

# Esquemas de Metadados FAIR

## fundamentos e aplicações

Filipi Miranda Soares

Universidade de São Paulo;  
University of Twente



Esta apresentação é baseada no artigo:

Soares, F. M., Pires, L. F., Santos, L. O. B. D. S., Calhau, R. F., Moreira dos Santos Maculan, B. C., Coyle, K., Wang, S., Folmer, E., Pignatari Drucker, D., Campos, M. L. D. A., Marcondes, C. H., Almeida, M. B., Braghetto, K. R., Dias, G. A., Salim, J. A., Corrêa, F. E., de Abreu Moreira, D., Botazzo Delbem, A. C., & Saraiva, A. M. (2024). Towards a Conceptual Model for FAIR Metadata Schemas. In E. Gallinucci, & H. Yasar (Eds.), *Companion Proceedings of the 43rd International Conference on Conceptual Modeling: ER Forum, Special Topics, Posters and Demos: Co-located with ER 2024* (Vol. 3849, pp. 42-55). (CEUR Workshop Proceedings; Vol. 3849). CEUR.  
<https://ceur-ws.org/Vol-3849/forum4.pdf>

**Carnegie Mellon University**

Software Engineering Institute

# The 43rd International Conference on Conceptual Modeling



Conceptual Modeling, AI, and Beyond

28-31 OCTOBER 2024 | PITTSBURGH, PENNSYLVANIA, USA

# O que são Esquemas de Metadados?

São estruturas formais que definem os elementos de metadados necessários para descrever, organizar e gerenciar dados em um domínio específico.

## Principais características:

- ✓ Definem **classes** e **propriedades** (elementos de metadados)
- ✓ Representam o **modelo semântico** de um domínio
- ✓ Estabelecem regras para a **estruturação e interpretação** dos dados
- ✓ Devem ser **documentados** e compreensíveis por humanos e máquinas

# Esquemas de Metadados vs. Ontologias

Característica	Esquemas de Metadados	Ontologias
<b>Finalidade</b>	Descrever e organizar dados em um domínio	Representar conhecimento formal de um domínio
<b>Elementos principais</b>	Classes e propriedades para descrever dados	Conceitos, relações, axiomas, regras
<b>Rigor semântico</b>	Moderado	Alto (usa lógica formal)
<b>Linguagens comuns</b>	XML, RDF, JSON-LD	OWL, RDF(S)
<b>Foco</b>	Organização e descoberta de dados	Entendimento, inferência e interoperabilidade semântica
<b>Exemplos</b>	Dublin Core, DCAT, Darwin Core	FOAF, BFO, RO, AGROVOC

# Por que desenvolver novos esquemas de metadados?

## 1. Necessidades específicas do domínio

- Esquemas existentes podem não representar adequadamente os conceitos e processos de determinados contextos (ex: agricultura, biodiversidade, saúde pública).

## 2. Evolução tecnológica e metodológica

- Novas formas de coleta, análise e compartilhamento de dados exigem modelos atualizados.

# Por que desenvolver novos esquemas de metadados?

## 3. Interoperabilidade e integração de dados

- Construir esquemas mais compatíveis com padrões abertos facilita a conexão entre sistemas e repositórios distintos.


## 4. Empoderamento comunitário

- Desenvolver esquemas participativos permite que comunidades representem seus próprios dados com autonomia e precisão.

# Princípios FAIR

Os Princípios FAIR foram propostos para melhorar a gestão e reutilização de dados científicos. Eles definem **características desejáveis para dados e metadados**:

	Princípio	Objetivo Principal
F	<b>Findable</b>	Tornar dados fáceis de encontrar
A	<b>Accessible</b>	Garantir acesso claro e confiável
I	<b>Interoperable</b>	Permitir integração com outros dados e sistemas
R	<b>Reusable</b>	Favorecer o reuso com clareza e licenças

 FAIR não significa “gratuito” ou “aberto”, mas sim que os dados/metadados sejam bem descritos, identificáveis e legíveis por máquina.

# Por que aplicar os Princípios FAIR a esquemas de metadados?

## 1. Esquemas também são objetos digitais

- Assim como os dados que descrevem, os esquemas precisam ser **encontráveis, acessíveis, interoperáveis e reutilizáveis**.

## 2. Facilita a descoberta e reuso por outras comunidades

- Permite que outros projetos encontrem, avaliem e reutilizem o esquema com confiança.



# Por que aplicar os Princípios FAIR a esquemas de metadados?

## 3. Melhora a interoperabilidade entre sistemas

- Ao adotar padrões abertos e identificadores persistentes, os esquemas favorecem a integração entre diferentes fontes de dados. **Exemplo ruim:**  
<https://hdl.handle.net/10568/69166>


## 4. Garante a sustentabilidade e governança

- Licenças claras, documentação versionada e autoridade definida fortalecem a manutenção a longo prazo.

# Por que aplicar os Princípios FAIR a esquemas de metadados?

## 5. Permite avaliações automatizadas de conformidade

- Estruturas FAIR facilitam validações automáticas por ferramentas, promovendo qualidade e confiança.

 *Aplicar FAIR aos esquemas garante que eles próprios sejam recursos sustentáveis, abertos e úteis na construção de uma Web de Dados robusta.*

# Objetivo do Artigo

**Propor um modelo conceitual para esquemas de metadados FAIR**, com base na linguagem OntoUML, que permita:

- ✓ **Interpretar os Princípios FAIR** no contexto da modelagem de esquemas de metadados, identificando os requisitos estruturais e semânticos que eles impõem.
- ✓ **Fornecer um modelo ontológico claro** que represente os principais componentes de um esquema de metadados alinhado aos princípios FAIR, como: autoridade, landing page, elementos, definições e serializações.

# Objetivo do Artigo

✓ **Auxiliar comunidades de prática** (por exemplo, biodiversidade, agricultura, ciências sociais) a **construir ou revisar** seus esquemas para que sejam **descobertos, reutilizados e interoperáveis**.

✓ **Viabilizar avaliações de conformidade FAIR** por meio de um modelo estruturado que pode servir como referência em processos de curadoria, documentação e publicação.

Responder à pergunta: Como os princípios FAIR se aplicam a metadata schemas?

# Fundamentação metodológica

## O que é OntoUML?

OntoUML é uma **linguagem de modelagem ontológica** baseada em UML, mas com restrições e estereótipos fundamentados em **ontologia formal**.

Seu objetivo é **representar com precisão os conceitos de um domínio**, respeitando sua natureza ontológica.

- Permite criar **modelos conceituais ricos e semanticamente precisos**
- Foco em representar **o significado dos elementos**, não apenas sua estrutura técnica

## Conexão com a UFO (Unified Foundational Ontology)

OntoUML se baseia na **UFO**, uma ontologia de alto nível desenvolvida por Giancarlo Guizzardi.

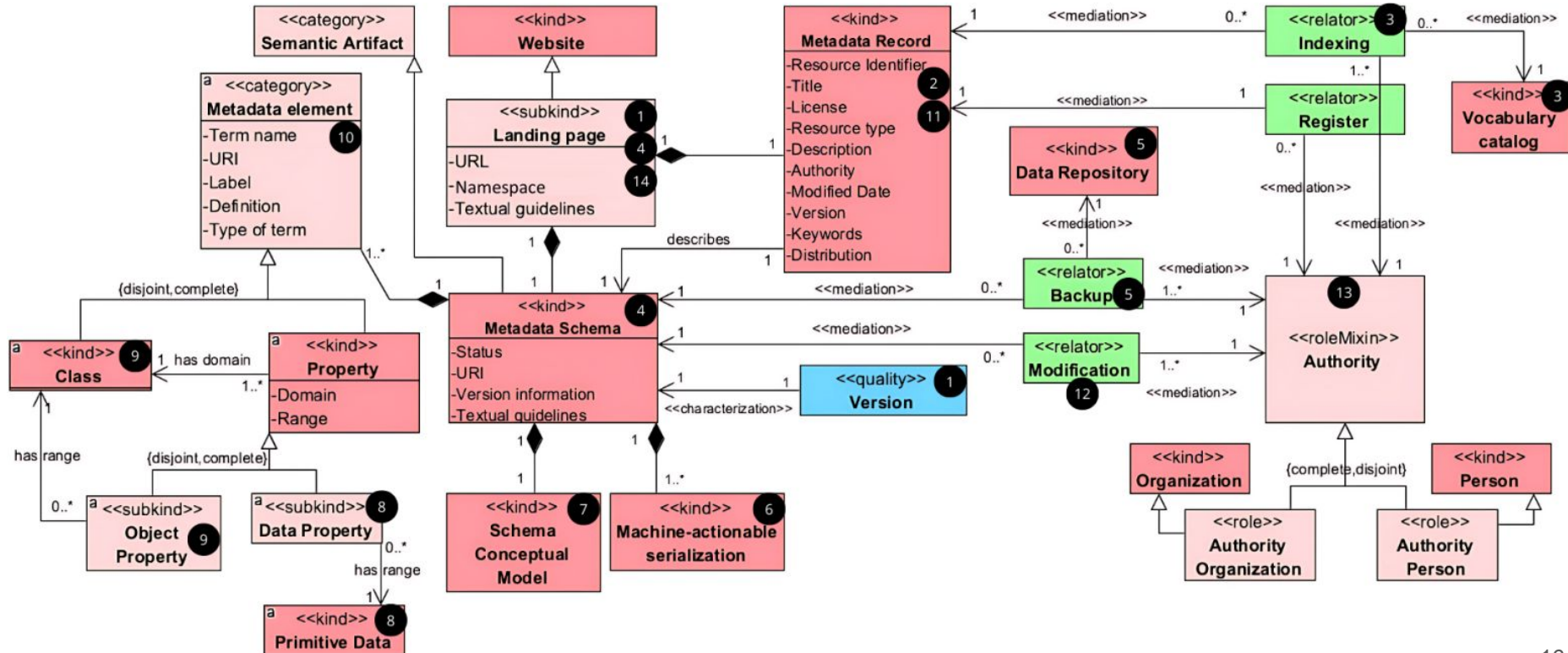
- A UFO fornece a **base filosófica e lógica** para os estereótipos de OntoUML. Principais estereótipos: Kind, Subkind, Role, Phase, Mixin, Relator e Quality.
- Ajuda a garantir **consistência e coerência ontológica** nos modelos

Análise ontológica - *ontological unpacking* (Guizzardi et al. 2021)

## Etapas:

- Análise de domínio (ex: GO FAIR Agro, Almes Core, PPI, WorldFAIR)
- Consulta à literatura (ex: ISO/IEC 11179, Marcia L. Zeng 2022, 2006 )
- Definição dos requisitos (vide [Tabela 1, Soares et al. 2024](#))
- Validação do modelo com avaliação FAIR

# Modelo Conceitual



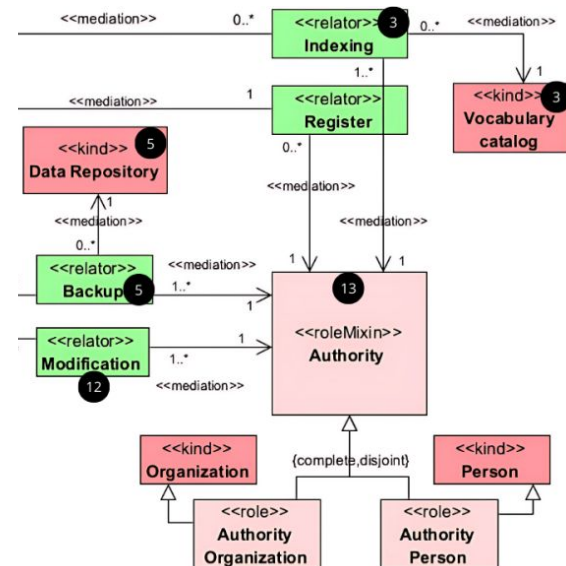


# Autoridade do Esquema

Entidade responsável pela publicação, curadoria e manutenção de um esquema de metadados FAIR. Pode ser uma organização, grupo de pesquisa, consórcio ou indivíduo com credibilidade reconhecida na comunidade.

## Principais responsabilidades:

- **Publicação oficial** do esquema com identificadores persistentes (ex: URI, DOI)
- **Governança e versionamento**, com registro claro de mudanças
- **Documentação pública**, com licenças abertas e histórico de versões
- **Transparência sobre decisões**, critérios de atualização e envolvimento comunitário
- **Estabilidade e confiança**, assegurando que o esquema continuará acessível e confiável ao longo do tempo

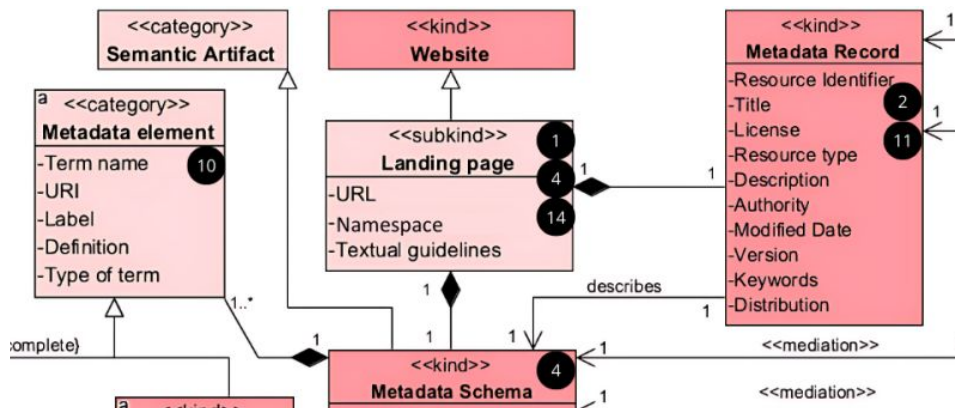



# Landing Page

É a **página web principal** onde o esquema de metadados pode ser acessado por humanos e máquinas. Deve ser estável, informativa e referenciável por identificadores persistentes.

## Acessibilidade Técnica:

- Uso de **protocolos abertos e padrão da web** (ex: **HTTPS**)
- **Conteúdo negociável**: deve possibilitar tanto **visualização humana** quanto **resolução automática** por agentes de software

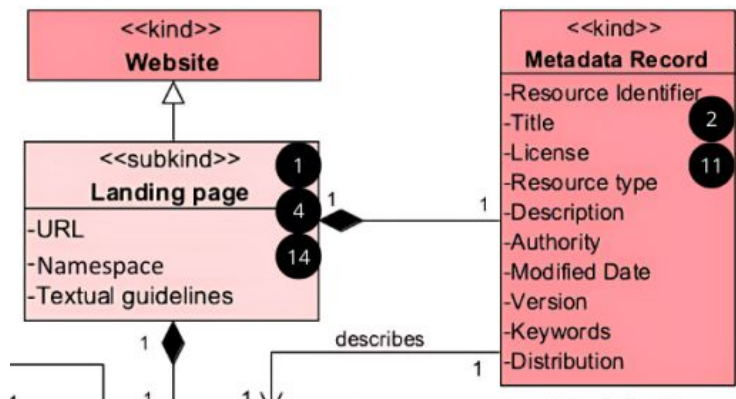


 A landing page é o ponto de entrada para o reuso do esquema. Se ela não é clara ou acessível, o esquema não é FAIR.

Exemplo: <https://w3id.org/AlmesCore#>

# Registro de Metadados

Recomendação do Cross Domain Interoperability Framework (CDIF, 2023); deve ser **legível por humanos e por máquinas**. → <https://search.google.com/test/rich-results?hl=pt-br>

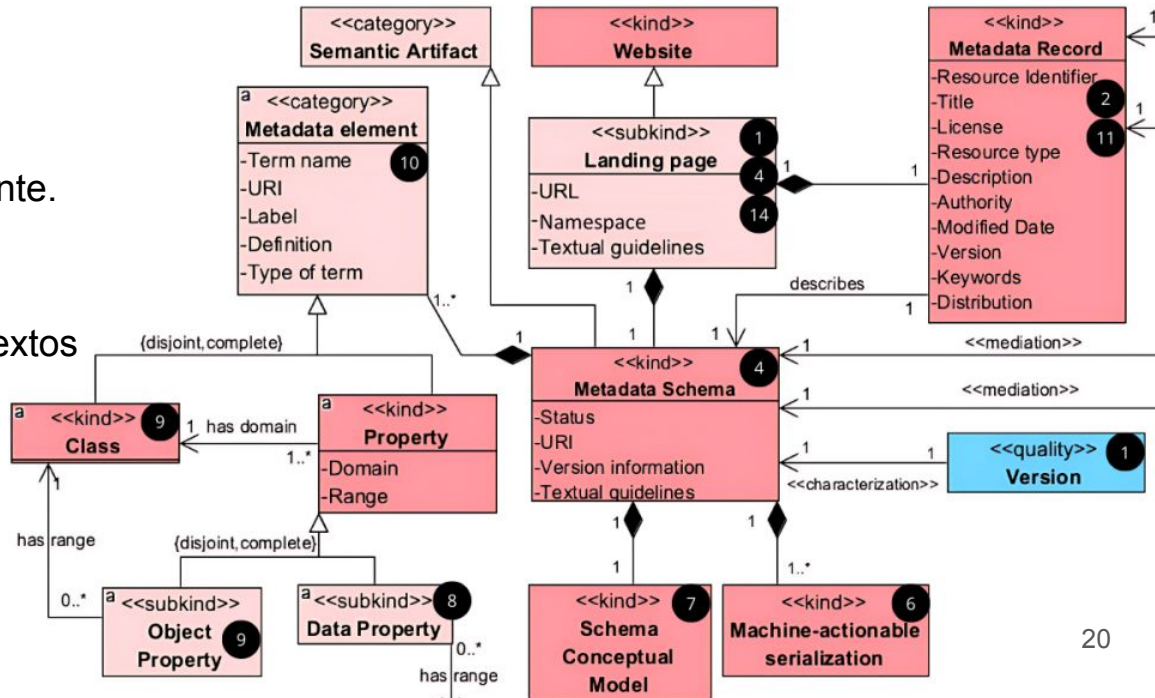


Exemplo: <https://github.com/AlmesCore/Almes-Core/edit/main/index.md>

# Esquema de Metadados

É um **artefato semântico** voltado à **organização, categorização e gestão da informação e do conhecimento** dentro de um domínio ou contexto específico.

- Atua como estrutura formal para descrever dados de forma consistente.
- É essencial para promover interoperabilidade e reuso em contextos FAIR.



# Elemento de Metadados

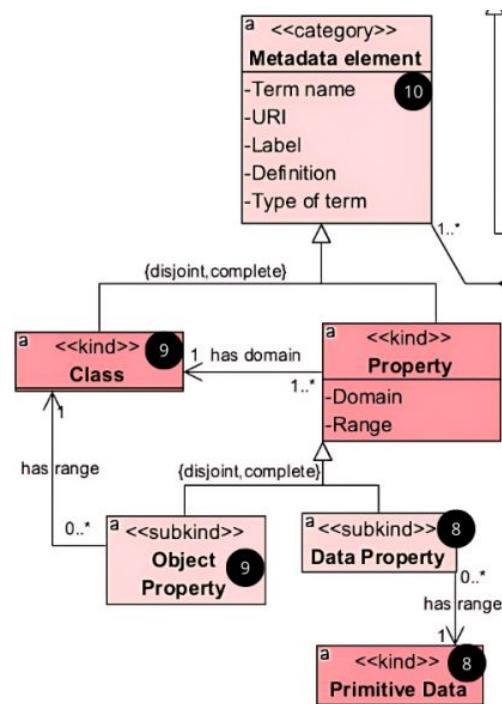
São as unidades fundamentais dos esquemas de metadados — elementos que descrevem aspectos relevantes de um recurso (ex: título, autor, data de publicação).

## 🧠 Class vs. Property (Abordagem Ontológica):

- **Classes:** Representam **tipos de entidades** ou categorias do mundo real (ex: *Dataset*, *Person*, *Place*)
- **Propriedades:** Representam **atributos ou relações** entre classes (ex: *hasAuthor*, *publicationDate*)

## 🔄 Domain e Range:

- **Domain:** Especifica a que classe a propriedade se aplica (ex: *hasAuthor* → domínio: *Dataset*)
- **Range:** Especifica o tipo de valor esperado (ex: *hasAuthor* → range: *Agent*)



# Elemento de Metadados



## Descrição formal e textual

- Elementos devem ser descritos:
  - Em **formato legível por humanos** (ex: glossário)
  - Em **formato formal legível por máquina** (ex: RDF/OWL)



*A clareza na definição de classes e propriedades é essencial para garantir coerência, interoperabilidade e automação no uso de esquemas de metadados.*

# Semântica dos Elementos

Para que metadados sejam **compreendidos, comparados e reutilizados** de forma precisa, seus elementos precisam ser descritos de maneira **semântica clara e padronizada**.

Componentes semânticos essenciais:

Elemento	Descrição
<b>Nome</b>	Rótulo único e descritivo, preferencialmente curto e claro (ex: <b>creator</b> )
<b>Definição</b>	Explicação precisa do significado do elemento e seu uso esperado
<b>Escopo</b>	Contexto em que o elemento é aplicável; evita ambiguidade ou sobreposição. Ex.: <a href="http://purl.org/dc/terms/created">http://purl.org/dc/terms/created</a>

# Semântica dos Elementos

## Diretrizes da ISO/IEC 11179 (Partes 4 e 5):

A norma internacional para registro de elementos de metadados propõe:

- **Regras para nomeação e definição** de elementos de dados
- **Separação entre conceito e representação** (ex: conceito “sexo” vs. valores “M/F”)
- Uso de **identificadores únicos e persistentes**
- Documentação de **restrições, domínios de valor e unidades**



# Serialização para Máquinas

Para que esquemas sejam **interpretáveis por agentes automáticos**, eles devem ser disponibilizados em **formatos formais, legíveis por máquina**. Isso viabiliza:

- Validação automática de dados
- Descoberta semântica
- Reuso interoperável
- Integração em aplicações e pipelines FAIR

Formatos comuns: RDF, OWL, JSON-LD



## Boas práticas recomendadas:

- Disponibilizar a serialização junto à **landing page**
- Garantir **conteúdo negociável** (ex: a URI de um elemento pode devolver HTML para humanos ou RDF para máquinas)

# Serialização para Máquinas

Identificadores Persistentes (URIs):

Cada classe ou propriedade deve possuir uma **URI global única e persistente**

Exemplos de domínios confiáveis para publicação:

- <https://w3id.org/> – mantido pela comunidade com apoio técnico da W3C
- <https://purl.org/> – redirecionamento confiável mantido pela OCLC - **instável!**
- <https://doi.org/> – para publicação formal e versionamento de artefatos - **pouco utilizado.**

# Modelo Conceitual do Esquema

É uma **representação abstrata dos fenômenos** que o esquema de metadados pretende descrever, capturando os **conceitos essenciais** e suas **relações**.

Também é conhecido como *metamodel* ou *modelo de alto nível*.

 **Funções principais de um modelo conceitual em esquemas FAIR:**

1. **Projeto do esquema (Schema Design)**
2. **Integração e mapeamento de dados (Data Integration)**
3. **Compreensão humana (Human Understanding)**

**Linguagens e Exemplos:**

UML é uma das linguagens mais populares para modelos conceituais de esquemas

- **EML (Ecological Metadata Language)** usa UML para descrever dados ecológicos
- **DDI-CDI** adota um modelo UML como “núcleo” de sua estrutura semântica

# Avaliação de FAIRness

## Método proposto no artigo

O modelo conceitual foi usado como **instrumento de diagnóstico** para avaliar o grau de conformidade de esquemas de metadados com os **Princípios FAIR**.

- Cada componente do modelo conceitual foi associado a um ou mais princípios FAIR (F, A, I, R)
- Criou-se uma **tabela de verificação** com perguntas-chave para cada componente (ex: o esquema tem namespace persistente? landing page está documentada? possui versionamento?)
- A avaliação foi realizada **manualmente**, com base na documentação pública dos esquemas

# Avaliação de FAIRness

## Casos analisados:

Esquema	FAIRness (%)	Comentários
DCAT	100%	Excelente documentação, URI persistente, serialização formal completa
ALM	91,75%	O esquema não estava indexado em vocabulary catalogs -- <b>Corrigido!</b> .
Darwin Core	87%	Utiliza namespace estável, mas falta definir serializações acionáveis por máquina e domain e ranges entre os elementos.

# Avaliação de FAIRness

Vamos avaliar juntos?

<http://purl.org/agmes/1.1/>

# Considerações finais

## Contribuições do modelo conceitual

- Limitações da abordagem manual
- Futuro: automação e validação empírica (testar o modelo com comunidades reais de prática para refinar sua aplicabilidade)
- **Aprimorar o modelo conceitual** com base no feedback de usuários e novos requisitos emergentes

# Referências

M. L. Zeng, J. Qin, Metadata, third edition ed., ALA Neal-Schuman, Chicago, 2022.

L. M. Chan, M. L. Zeng, Metadata interoperability and standardization - a study of methodology part i: Achieving interoperability at the schema level, D-Lib Magazine 12 (2006). doi:10.1045/june2006- chan.

G. Guizzardi, A. Bernasconi, O. Pastor, V. C. Storey, Ontological unpacking as explanation: The case of the viral conceptual model, in: A. Ghose, J. Horkoff, V. E. Silva Souza, J. Parsons, J. Evermann (Eds.), Conceptual Modeling, Springer International Publishing, Cham, 2021, pp. 356–366.

Cross-Domain Interoperability Framework (CDIF) Working Group, S. Richard, A. Gregory, S. Hodson, D. Fils, et al., Cross domain interoperability framework (cdif): Discovery module (v01 draft for public consultation), 2023. doi:10.5281/ZENODO.10252564.



Obrigado!

Contato: [f.mirandasoares@utwente.nl](mailto:f.mirandasoares@utwente.nl)

Agradecimentos especiais:



UNIVERSITY  
OF TWENTE.