records • Finding aids • Differentiations between versions • Specialized indexes • Curatorial information • Hyperlinked relationships between resources • Annotations by creators and users
Preservation	Metadata related to the preservation management of collections and information resources	<ul style="list-style-type: none"> • Documentation of physical condition of resources • Documentation of actions taken to preserve physical and digital versions of resources, e.g., data refreshing and migration • Documentation of any changes occurring during digitization or preservation
Technical	Metadata related to how a system functions or metadata behaves	<ul style="list-style-type: none"> • Hardware and software documentation • Technical digitization information, e.g., formats, compression ratios, scaling routines • Tracking of system response times • Authentication and security data, e.g., encryption keys, passwords
Use	Metadata related to the level and type of use of collections and information resources	<ul style="list-style-type: none"> • Circulation records • Physical and digital exhibition records • Use and user tracking • Content reuse and multiversioning information • Search logs • Rights metadata

Table 3. **Attributes and Characteristics of Metadata**

Attribute	Characteristics	Examples
Source of metadata	Internal metadata generated by the creating agent for an information object at the time when it is first created or digitized	<ul style="list-style-type: none"> • File names and header information • Directory structures • File format and compression scheme
	Metadata intrinsic to an item or work	<ul style="list-style-type: none"> • A title or other inscription added to an art work by its creator • A title or subtitle on the title page of a manuscript or printed book
	External metadata relating to an original item or information object, that is created later, often by someone other than the original creator	<ul style="list-style-type: none"> • URLs and other digital statements of provenance • "Tracked changes" • Registrarial and cataloging records • Rights and other legal information
Method of metadata creation	Automatic metadata generated by a computer	<ul style="list-style-type: none"> • Keyword indexes • User transaction logs • Audit trails
	Manual metadata created by humans	<ul style="list-style-type: none"> • Descriptive metadata such as catalog records, finding aids, and specialized indexes
Nature of metadata	Nonexpert metadata created by persons who are neither subject specialists nor information professionals, e.g., the original creator of the information object or a folksonomist	<ul style="list-style-type: none"> • META tags created for a personal Web page • Personal filing systems • Folksonomies
	Expert metadata created by subject specialists and/or information professionals, often not the original creator of the information object	<ul style="list-style-type: none"> • Specialized subject headings • MARC records • Archival finding aids • Catalog entries for museum objects • Ad hoc metadata created by subject experts, e.g., notations by scholars or researchers
Status	Static metadata that does not or should not change once it has been created	<ul style="list-style-type: none"> • Technical information such as the date(s) of creation and modification of an information object, how it was created, file size
	Dynamic metadata that may change with use, manipulation, or preservation of an information object	<ul style="list-style-type: none"> • Directory structure • User transaction logs
	Long-term metadata necessary to ensure that the information object continues to be accessible and usable	<ul style="list-style-type: none"> • Technical format and processing information • Rights information • Preservation management documentation
	Short-term metadata, mainly of a transactional nature	<ul style="list-style-type: none"> • Interim location information
Structure	Structured metadata that conforms to a predictable standardized or proprietary structure	<ul style="list-style-type: none"> • MARC • TEI • EAD • CDWA Lite • Local database formats
	Unstructured metadata that does not conform to a predictable structure	<ul style="list-style-type: none"> • Unstructured note fields and other free-text annotations

Attribute	Characteristics	Examples
Semantics	Controlled metadata that conforms to a standardized vocabulary or authority form, and that follows standard content (i.e., cataloging) rules	<ul style="list-style-type: none"> • LCSH, LCNAF, AAT, ULAN, TGM, TGN • AACR (RDA), DACS, CCO
	Uncontrolled metadata that does not conform to any standardized vocabulary or authority form	<ul style="list-style-type: none"> • Free-text notes • HTML META tags and other user-created tags
Level	Collection-level metadata relating to collections of original items and/or information objects	<ul style="list-style-type: none"> • Collection- or group-level record, e.g., a MARC record for a group or collection of items; a finding aid for an intact archival collection • Specialized index
	Item-level metadata relating to individual items and/or information objects, often contained within collections	<ul style="list-style-type: none"> • Catalog records for individual bibliographic items or unique cultural objects • Transcribed image captions and dates • "Tombstone" information for works of art and material culture • Format information

O ciclo de vida de um objeto de informação

Funções primárias de Metadados

- Criação, multiversiionamento, reuso e recontextualização de objetos de informação (*criação do objeto digital ou digitalizado*)
- Organização e descrição
- Validação (afinamento de metadados)
- Busca e recuperação (local ou virtualmente)
- Utilização e preservação (Anotações de uso, direitos de uso, etc)
- Disposição (Adição ou eliminação de registro e relacionados)

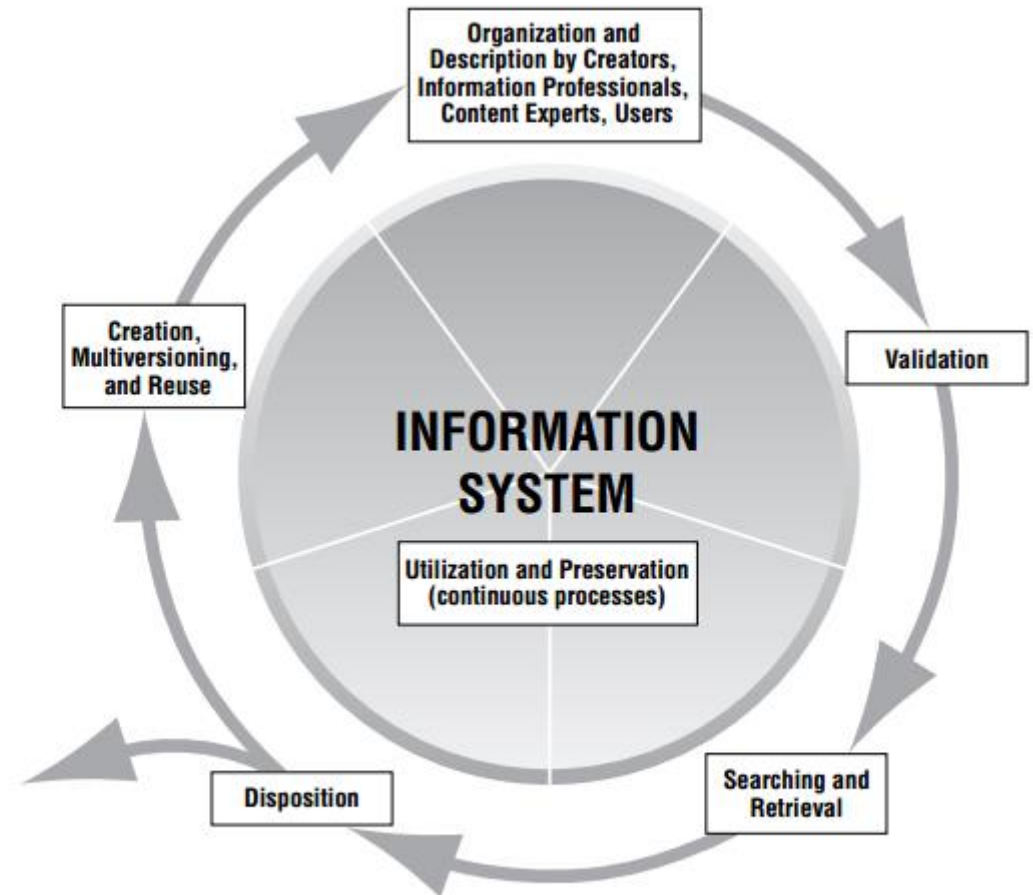


Figure 1. The Life Cycle of an Information Object

Fatos pouco conhecidos sobre metadados

- Metadados não precisam necessariamente serem digitais
- Metadados vão muito além da descrição de um objeto de informação
- Metadados podem vir de uma diversa gama de fontes
- Metadados continuam a serem atualizados ao longo da vida útil de um objeto de informação ou sistema
- Um metadado de **um objeto de informação pode simultaneamente ser o dado de outro objeto de informação**, dependendo dos tipos de 'agregações' e dependência entre objetos de informação e sistemas

Porque metadados são importantes?

Embora a criação e manutenção de metadados pode ser extremamente custosa, o desenvolvimento da web e sistemas digitais em rede proporcionam diversas oportunidades para profissionais da informação:

- Propagar e expandir acessibilidade
- Reter contextos
- Expandir o uso
- Metadados educacionais
- Desenvolvimento e aperfeiçoamento de sistemas
- Multiversiõamento
- Questões legais
- Preservação e persistência
- Melhorias de sistema e economia

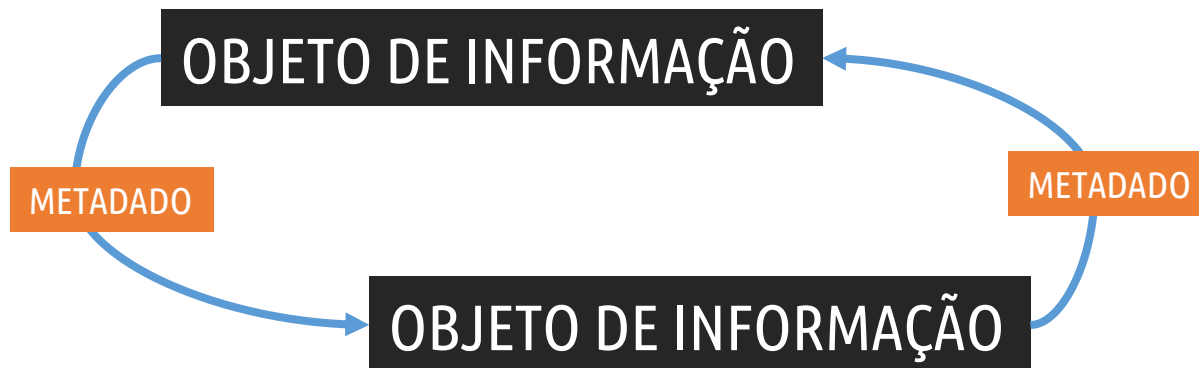
Conclusões e questões pendentes

Metadados são como interesses – eles aparecem com o tempo. Sábios investimentos geram o melhor resultado em capital intelectual. Metadados bem feitos e manejados resultam em uma melhor gestão da informação, e o melhor aproveitamento do usuário final, com retorno imediato e também a longo prazo.

- Questões chave que devem ser resolvidas pelos profissionais de informação:
- Identificar qual o melhor esquema de metadados (MARC, EAD, etc)
- Decidir qual aspectos de metadados são essenciais para o objetivo desejado e como 'granular': "How much is enough and how much is too much"
- Assegurar-se que o esquema de metadados e vocabulários controlados que estão sendo aplicados estejam atualizados e que são compostos pelas terminologias apropriadas

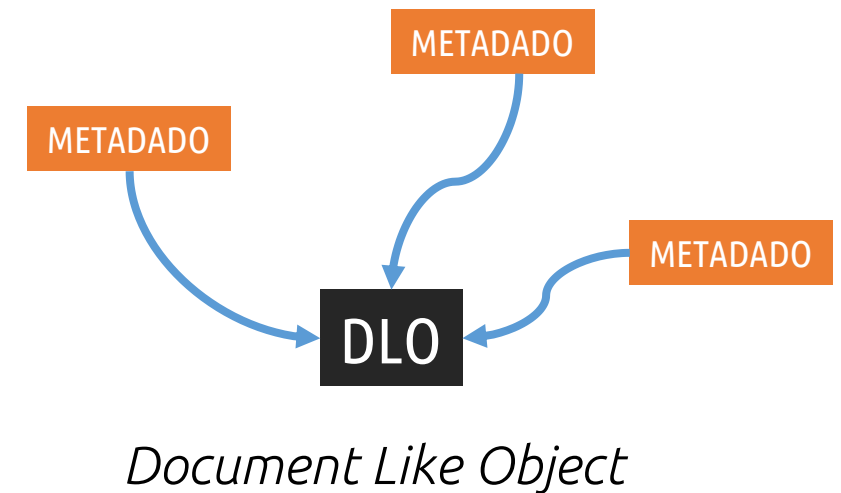
Gilliland

“Um metadado de um objeto de informação pode ser simultaneamente ser o dado de outro objeto de informação, dependendo dos tipos de ‘agregações’ e dependência entre objetos de informação e sistemas”



Méndez

“Os metadados são também dados, mas dados representacionais, que somados à própria informação adquirem valor semântico para substituí-la ou representá-la”



Aplicações atuais (referência 2015)

- Conta Google
- Anúncios digitais (AdWords, Social Ads, Mídia Programada, publicidade persistente)
- Monitoramento da rede (conteúdo publicado e redes sociais)
- Documentário [Do Not Track](#)
- Waze