

1er. Coloquio sobre RDA  
en América Latina

BIBLIOTECA  
NACIONAL  
DE MÉXICO

# CONTRIBUIÇÕES DO BIBFRAME PARA O ENRIQUECIMENTO SEMÂNTICO DE OBJETOS PUBLICADOS NA WEB

Msc. Claudiane Emanuele Nazário - UFMG/Brasil

Dra. Célia da Consolação Dias – UFMG/Brasil

# CONTRIBUCIONES DEL BIBFRAME PARA EL ENRIQUECIMIENTO SEMÁNTICO DE OBJETOS PUBLICADOS EN LA WEB

14 al 16 de noviembre de 2018

## Introdução

- A Library of Congress, iniciou em 2011 a construção de um padrão de metadados denominado **Bibliographic Framework Initiative (BIBFRAME)**, cuja proposta era de substituir o MARC21. O objetivo deste modelo de metadados era de publicar os registros bibliográficos na web, utilizando os princípios do Linked Data, de acordo com a proposta do FRBR.
- **Objetivo deste trabalho:** avaliar as contribuições do BIBFRAME para o enriquecimento semântico de objetos publicados na WEB através do Linked Data.
- **Metodologia:** criação e aplicação de uma Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos (**Matriz TRESO**) composta por 7 critérios e desenvolvida durante a pesquisa de mestrado em Gestão & Organização do Conhecimento na Escola de Ciência da Informação da Universidade Federal de Minas Gerais, Brasil.

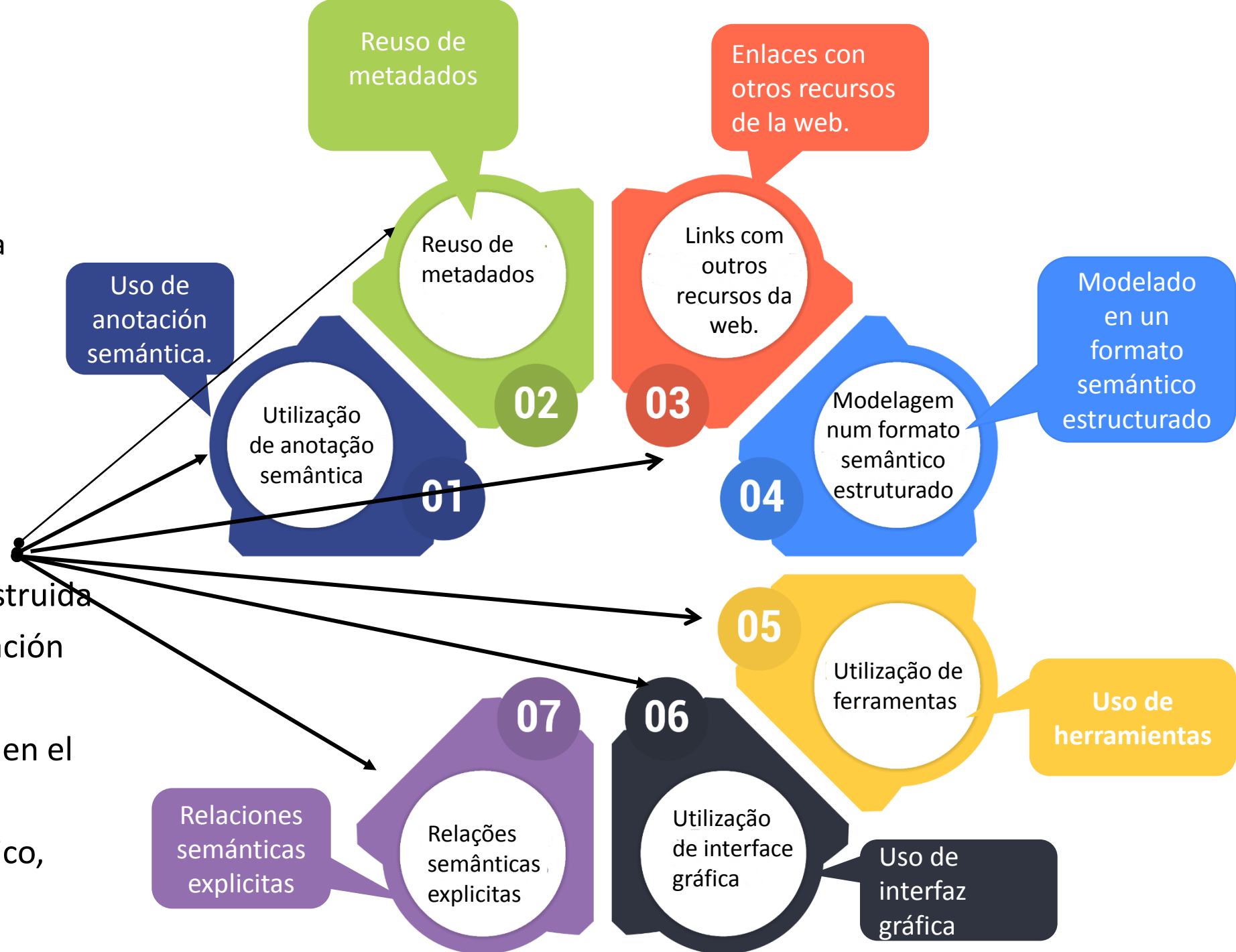
## Introducción

- La Biblioteca del Congreso, inició en 2011 la construcción de un patrón de metadatos denominado Bibliographic Framework Initiative (BIBFRAME), cuya propuesta era sustituir el MARC21. El objetivo de este modelo de metadatos era publicar los registros bibliográficos en la web, utilizando los principios del Linked Data, de acuerdo con la propuesta del FRBR.
- **Objetivo de este trabajo:** evaluar las contribuciones del BIBFRAME para el enriquecimiento semántico de objetos publicados en la WEB a través de Linked Data.
- **Metodología:** creación y aplicación de una Matriz de Técnicas y Recursos para el Enriquecimiento Semántico de Objetos (Matriz TRESO) compuesta por 7 criterios y desarrollada durante la investigación de maestría en Gestión y Organización del Conocimiento en la Escuela de Ciencia de la Información de la Universidad Federal de Minas Gerais, Brasil.

# Matriz TRESO

A Matriz TRESO foi construída para possibilitar a **verificação de quais os recursos e técnicas estão presentes no modelo de dados para o enriquecimento semântico**, considerando 7 critérios.

La Matriz TRESO fue construida para posibilitar la verificación de cuáles los recursos y técnicas están presentes en el modelo de datos para el enriquecimiento semántico, considerando 7 criterios.



## **Critério 1 - Anotação semântica**

- como o modelo utiliza o recurso de anotação semântica **para atribuir significado aos recursos**.
- A anotação é realizada por meio de *tags* e de propriedades que possibilitam o uso de texto livre e de linguagem natural.
- **A utilização de linguagem natural pode promover incorreções, ambigüidade e dificuldade de interpretação por homens e máquinas**, sendo considerada **limitada** se comparada à anotação semântica **utilizando ontologias e vocabulários**.
- **Nota após a avaliação: Nota 2 pelo uso de linguagem natural**, pois a excelência neste quesito seria atribuída caso a anotação, utilizasse ontologias.

## **Criterio 1 - Anotación semántica**

- Modo en que el modelo utiliza la característica de anotación semántica para asignar significado a los recursos.
- Se observó que en el BIBIFRAME la anotación es realizada por medio de tags y de propiedades que posibilitan el uso de texto libre y de lenguaje natural.
- **La utilización del lenguaje natural puede promover incorrecciones, ambigüedad y dificultad de interpretación por hombres y máquinas**, siendo considerada limitada si se compara con la anotación semántica **utilizando ontologías y vocabularios**.
- **Nota después de la evaluación: Nota 2 por el uso del lenguaje natural**, pues la excelencia en este aspecto sería atribuida si la anotación, utilizaba ontologías.

## Critério 2 - Reuso de metadados

Visa avaliar a reutilização de metadados de outras fontes para facilitar a publicação de dados em Linked Data.

- Versão 2.0 do BIBFRAME: as propriedades identifier, language, source e title, possuem a mesma nomenclatura e função presentes no padrão Dublin Core...
- ... a propriedade CopyrightRegistration possui a mesma finalidade da propriedade Rights management do Dublin Core.
- **Nota após a avaliação:** Constatou-se que o modelo realiza a duplicação e não a reutilização do dado em sua fonte de origem → **Nota: Zero.**

## Criterio 2 - Reuso de metadatos

- Evaluar la reutilización de metadatos de otras fuentes para facilitar la publicación de datos en Linked Data.
- En la versión 2.0 de BIBFRAME, las propiedades identifier, language, source e title, poseen la misma nomenclatura y función presentes en el estándar Dublin Core. Y la propiedad CopyrightRegistration tiene el mismo propósito de la propiedad Rights management de Dublin Core.
- **Nota después de la evaluación:** Se constató que el modelo realiza la **duplicación y no la reutilización del dato** en su fuente de origen → **Nota: Cero.**



### Critério 3 - Links entre dados e metadados do modelo com outros recursos da web

- A quantidade de conexões com outros *datasets* está relacionada à facilidade para publicar os dados em *Linked Data* e realizar a interoperabilidade de dados.
- Para esta avaliação foram verificadas a quantidade de conexões do BIBFRAME no LOD (Linked Open Data):

Critério	BIBFRAME
Utilização do Vocabulário no LOD	0 datasets
Links de saída	2 links
Links de entrada	1 link
Classes utilizadas	0 propriedade
Propriedades utilizadas	0 propriedade

- **Nota após a avaliação:** O BIBFRAME se conecta apenas com **2 datasets**, o que demonstra uma baixa taxa de conexão com outros vocabulários do LOD → **Nota 2.**

### Criterio 3 - Enlaces entre datos y metadatos del modelo con otros recursos de la web

- La cantidad de conexiones con otros datasets está relacionada con la facilidad para publicar los datos en Linked Data y realizar la interoperabilidad de datos.
- Para esta evaluación se verificó la cantidad de conexiones de BIBFRAME en el LOD (Linked Open Data) :

Critério	BIBFRAME
Uso del vocabulario en el LOD	0 conjuntos de datos
Enlaces de salida	2 enlace
Enlaces de entrada	1 enlace
Clases utilizadas	0 propiedad
Propiedades utilizadas	0 propiedad

- **Nota después de la evaluación:** el BIBFRAME se conecta sólo con dos datasets, lo que demuestra una baja tasa de conexión con otros vocabularios del LOD → **Nota 2.**

## **Critério 4 - Modelagem de dados num formato semântico estruturado**

- O BIBFRAME é modelado por meio do RDS/XML. As **relações existentes entre os objetos e suas propriedades são descritos por meio de grafos RDF**, o que favorece uma maior liberdade e simplificação das representações...
- Não são definidas a natureza e a tipologia das relações do tipo todo-parte, gênero-espécie, etc.
- **Nota após a avaliação:** Os dados serem modelados num formato semântico estruturado, como o RDF/XML, esse modelo não faz uso de ontologias para o desenvolvimento do seu processo de modelagem. → **Nota: 3**
- O uso de ontologias é superior ao RDF/XML uma vez que permitem a identificação dos tipos de relacionamento entre os conceitos...
- ...possibilita a realização de inferências sobre esses relacionamentos.

## **Criterio 4 - Modelado de datos en un formato semántico estructurado**

- El BIBFRAME es modelado a través de RDS / XML. En el BIBFRAME las relaciones existentes entre los objetos y sus propiedades se describen por medio de grafos RDF, lo que favorece una mayor libertad y simplificación de las representaciones. ..
- No se definen la naturaleza y la tipología de las relaciones del tipo todo-parte, género-especie, etc.
- **Nota después de la evaluación:** Los datos se modelan en un formato semántico estructurado, como el RDF / XML, este modelo no hace uso de ontologías para el desarrollo de su proceso de modelado. → **Nota: 3**
- El uso de ontologías es superior al RDF / XML ya que permiten la identificación de los tipos de relación entre los conceptos...
- ... posibilita la realización de inferencias sobre esas relaciones.

## **Critério 5 - Utilização de ferramentas para o enriquecimento semântico**

- O BIBFRAME fornece uma ferramenta em Excel para a conversão de dados em formato MARC.
- Serão fornecidas especificações, ferramentas e serviços de *software* para que as comunidades usuárias possam fazer *download* e converter seus dados do formato MARC para BIBFRAME. (Library of Congress, 2017).
- **Nota após a avaliação:** A ferramenta disponibilizada pelo modelo realiza somente a conversão de formatos e não realiza outras atividades que possibilitam o enriquecimento semântico, tais como: extração de dados, classificação, anotação, publicação e disseminação. → **Nota 2.** Baixo atendimento

## **Criterio 5 - Utilización de herramientas para el enriquecimiento semántico**

- BIBFRAME proporciona una herramienta en Excel para la conversión de datos en formato MARC.
- se proporcionarán especificaciones, herramientas y servicios de software para que las comunidades usuarias puedan descargar y convertir sus datos del formato MARC a BIBFRAME. (Biblioteca del Congreso, 2017).
- **Nota después de la evaluación:** La herramienta disponibilizada por el modelo realiza solamente la conversión de formatos y no realiza otras actividades que posibilitan el enriquecimiento semántico, tales como: extracción de datos, clasificación, anotación, publicación y diseminación. → **Nota 2.**



## **Critério 6 - Utilização de interface gráfica**

▪ considera a utilização de interface gráfica para processo de enriquecimento semântico e publicação de dados. Segundo Uren et. al. (2005) para realizar um processo colaborativo é crucial fornecer aos publicadores, interfaces gráficas que simplifiquem o processo de anotação.

▪ **Nota após a avaliação:** Não foi possível identificar a utilização de interface gráfica para o enriquecimento semântico → **Nota: Zero**

## **Criterio 6 - Utilización de interfaz gráfica**

▪ Considera la utilización de interfaz gráfica para el proceso de enriquecimiento semántico y la publicación de datos. Según Uren et. al. (2005) para realizar un proceso colaborativo es crucial proporcionar a los publicadores, interfaces gráficas que simplifiquen el proceso de anotación.

▪ **Nota después de la evaluación:** No fue posible identificar el uso de interfaz gráfica para el enriquecimiento semántico. → **Nota: Cero**

## **Critério 7 – Relacionamento entre os metadados do modelo e termos de outros vocabulários**

- As relações semânticas não são hierárquicas, associativas e de equivalência como algumas que estão presentes, em uma ontologia...
- são representadas por meio de grafos RDF → os dados estão conectados pela expressão **sujeito-predicado-objeto**, mas sem necessariamente estruturar as relações de forma semântica.
- **Nota após a avaliação:** BIBFRAME é modelado em RDF e não explicita o tipo de relacionamento →  
**Nota: Zero**

## **Criterio 7 - Relación entre los metadatos del modelo y los términos de otros vocabularios**

- Las relaciones semánticas no son jerárquicas, asociativas y de equivalencia como algunas que están presentes, en una ontología,
  - representadas por medio de grafos RDF, donde los datos están conectados por la expresión sujeto-predicado-objeto, pero sin necesariamente estructurar las relaciones de forma semántica.
- Nota después de la evaluación:** el BIBFRAME es modelado en RDF y no explicita el tipo de relación →  
**Nota: Zero**

## Pontuação obtida pelo BIBFRAME nos critérios da Matriz TRESO

CRITÉRIO DE AVALIAÇÃO	PONTOS
1) Utilização de recursos de anotação semântica para o enriquecimento de objetos publicados em Linked Data.	2
2) Reuso de metadados para facilitar as atividades de publicação de dados em Linked Data, otimizando o trabalho do publicador.	0
3) Links entre as combinações semânticas dos dados e metadados com outros recursos da web.	2
4) Modelagem de dados num formato semântico estruturado, permitindo sua manipulação por aplicações que consomem esses modelos de dados, ampliando as possibilidades de conexão com outros Datasets do LOD.	3
5) Utilização de ferramentas para o enriquecimento semântico.	2
6) Utilização de interface gráfica para apoiar a execução do processo de enriquecimento semântico e publicação de dados.	0
7) Relações de sinonímia (equivalência), associação e hierarquia entre o metadado e o termo correspondente em outros vocabulários utilizados.	0
<b>RESULTADO – ATENDIMENTO PARCIAL</b>	<b>9</b>

## Puntuación obtenida por el BIBFRAME en los criterios de la Matriz TRESO

CRITÉRIO DE AVALIAÇÃO	PUNTOS
1) Utilización de recursos de anotación semántica para el enriquecimiento de objetos publicados en Linked Data.	2
2) Reuso de metadatos para facilitar las actividades de publicación de datos en Linked Data, optimizando el trabajo del publicador.	0
3) Enlaces entre las combinaciones semánticas de los datos y metadatos con otros recursos de la web.	2
4) Modelado de datos en un formato semántico estructurado, permitiendo su manipulación por aplicaciones que consumen estos modelos de datos, ampliando las posibilidades de conexión con otros Fases de LOD.	3
5) Utilización de herramientas para el enriquecimiento semántico.	2
6) Uso de interfaz gráfica para apoyar la ejecución del proceso de enriquecimiento semántico y la publicación de datos.	0
7) Relaciones de sinonimia (equivalencia), asociación y jerarquía entre el metadato y el término correspondiente en otros vocabularios utilizados.	0
<b>RESULTADO - ATENCIÓN PARCIAL</b>	<b>9</b>

## RECOMENDAÇÕES PARA O BIBFRAME

- Utilização de vocabulários controlados e ontologias para facilitar a compreensão da anotação pelo usuário,
- Reutilização de metadados de outros modelos em seu contexto original ao invés de duplicar.
- Realizar a conexão com outros recursos da web, na medida em que o portal LOV demonstrou uma baixa utilização do modelo no LOD.
- A modelagem de dados em um formato semântico estruturado, permitindo sua compreensão por homens e máquinas e sua reutilização;
- Explicitar os relacionamentos entre os metadados do modelo e termos de outros vocabulários
- Utilizar ferramentas para facilitar o trabalho do publicador e permitir a manipulação de grandes volumes de dados.

## RECOMENDACIONES PARA EL BIBFRAME

- Uso de vocabularios controlados y ontologías para facilitar la comprensión de la anotación por el usuario,
- Reutilización de metadatos de otros modelos en su contexto original en lugar de duplicar.
- Realizar la conexión con otros recursos web, en la medida en que el portal LOV demostró una baja utilización del modelo en el LOD.
- El modelado de datos en un formato semántico estructurado, permitiendo su comprensión por hombres y máquinas y su reutilización;
- Explicar las relaciones entre los metadatos del modelo y los términos de otros vocabularios
- Utilizar herramientas para facilitar el trabajo del publicador y permitir la manipulación de grandes volúmenes de datos.

## CONSIDERAÇÕES FINAIS | CONSIDERACIONES FINALES

- **Necessidade de ajustes:** nos recursos e técnicas para o enriquecimento semântico de objetos publicados no Linked Data utilizado no BIBFRAME.
- **alguns pontos a evoluir:** revisão das propriedades do modelo, privilegiando sempre que possível à reutilização de metadados de outros modelos, bem como a utilização de ontologias e vocabulários para modelagem de seus dados.
- **Benefícios ao BIBFRAME:** utilização de ferramentas tecnológicas para facilitar o processo de enriquecimento semântico.
- **Recomendações deste estudo :** otimização do processo de enriquecimento semântico de objetos para sua publicação em linked data → maior interoperabilidade de dados disponibilizados pelo modelo.

- **Necesidad de ajustes:** en los recursos y técnicas para el enriquecimiento semántico de objetos publicados en el Linked Data utilizado en el BIBFRAME.
- **Puntos a evolucionar:** revisión de las propiedades del modelo, privilegiando siempre que sea posible a la reutilización de metadatos de otros modelos, así como la utilización de ontologías y vocabularios para modelado de sus datos.
- **Beneficios al BIBFRAME:** utilización de herramientas tecnológicas para facilitar el proceso de enriquecimiento semántico.
- **Recomendaciones:** optimizar su proceso de enriquecimiento semántico de objetos para su publicación en linked data → mayor interoperabilidad de datos disponibilizados por el modelo.

## **SOBRE AS AUTORAS | SOBRE LAS AUTORES**

### **Msc. Claudiane Emanuele Nazário**

Mestre em Ciência da Informação – Escola de Ciência da Informação - Universidade Federal de Minas Gerais – Brasil. Certificação CDIA+ - Certified Document Imaging Architect. Especialização em Gestão da Informação e Conhecimento no Centro Federal de Educação Tecnológica de Minas Gerais - CEFET MG.

### **Dra. Célia da Consolação Dias**

Professora Adjunta do Departamento de Organização e Tratamento da Informação - Escola de Ciência da Informação - Universidade Federal de Minas Gerais – Brasil - Doutorado e mestrado em Ciência da Informação na Escola de Ciência da Informação da Universidade Federal de Minas Gerais.

### **Msc. Claudiane Emanuele Nazário**

Maestría en Ciencia de la Información - Escuela de Ciencia de la Información - Universidad Federal de Minas Gerais - Brasil. - Certificación CDIA + - Certified Document Imaging Architect. Especialización en Gestión de la Información y Conocimiento en el Centro Federal de Educación Tecnológica de Minas Gerais - CEFET MG.

### **Dra. Célia da Consolação Dias**

Profesora Adjunta del Departamento de Organización y Tratamiento de la Información - Escuela de Ciencia de la Información - Universidad Federal de Minas Gerais - Brasil - Doctorado y maestría en Ciencia de la Información en la Escuela de Ciencia de la Información de la Universidad Federal de Minas Gerais.



# AGRADECIMENTOS | RECONOCIMIENTOS

Dúvidas?

Contatos: Célia Dias: [celiadias@gmail.com](mailto:celiadias@gmail.com)  
Claudiane: [cnazario@gamarratecnologia.com.br](mailto:cnazario@gamarratecnologia.com.br)

Preguntas?

Obrigada !!! /Gracias!!

Gracias a FAPEMIG

