

UNIVERSIDADE FEDERAL DE MINAS GERAIS – UFMG
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI

Elaine Aparecida de Lima

**A IDENTIFICAÇÃO DE METADADOS DE ASSUNTO PARA A LEGISLAÇÃO FEDERAL
BRASILEIRA: O CASO DO SISTEMA LEGIN WEB DA CÂMARA DOS DEPUTADOS**

Curso patrocinado pelo
Centro de Formação, Treinamento e Aperfeiçoamento (Cefor)
da Câmara dos Deputados.

Belo Horizonte

2013

ELAINE APARECIDA DE LIMA

**A IDENTIFICAÇÃO DE METADADOS DE ASSUNTO PARA A LEGISLAÇÃO FEDERAL
BRASILEIRA: O CASO DO SISTEMA LEGIN WEB DA CÂMARA DOS DEPUTADOS**

Monografia apresentada ao programa de Especialização do Núcleo de Informação Tecnológica e Gerencial – NITEG, no curso de especialização em Arquitetura e Organização da Informação da Escola de Ciência da Informação, da Universidade Federal de Minas Gerais, como requisito para a obtenção do certificado de Especialista em Arquitetura e Organização da Informação

Linha de Pesquisa: Arquitetura e Organização da Informação

Orientador: Prof. Leandro Ribeiro
Negreiros

Belo Horizonte

2013

L628i

Lima, Elaine Aparecida de.

A identificação de metadados de assunto para a legislação federal brasileira: o caso do Sistema Legin Web da Câmara dos Deputados / Elaine Aparecida de Lima. – Belo Horizonte: [s.n.], 2013.

Orientador: Prof. Leandro Ribeiro Negreiros.

Monografia (especialização) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

1. Metadados de assunto. 2. Legislação federal brasileira. 3. Sistema de Legislação Informatizada (Legin Web) - Câmara dos Deputados. 5. Informação – Granularidade. I. Negreiros, Leandro Ribeiro. II. Universidade Federal de Minas Gerais. Escola de Ciência da Informação. III. Título.



UFMG

Universidade Federal de Minas Gerais
Escola de Ciência da Informação
Núcleo de Informação Tecnológica e Gerencial – NITEG
Curso de Especialização em Arquitetura e Organização da Informação

ATA DA DEFESA DE MONOGRAFIA DE **ELAINE APARECIDA DE LIMA**, matrícula:
2011717447

Às 14:00 horas do dia 30 de abril de 2013, reuniu-se nas Instalações do CEFOR/Câmara dos Deputados em Brasília-DF a Comissão Examinadora, para julgar o trabalho intitulado **A identificação de metadados de assunto para a legislação federal brasileira: o caso do Sistema Legim Web da Câmara dos Deputados**, requisito final para obtenção do Grau de ESPECIALISTA em ARQUITETURA E ORGANIZAÇÃO DA INFORMAÇÃO. Abrindo a sessão, a Presidente da Comissão, Prof. Me. Leandro Ribeiro Negreiros, após dar conhecimento aos presentes do teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa da candidata. Logo após, a Comissão se reuniu sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

Prof. Me. Leandro Ribeiro Negreiros – Orientador

Profa. Dra. Gercina Borém de Oliveira Lima

Thiago Gomes Eirão

Pelas indicações, a candidata foi considerada *aprovada* com nota *100*.

O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a sessão, da qual foi lavrada a presente ATA que será assinada por todos os membros participantes da Comissão Examinadora.

Brasília, 30 de abril de 2013

Prof. Me. Leandro Ribeiro Negreiros
Escola de Ciência da Informação/UFMG (Orientador)

Profa. Dra. Gercina Borém de Oliveira Lima
Escola de Ciência da Informação/UFMG

Thiago Gomes Eirão
Representante do CEFOR/Câmara dos Deputados

Obs: Este documento não terá validade sem a assinatura e carimbo do Coordenador do Curso.

AGRADECIMENTOS

*Meus sinceros agradecimentos a todos que me proporcionaram realizar esta jornada,
me apoiaram e dividiram comigo este gratificante aprendizado.*

RESUMO

Procura apontar aspectos que devem ser considerados na identificação de metadados de assunto granulares para a legislação federal brasileira. O objeto de estudo foi o Sistema de Legislação Informatizada (Legin Web) disponível no Portal da Câmara dos Deputados. Os objetivos específicos foram: identificar tipos de assuntos amplamente utilizados na indexação da legislação federal brasileira e aspectos do contexto de busca de informação que interferissem na identificação dos metadados de assunto; analisar possibilidades de metadados de assunto para a legislação federal com base em padrões de metadados e modelos de organização da informação abordados na literatura; e, com isso, propor metadados de assunto para a legislação federal brasileira. A ideia é usar esses metadados para diminuir a imprecisão dos resultados das pesquisas na legislação federal, tornando o processo mais rápido e eficiente. O estudo consistiu em pesquisa descritiva, estudo de caso e usou triangulação. O método envolveu técnicas de levantamento para aferição de padrões de repetição de tipos de assunto da legislação federal brasileira e questionário para definição de tipos de pesquisa por assunto recorrentes na legislação. Os dados quantitativos usaram duas fontes: o campo de indexação do Legin e relatório fornecido pelo Centro de Informática (Cenin) da Câmara sobre tipos de pesquisa realizados por usuários internos e externos do Portal da Câmara. Foram utilizadas literatura técnica e científica sobre metadados, padrões de metadados, metadados de assunto, qualidade e metodologia de criação de metadados, e características temáticas da legislação federal brasileira. Os resultados obtidos consistiram nos aspectos a serem considerados na identificação de metadados de assunto, que corresponderam, resumidamente, a estudo por equipe multidisciplinar de padrões e funcionalidades de metadados, regras e controles atribuídos a metadados de assunto, características de estruturação temática da legislação, priorização da valoração dos dados conforme a comunidade usuária e tecnologias relacionadas à estruturação de metadados. Também foram apontados alguns metadados, assim como regras, controles atribuídos e formas de utilizá-los, com base na decomposição de estrutura e conteúdo da legislação federal brasileira.

Palavras-chave: Metadados de assunto. Legislação federal brasileira. Sistema de Legislação Informatizada (Legin Web). Câmara dos Deputados.

ABSTRACT

It highlights aspects that should be considered in identifying granular subject metadata to Brazilian federal law. The object of study was the Sistema de Legislação Informatizada (Legin Web) available on the website of the Câmara dos Deputados. Specific objectives were to: identify types of subjects widely used in indexing the Brazilian federal law and aspects of the context of information search that interfering with identifying subject metadata; analyze possibilities of subject metadata to federal legislation based on metadata standards and models of information organization addressed in the literature, and, therefore, propose subject metadata to Brazilian federal law. The idea is to use this metadata to reduce the imprecision of the results of research in the federal legislation, making the process faster and more efficient. The study consisted of descriptive research, case study and used triangulation. The method involved lifting techniques for measuring patterns of repetition of types of subjects of Brazilian federal law and questionnaire to define types of research by subject recurring in legislation. Quantitative data used two sources: the field of indexing and Legin report provided by the Centro de Informática (Cenin) of the Câmara on types of research conducted by both internal and external users of the Portal Câmara. It used technical and scientific literature on metadata, metadata standards, metadata subject, quality and methodology of creating metadata and thematic characteristics of the Brazilian federal law. The results consisted of the aspects to be considered in identifying metadata subject, which corresponded, briefly, to the study by a multidisciplinary team of metadata standards and features, rules and controls assigned to the subject metadata, thematic structuring features of the legislation, prioritization of value data as the user community and related technologies for structuring metadata. Also were highlighted some metadata, like rules, forms and controls assigned to use them, based on the decomposition of the structure and content of Brazilian federal law.

Keywords: Subject metadata. Brazilian federal law. Sistema de Legislação Informatizada (Legin Web). Câmara dos Deputados.

LISTA DE TABELAS

1 - Tipos de normas federais mais recorrentes e Emendas Constitucionais em relação ao total de registros do Legin Web (Total de registros = 225.843)

2 - Decomposição da estrutura de atos normativos federais

3 - Decomposição da estrutura do campo de indexação de atos normativos federais

4 - Decomposição do conteúdo do campo de indexação de atos normativos federais (assuntos sem repetição dentro do ato ou exclusivos)

5 - Usuários internos e externos da Câmara dos Deputados: tipos de buscas realizadas no Legin Web de 01/09/2012 a 30/09/2012

6 - Metadados de assunto para a legislação federal brasileira com base em estudo do Legin Web

SUMÁRIO

1	INTRODUÇÃO	10
1.1	OBJETIVOS GERAL E ESPECÍFICOS.....	14
2	REFERENCIAL TEÓRICO	16
2.1	OS METADADOS: DEFINIÇÕES E FUNCIONALIDADES.....	17
2.2	METADADOS DE ASSUNTO E PADRÕES DE METADADOS.....	28
2.3	CRIAÇÃO E QUALIDADE DE METADADOS.....	42
2.4	LEGISLAÇÃO FEDERAL: CONCEITOS E ESTRUTURA.....	47
2.5	BUSCAS DE INFORMAÇÃO NA WEB E BUSCAS DE LEGISLAÇÃO.....	53
3	MÉTODO DE PESQUISA	57
4	DISCUSSÃO DOS RESULTADOS	62
4.1	ANÁLISE DOS DADOS QUANTITATIVOS.....	62
4.2	ANÁLISE DOS DADOS QUALITATIVOS.....	72
5	CONSIDERAÇÕES FINAIS	80
	REFERÊNCIAS	87
	APÊNDICE A	92

1 INTRODUÇÃO

O Poder Legislativo federal brasileiro é exercido pelo Congresso Nacional, composto por duas Casas Legislativas: a Câmara dos Deputados e o Senado Federal, portanto um sistema bicameral. São funções típicas do Poder Legislativo legislar e fiscalizar, e atípicas, administrar e julgar, além da função primária de representar o povo, a Nação. A Câmara dos Deputados reúne os representantes do povo. (SANTOS; NÓBREGA NETO; CARNEIRO, 2007, p. 29 et seq.)

No exercício de suas funções, a Câmara dos Deputados consome e gera várias informações. Com o objetivo de melhor tratá-las e disseminá-las, começou um processo tardio de automação, por volta do final dos anos noventa, que foi movido pela necessidade de atender inúmeras demandas reprimidas de disponibilização de informações em meio digital, tanto para a consecução das atividades fins e meio da Câmara, como para o atendimento das demandas informacionais da sociedade.

Assim, a rápida proliferação de sistemas e a falta de planejamento de ações concatenadas, que evitassem a duplicação de esforços e olhassem para a Câmara como um todo, acabaram por deixar de lado o estudo mais aprofundado da essência das informações a serem disponibilizadas, o porquê de disponibilizá-las e também para quem, onde e como.

Além disso, no momento, a Câmara dos Deputados passa por um processo de perda de memória porque muitos servidores, que participaram do início dos trabalhos na instituição, estão se aposentando. Isso torna a atual conjuntura uma oportunidade única para a captação de informações importantes a respeito dos processos de trabalho, bem como das peculiaridades da gestão das informações da Câmara.

De fato, a forma como se deu a gestão da informação (ou a falta dela) para a automação na Câmara dos Deputados não resultou, necessariamente, em uma recuperação da informação menos onerosa, imprecisa e lenta. E apesar dos vários esforços realizados com o intuito de resolver essas deficiências, muitos problemas ainda persistem.

De qualquer forma, esse não é um cenário exclusivo da Câmara. Rosenfeld e Morville¹ (2002, apud MORESI, RAMOS e PRADO, 2010, p. 103) discutem o porquê da importância de uma Arquitetura da Informação (AI), sob a perspectiva da recuperação da informação e apontam os seguintes aspectos que devem ser considerados:

- o custo de encontrar a informação: “Quanto custará se cada empregado da sua organização passar mais de cinco minutos por dia para encontrar uma informação de que precisa?” e
- o custo de não encontrar a informação: “quantas más decisões são tomadas por dia na sua organização por que os empregados não encontram as informações de que necessitam?”.

Assim, a Câmara dos Deputados busca formas de rever seus sistemas e políticas de informação, sob a ótica de todas essas variáveis, com o objetivo de viabilizar a recuperação da informação de modo mais eficiente e rápido.

Especificamente, sobre a questão do ambiente informacional das organizações, Moresi, Ramos e Prado (2010, p. 102) afirmam que é possível enfatizá-lo por meio da *ecologia da informação*, que está baseada em como as pessoas criam, distribuem, compreendem e usam a informação, sendo que nesse contexto a tecnologia assume o papel de disseminadora de informações e dados, e não é um fator essencial na gestão da informação.

Segundo os autores, essa visão ecológica é subdividida em três ambientes (informacional, organizacional e externo) e a arquitetura da informação faz parte do ambiente informacional e constitui em “um guia para estruturar e localizar a informação dentro de uma organização”.

Moresi, Ramos e Prado (ibid., p. 103) também correlacionam a arquitetura da informação com metadados, quando dizem que a arquitetura da informação tem como foco a informação em todas as formas e tamanhos, como *web sites*, documentos, aplicativos, imagens e também os metadados utilizados para descrever e representar o conteúdo de objetos, como documentos, pessoas, processos e organizações, com o fim de tornar acessível o que já existe na organização.

Porém, a definição de metadados provavelmente mais conhecida, condizente com a descrição apresentada por Moresi, Ramos e Prado e citada

¹ ROSENFELD, L; MORVILLE, P. **Information architecture for the world wide web**. 2. ed. [s.l.]: O'Reilly & Associations, 2002.

por Almeida (1998, p. 2), de que “metadados podem ser entendidos como dados sobre dados, ou dados que descrevem dados” não é suficiente para dar a dimensão de todos os papéis hoje desempenhados pelos metadados.

Entre suas muitas atribuições, atualmente, os metadados são apontados na literatura como responsáveis por inúmeras funções que vão além de simplesmente descrever o dado como: permitir decisões quanto ao uso dos dados, levar a uma maior qualidade do que se está produzindo em termos de dados e informação, tornar explícitos e estruturar os dados com o fim de melhor recuperá-los.

As possibilidades propiciadas pelos metadados levam autores como Megrey (2007, p. 11) a afirmar que “comprometer-se com um padrão de metadados não é uma decisão trivial”. O referido autor ainda complementa alertando sobre a necessidade de uma análise preliminar que avalie os benefícios e desvantagens da escolha de um padrão de metadados.

No caso da Câmara, é possível observar em função do atendimento de muitas demandas informacionais ser custoso ou praticamente inviável, a necessidade de maior “granularidade” de metadados, ou número maior de metadados para representar documentos e informações.

Segundo Rosato² (2008 apud ALVES, SIMIONATO e SANTOS, 2012, p. 3), o termo “granularidade” é originário da Ciência da Computação e refere-se ao nível de detalhamento da descrição de um recurso informacional. Segundo os autores, a granularidade é dividida em dois níveis: “a granularidade fina (*fine-granularity*), que significa que a descrição apresenta um alto nível de detalhamento; e a granularidade grossa (*coarse-granularity*), que significa que a descrição possui um baixo nível de detalhamento”. Os autores também afirmam, com base em vários estudos, que o princípio do refinamento engloba, além da própria granularidade e vinculadas a ela, duas outras noções importantes:

O acréscimo de qualificadores que são elementos de metadados que qualificam e especificam o significado do atributo, parecidos com subcampos; e a adoção de esquemas de codificação externos

² ROSATO, F. **Equilíbrio e Granularidade de Serviços (II)**. *Post* publicado em 18 de julho de 2008 no Blog: Aquele blog de SOA. Disponível em: <<http://www.aqueleblogdesoa.com.br/2008/07/equilibrio-e-granularidade-de-servicos-ii/>>.

Acesso em: 20 ago. 2012.

(esquemas de valores de dados e conteúdo de dados), que padronizam a representação dos valores nos metadados.

Com o objetivo de verificar a dificuldade de uma pesquisa devido à “granularidade grossa” ou falta de detalhamento de metadados, foi realizada uma pesquisa experimental no Legin, a qual constatou a partir de uma busca no campo de indexação da referida base, pelo assunto “Getúlio Vargas”, que seriam necessários 5 minutos para concluir a pesquisa e selecionar manualmente todo o material que tratasse de “Getúlio Vargas”.

Dos 21 resultados encontrados, apenas seis tratavam do assunto procurado. Os outros se referiam a “Hospital Getúlio Vargas”, “Fundação Getúlio Vargas”, “Rádio Comunitária de Getúlio Vargas”, município de Getúlio Vargas, no Rio Grande do Sul, “Associação de Pais e Amigos dos Excepcionais de Getúlio Vargas, RS”. Ao contrário, se houvesse o metadado de assunto pessoa ou nome pessoal, o que significa um tratamento mais granular da informação, a pesquisa levaria somente 0,046³ segundos, ou o tempo de resposta do sistema, depois de elaborada a expressão de busca e processada a pesquisa, já que os resultados que não consistissem em nomes pessoais seriam excluídos automaticamente.

Dessa forma, um dos dois pressupostos deste trabalho é que o Legin apresenta problemas referentes à suficiência e relevância dos metadados de assunto utilizados para a legislação federal brasileira, o que impossibilita uma recuperação temática adequada aos propósitos de busca de informação existentes na Câmara dos Deputados.

O segundo pressuposto é a legislação federal brasileira tem características estruturais e temáticas que determinam os tipos de questões a serem realizadas nos seus respectivos sistemas de informação, assim como indicam os metadados necessários à obtenção das respostas a essas questões. Exemplificando, no contexto do orçamento, poderiam ser criados metadados específicos para tratar assuntos referentes a leis orçamentárias, diferenciando, assim, tais assuntos no caso de ocorrerem fora desse contexto de orçamento.

³ O tempo 0,046s foi informado no Portal da Câmara, acima da própria lista de resultados exibidos referentes à pesquisa sobre “Getúlio Vargas”, realizada no Legin Web. O número 0,046s constitui no tempo que o sistema usou para realizar a pesquisa e apresentar o resultado.

Assim, metadados de assunto mais granulares para a legislação federal brasileira serão o foco deste estudo. O objeto ou base para a análise será o Sistema de Legislação Informatizada (Legin Web), desenvolvido pela Câmara dos Deputados, o qual inclui a legislação federal e interna da instituição. A princípio, a Câmara e seu histórico de automação serão contextualizados, assim como os problemas relacionados à recuperação de assuntos no Legin, os quais justificaram este trabalho. Também será feita uma breve apresentação dos metadados na perspectiva da Câmara.

1.1 OBJETIVOS GERAL E ESPECÍFICOS

Em virtude dos pressupostos já expostos, é preciso analisar a legislação federal sob várias perspectivas, com o fim de rever os metadados de assunto utilizados até então, para saber quais aspectos foram utilizados na estruturação desses metadados, e em função das lacunas encontradas, definir os aspectos não considerados inicialmente na elaboração dos sistemas.

Aqui, os aspectos determinantes dos metadados de assunto serão tratados conforme as seguintes variáveis: o documento em si, as práticas de indexação utilizadas, o contexto de busca de informação e padrões de metadados e modelos conceituais de organização da informação geralmente presentes na literatura.

Em virtude disso, o objetivo geral deste estudo é apontar os aspectos que devem ser considerados na identificação de metadados de assunto para a legislação federal brasileira na Câmara dos Deputados. E com o intuito de atingir o objetivo geral foram definidos os seguintes objetivos específicos:

- a) Levantar aspectos que podem ser considerados na criação ou seleção de metadados de assunto para a legislação federal brasileira;
- b) Sugerir metadados de assunto com base em tipos de assuntos amplamente utilizados na indexação da legislação federal brasileira e na organização estrutural desses atos normativos;
- c) Propor metadados de assunto voltados para necessidades apontadas nos contextos ou situações de buscas por assunto na legislação federal brasileira, disponível no Portal da Câmara;

d) Analisar possibilidades de metadados de assunto para a legislação federal brasileira com base em padrões de metadados e modelos de organização da informação abordados na literatura.

Por fim, de modo macro este estudo visa contribuir para a Arquitetura da Informação de ambientes informacionais da Câmara dos Deputados, pois busca “tornar o complexo claro”, outra forma de definição de AI, proposta pelo desenhista e arquiteto gráfico Richard Saul Wurman (1997), que cunhou o termo em 1976. (REIS, 2007, p. 3)

2 REFERENCIAL TEÓRICO

Em relatório de 2003, sobre o impacto em bibliotecas, museus, arquivos e em organizações relacionadas de tendências impulsionadas pela Web, a *Online Computer Library Center (OCLC)* afirmou que os ambientes estão cada vez mais interconectados e que isso é resultado de quatro áreas de pressão: a diversidade e o número de sistemas das organizações de informações, a crescente tendência para acordos de compartilhamento de recursos em diversos níveis, e a necessidade de interagir com ambientes em outros sistemas. Essa constatação termina com a afirmação de que todos esses sistemas complexos precisam ser interoperáveis (OCLC, 2004, p. 15). No mesmo relatório, na sequência, a OCLC fala de arquitetura e serviços de rede:

À medida que o ambiente se torna mais complexo, vemos um distanciamento dos “fornos” de aplicativos em direção a uma decomposição de aplicativos, de forma que possam ser recombinaados para atender as novas necessidades com mais flexibilidade. Pense nisso como redirecionamento das arquiteturas para outra finalidade. Essa perspectiva apresenta os seguintes tipos de serviços: [...] de apresentação responsáveis por aceitar entradas dos usuários e produzir as saídas do sistema; [...] de aplicativos responsáveis por gerenciar transações entre componentes; de repositórios de conteúdo de dados e metadados e serviços comuns com potencial para serem compartilhados por diversos aplicativos. [...] Portanto, isso levanta a questão de garantir uma estrutura adequada de padrões para que isso aconteça.

Essa é uma forma de olhar para a Web, mas não só para ela, como para as próprias organizações que não deixam de ser sistemas semelhantes, em uma escala menor.

Os sistemas de informação da Câmara compõem um microcosmo parecido. Existem inúmeros na instituição, que operacionalizam atividades do Plenário, de Comissões, das áreas administrativas e até o gerenciamento do próprio processo legislativo. Só para citar, temos o Sistema de Informações Legislativas (Sileg), que inclui as proposições legislativas, o Sistema de Taquigrafia (Sitaq), com os discursos dos parlamentares, notas taquigráficas e mesmo o próprio Legin Web. Muitos dos metadados usados por esses sistemas são comuns, como no caso do metadado “Deputado”, que pode ser o

autor de uma proposição, de um discurso, e como o autor da proposição está também relacionado ao ato normativo originado dessa proposição.

Apesar disso, os metadados ainda estão em fase de padronização, então é difícil realizar o reaproveitamento de dados de um sistema para o outro, sem perdas, problemas de importação e custos, muitas vezes, altos. Isso porque, em princípio, os sistemas não foram construídos com base em tecnologias de estruturação de dados, que tivessem como fim a interoperabilidade e reutilização dos dados.

Assim, de modo geral, no caso dos sistemas de informação da Câmara, percebe-se a falta de uma análise mais profunda a respeito das informações prestadas pela Câmara ao seu público, assim como de um estudo teórico não só de padrões, mas dos fatores que contribuem para a criação de sistemas interoperáveis e para uma arquitetura da informação.

De fato, muitas bases da Câmara, assim como os metadados que as compõem, foram criados a partir de princípios pragmáticos de tratamento da informação, com o objetivo de atender demandas pontuais de ordem operacional, sem uma reflexão mais prospectiva em relação ao processo de produção e recuperação da informação. O referencial teórico deste estudo contribui neste sentido, pois busca um maior conhecimento sobre metadados, o sistema Legim Web e a legislação federal brasileira, e a comunidade usuária desse sistema via Portal da Câmara.

2.1 OS METADADOS: DEFINIÇÕES E FUNCIONALIDADES

A utilização de metadados para a descrição e recuperação de conteúdos informacionais não é novidade na Ciência da Informação. Na verdade, alguns padrões de metadados da área tiveram sua origem na década de 60. No entanto, foi a partir dos anos 90, com o crescimento acelerado e desordenado da Web, que o tema metadados ressurge com força, no sentido não só de garantir a recuperação, mas controlar o problema da “baixíssima precisão” de resultados da Web, o que, entretanto, não é uma questão exclusiva desse tipo de ambiente, caracterizado pelo compartilhamento e gestão descentralizada.

Sneiderman⁴ (1997 apud MARCONDES, 2006, p. 97) descreve o problema da falta de precisão de resultados na Web da seguinte forma:

A primeira tentativa de dar conta da explosão informacional em que se transformou a Web foram os catálogos, como o Yahoo (o primeiro catálogo da Web), e os chamados mecanismos de busca, como AltaVista, [...], e, mais recentemente, o Google. Enquanto em catálogos como o Yahoo, a descoberta, avaliação e descrição e inclusão dos recursos web na base de dados são feitas por profissionais de informação, os mecanismos de busca para indexarem a Web, possuem programas que visitam página por página [...] extraindo daí palavras-chave [...] associadas ao URL da página. [...] Naturalmente, por ser uma indexação automática com base em palavras isoladas, sem nenhum controle terminológico, efetuada em páginas sobre os mais variados assuntos, diferentes idiomas e totalmente desprovida de qualquer informação contextual, os resultados têm baixíssima precisão.

Tal problema pode ser visto como um dos motivos que levou a National Information Standards Organization (NISO) (2004, p. 1, tradução nossa) a afirmar que:

Uma razão para a criação de metadados é facilitar a descoberta de informação relevante. Além disso, metadados podem ajudar a organizar recursos eletrônicos, facilitar a interoperabilidade e integração de recursos legados, proporcionar identificação digital, e dar suporte ao arquivamento e preservação. [...] Metadados têm a mesma função na descoberta de recursos que uma boa catalogação por:

- Permitirem que os recursos possam ser encontrados por critérios relevantes;
- Identificarem recursos;
- Trazerem recursos semelhantes em conjunto;
- Distinguirem recursos diferentes;
- Darem informações de localização.

Na descrição acima, apesar de ficar claro que a descoberta dos recursos é o objetivo fim dos metadados, observa-se uma mudança em relação às funcionalidades atribuídas aos metadados, que antes basicamente remetiam à descrição e recuperação dos dados, e agora apontam outro papel chave dos metadados, a estruturação dos dados, com a finalidade de permitir a interoperabilidade entre sistemas e a consequente reutilização desses mesmos dados.

⁴ SNEIDERMAN, B. Clarifying search: a user-interface framework for text searches. **D-Lib Magazine**, jan. 1997.

Assim, surgem muitas definições de metadados, que tornam pouco representativa a famosa “dados sobre dados” (Weibel⁵, 1995 apud MARCONDES, 2006, p. 97). De fato, aparecem definições que acrescentam outras funcionalidades aos metadados como a da NISO (op. cit., p. 1), adotada neste estudo, devido ao seu escopo amplo, a qual conceitua metadados como “informação estruturada que descreve, explica, localiza, ou então, facilita a recuperação, o uso ou o gerenciamento de uma fonte de informação” (tradução de Gleisy R. B. Fachin). A própria NISO ainda acrescenta que o termo “metadados” é usado de formas distintas em comunidades diferentes. A respeito disso afirma:

Alguns usam para se referir à informação compreensível por máquina, enquanto outros usam apenas para registros que descrevem recursos eletrônicos. Em ambiente de biblioteca, metadados são normalmente usados para qualquer esquema formal de descrição de recursos, aplicável a qualquer tipo de objeto, digital ou não. (tradução nossa)

Ainda no mesmo estudo, a NISO relata que existem três tipos principais de metadados:

- Metadados descritivos descrevem um recurso para fins de descoberta e identificação. Eles podem incluir elementos como título, resumo, autor e palavras-chave.
- Metadados estruturais indicam a composição ou forma de estruturação dos objetos, por exemplo, como páginas são ordenadas para formar capítulos.
- Metadados administrativos fornecem informações para ajudar a gerenciar um recurso, tais como quando e como ele foi criado, tipo de arquivo, outras informações técnicas e quem pode acessá-lo. Existem vários subconjuntos de dados administrativos, dois que, às vezes, são listados como tipos de metadados diferentes são:
 - Metadados de gestão de direitos (*Rights management metadata*), que lidam com direitos de propriedade intelectual, e
 - Metadados de preservação (*Preservation metadata*), que contêm informações necessárias para arquivar e preservar um recurso.
 (tradução nossa)

Em um estudo amplo, Campos (2007, p. 17) aponta dez categorias funcionais dos metadados e as correlaciona com algumas tecnologias: catalogação digital (*Machine Readable Cataloging Record - MARC e Dublin Core*), XML (*Extensible Markup Language*), RDF (*Resource Description Framework*), XMLS (*XML Schema*), RDFS (*RDF Vocabulary Description*

⁵ WEIBEL, S. Metadata: the foundations of resource description. **D-Lib Magazine**, jul. 1995.

Language Schema), ontologias, *data warehouses*, Web Semântica, Serviços Web e outras. As categorias apontadas foram assim nomeadas: descritividade, padronização, interoperabilidade, modularidade, reflexibilidade, visibilidade, flexibilidade, automatização, administração e preservação.

Algumas das categorias funcionais mencionadas pelo autor são autoexplicativas e outras precisam ser explicitamente definidas. Assim, seguem esclarecimentos a respeito das dez categorias funcionais dos metadados propostas por Campos, com base em suas próprias considerações (op. cit., p. 21 et seq.):

- 1) Descritividade: função principal dos metadados, que é descrever o recurso ou objeto informacional propiciando a identificação, localização, recuperação, manipulação e uso dos dados. O autor ainda coloca que o “entendimento da função descritiva do metadado pode ser refinado a partir da identificação da característica que está sendo retratada. Descrições de contexto, estrutura e apresentação possibilitam a implementação de várias funcionalidades dos metadados”;
- 2) Padronização: “condição necessária para o entendimento das representações (estruturas, convenções, sintaxe e significado consensual dos metadados)”;
- 3) Interoperabilidade: “habilidade para transferir e utilizar informações entre sistemas com eficiência e uniformidade, exigindo padronização e flexibilidade em certo nível”;
- 4) Modularidade: trata das características de conteúdo, contexto, estrutura e apresentação de um objeto digital, as quais permitem a representação em bloco do objeto, por meio de dimensões e módulos. Segundo o autor “o ganho em praticidade é evidente e está vinculado à utilização de metadados direcionados para cada um desses módulos”;
- 5) Reflexibilidade: é um “ato de auto-orientação e descrição dos objetos informacionais” que permite maior visibilidade dos próprios objetos e versatilidade no seu uso. Está ligada a ideia do uso de metadados para descrever os contextos relacionados aos próprios sistemas, regras e processos que representam, organizam e tratam esses objetos. A reflexibilidade cria condições para a flexibilização e automatização de usos presentes e potenciais

dos objetos e relaciona-se com a documentação, confiabilidade, facilidade de manutenção e interoperabilidade;

6) Visibilidade: relacionada à capacidade dos metadados de estabelecer vínculos visíveis com o modelo de negócio, facilitando a navegação e o planejamento de consultas frente às necessidades dos usuários das áreas de negócio, que geralmente não possuem perfis técnicos que permitam a compreensão detalhada dos dados;

7) Flexibilidade: é um modo de lidar com a diversidade de estruturas e padrões e abrandá-la, com o fim da interoperabilidade e automatização dos processos. A flexibilidade também mantém uma ligação com a padronização. “Uma padronização total não seria viável e tenderia a reduzir a criatividade e inovação. Os metadados são intensamente explorados, em parte, devido à necessidade de padronizar mantendo a flexibilidade”;

8) Automatização: se for considerada a conceituação de ontologias como sistemas de metadados (ou vocabulários de metadados), constituindo a “artéria central de metadados para a Web semântica”, proposta por Greenberg⁶ et al. (2003 apud CAMPOS, 2007, p. 32) pode-se dizer que os metadados permitem a automatização de sistemas inclusive em nível semântico;

9) Preservação: metadados são utilizados em ambientes digitais para a preservação lógica, relacionada à garantia da conversão dos formatos originais, e para a preservação intelectual ligada à integridade e autenticidade das informações;

10) Administração: metadados revelam “as condições corretas ou ideais de seu gerenciamento, as relações do objeto com outros na coleção, sua função, utilização, comportamento, contexto de criação e condições de preservação”.

Outras funcionalidades dos metadados também podem ser observadas na literatura. Marcondes (2006, p. 99) afirma que a Web suscitou a necessidade de criação de um padrão de metadados que propiciasse uma descrição simples o suficiente para que um autor pudesse descrever seu documento, o que fez da simplicidade uma característica importante de um conjunto de metadados.

⁶ GREENBERG, J.; SUTTON, S.; CAMPBELL, D. Metadata: a fundamental component of the semantic web. **Bulletin of the American Society for Information Science**, v. 29, n. 4, 2003.

Mais questões surgem em paralelo ao crescimento das funcionalidades atribuídas aos metadados, como o aumento da preocupação com a criação ou escolha de um padrão de metadados e com a própria definição de metadados. Dessa forma, Gill⁷ (2000) e Vellucci⁸ (1998) (apud CAMPOS, 2007, p. 19) afirmam, respectivamente, que a definição de metadados, “dados sobre dados” tem “levado à utilização do termo metadados onde o termo dados teria sido anteriormente considerado suficiente (como descrições de pessoas ou objetos)”, e à ideia equivocada de que “metadados são simplesmente dados catalográficos ou bibliográficos com outro nome”.

Em função disto, aparecem outras definições e mesmo acepções de metadados na literatura que valorizam o contexto e o usuário do dado e/ou informação. Entre elas, temos a de Almeida (1999, p. 2).

Os metadados, de um modo simplificado, além de conterem as informações básicas para a pesquisa, apresentam também, estruturadamente, numa comparação simplificada, as informações do índice analítico. O usuário, de posse dessas informações pertinentes pode, de uma só vez, e de uma forma padronizada, avaliar os dados à luz de suas próprias necessidades. Assim, a estrutura de metadados, além de substituir a ação física da busca por uma pesquisa qualitativa da informação no computador do usuário, onde ele estiver (processamento realmente distribuído), representa ainda um ganho na quantidade e qualidade da informação relacionada aos dados.

E o autor ainda acrescenta “vê-se qualquer que seja o viés pelo qual se analisa o conceito de metadados, todos desembocam numa clareza e qualidade maior do conjunto de dados, implicando, em última análise, num ganho para a própria Sociedade”.

Consolidando várias definições, Vellucci⁹ (1998 apud CAMPOS, op. cit., p. 19) conceituou metadados digitais como:

[...] dado que descreve atributos de um recurso, caracteriza suas relações, apóia sua descoberta e uso efetivo, e existe em um ambiente eletrônico. Usualmente consiste em um conjunto de elementos, cada qual descrevendo um atributo do recurso, seu gerenciamento, ou uso.

⁷ GILL, Tony. **Metadata and the World Wide Web**. 2000.

⁸ VELLUCCI, S. L. Metadata. **Annual Review of Information Science and Technology**, v. 33, 1998.

⁹ VELLUCCI, S. L. Metadata. **Annual Review of Information Science and Technology**, v. 33, 1998.

A Web profunda ou *deep Web*, invisível aos mecanismos de busca, também contribuiu para a reformulação do significado de metadados. Conforme Marcondes (2006, p. 98), a Web profunda constitui nos objetos informacionais que estão disponíveis por meio de páginas geradas *on-the-fly* (dinamicamente) pelos sistemas de banco de dados, como resultado de uma consulta específica, só para o usuário que fez a consulta, e não como arquivo permanentemente disponível em um servidor. Isso impulsionou o surgimento de bibliotecas digitais, as quais passaram a disponibilizar os metadados dos objetos informacionais ou associados a sua URI (*Uniform Resource Identifier*) ou URL (*Uniform Resource Locator*).

Neste contexto, os metadados podem ser entendidos não só como solução de recuperação, mas como parte imprescindível do tratamento dos dados disponibilizados. Por si só, o tratamento dos dados remete a outras concepções de metadados. Ferreira¹⁰ (1976 apud ALMEIDA, 1999, p. 2) cita “o prefixo meta, do grego *méta*, significa, entre outras coisas, além, transcendência e reflexão crítica sobre”. O próprio Almeida depois conclui:

Todos estes significados são verdadeiros e representam, cada um a seu modo, uma dimensão do conceito de metadados:

- além ou que transcendem indica a própria natureza dos metadados: uma descrição completa dos dados e do que eles representam, permitindo decisões quanto a seu possível uso;
- uma *reflexão crítica* representa o desejo intrínseco de se induzir um maior conhecimento sobre o que significam os dados, conduzindo, automática e subliminarmente, a um cuidado maior, ou em termos mais atuais, a uma qualidade maior do que se está produzindo, a partir de considerações sobre a aplicabilidade e a validade do conhecimento adquirido.

Os metadados ainda estão sempre associados à ideia de conjunto, ou conjuntos de metadados, remetendo tanto à interoperabilidade, como à granularidade e à reutilização de dados. Vale ressaltar que isso foge, muitas vezes, da percepção de metadados dos profissionais de informação, que tendem a enxergar, conforme Zeng (2010, p. 187) cita, um registro de metadados como uma unidade básica. O autor afirma:

¹⁰ FERREIRA, A. B. de H. **Novo dicionário da língua portuguesa**. Rio de Janeiro: Nova Fronteira, 1976.

Similar à montagem de um quebra-cabeça, os tipos diferentes de elementos de metadados (descritivo, administrativo, técnico, de uso e de preservação) e de diferentes esquemas, vocabulários e aplicações podem ser combinados entre operações. Um registro de metadados [...] deve ser visto como um conjunto de descrições. Por esta razão, a combinação de descrições de metadados deveria ser racional e executável.

Marcondes (2006, p. 98) também trata de “conjuntos de metadados”. O autor menciona que um mesmo documento eletrônico pode ser referenciado por conjuntos de metadados de diferentes instituições, os quais apresentem diferentes finalidades, além da descrição, como melhorar a recuperação, preservação digital dos documentos e questões referentes a direitos autorais.

Também é ponto passivo na literatura que padrões de metadados estão sendo desenvolvidos por uma variedade de comunidades para fins específicos. Zeng (op. cit., p. 177), assim como a NISO (2004, p. 1 et seq.) citam vários exemplos de padrões de metadados, entre eles:

- a) *Dublin Core Metadata Element Set* (DCMES) – Conjunto de Elementos de Metadados Dublin Core;
- b) *Dublin Core Metadata Terms* – Termos de Metadados Dublin Core;
- c) *Electronic Theses and Dissertations Metadata Standard* (ETD-MS) – Padrão de Metadados de Teses e Dissertações Eletrônicas;
- d) *Learning Object Metadata* (LOM) – Metadados de Objetos de Aprendizagem;
- e) *Metadata Object Description Schema* (MODS) – Esquema de metadados descritivos derivado do MARC 21 que usa o XML. No entanto, os elementos MODS são mais fáceis de aplicar do que o MARC 21, mais ricos que o Dublin Core, e os nomes dos campos ou parágrafos são expressos em palavras, ou baseados na linguagem e não em números, como no MARC 21. Devido a usar XML, o MODS ainda tem a possibilidade de utilizar um atributo ID (identificador unívoco) opcional para facilitar a ligação no nível de elemento/metadado;
- f) ISO/TS 19115:2003 *Geographic information Metadata* – Metadados de Informação Geográfica;
- g) *Federal Geographic Data Committee* (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM), oficialmente conhecido como FGDC-STD-001-1998 – Metadados de informação geoespacial;
- h) *Text Encoding Initiative* (TEI) - é um projeto internacional para desenvolver diretrizes para marcação de textos eletrônicos como romances, peças de teatro

e poesia, para dar suporte principalmente à investigação nas ciências humanas;

i) *Metadata Encoding and Transmission Standard (METS)* - é um esquema XML para a criação de documentos XML, que expressam a estrutura de objetos complexos de bibliotecas digitais, os metadados descritivos e administrativos associados, e os nomes e localizações dos arquivos que contêm o objeto digital. *Metadados estruturais* são necessários para assegurar que arquivos digitalizados separadamente (por exemplo, diferentes páginas de um livro) sejam estruturados de forma adequada. *Metadados técnicos* fornecem informações sobre o processo de digitalização para que os estudiosos possam determinar quão precisa é a reflexão do original proporcionada pela versão digital. Outros metadados técnicos são necessários para fins internos, como para realizar periodicamente a atualização e migração dos dados, assegurando a durabilidade dos recursos;

j) *Encoded Archival Description (EAD)* - foi desenvolvido como uma forma de marcação dos dados contidos em instrumentos de pesquisa, de modo que eles possam ser pesquisados e exibidos on-line. Em arquivos e coleções especiais, o instrumento de pesquisa é uma importante ferramenta para descrever recursos;

k) *<indec> Framework (Interoperability of Data in ECommerce Systems)* - é um modelo de dados para propriedade intelectual e sua transferência;

l) *Categories for the Descriptions of Works of Art (CDWA)* - é um modelo conceitual para descrever e acessar informações sobre objetos e imagens;

m) *VRA Core Categories* - foi expandido com base no trabalho do CDWA, definiu um único conjunto de elementos de metadados que pudessem ser usados para descrever o trabalho (a fotografia real, pintura, escultura, construção, etc), bem como suas imagens (representações visuais) deles;

n) *MPEG-7, Multimedia Content Description Interface (ISO/IEC 15938)* - define os elementos de metadados, estrutura e relacionamentos que são utilizados para descrever objetos audiovisuais, incluindo ainda pinturas, gráficos, modelos 3D, música, áudio, fala, vídeo ou coleções multimídia;

o) *MPEG-21, Multimedia Framework (ISO/IEC 21000)* - desenvolvido em resposta à necessidade de uma abrangente estrutura para garantir a interoperabilidade de objetos digitais multimídia.

A NISO (ibid., p. 2, tradução nossa) relata que a estruturação de metadados para fins específicos, ou mesmo voltados para conjuntos de dados, permitiu que as ciências pudessem compartilhar seus dados “de uma maneira que nunca foi possível antes da Internet”, a respeito disso ela diz:

Esquemas de metadados [conhecidos como ‘*Schema*’] são conjuntos de elementos de metadados concebidos para um determinado fim, tal como a descrição de um tipo particular de recurso de informação. A definição ou significado dos próprios elementos é conhecida como a semântica do esquema. Os valores dados aos elementos de metadados são o conteúdo. Esquemas de metadados geralmente especificam nomes de elementos e sua semântica. Opcionalmente, podem especificar as regras de conteúdo, ou regras para especificação ou formulação do conteúdo (por exemplo, como identificar um título principal), as regras de exibição do conteúdo (por exemplo, regras de acumulação), e valores de conteúdo permitidos (por exemplo, os termos utilizados devem ser retirados de vocabulário controlado especificado). Podem existir também regras de sintaxe indicando como os elementos e seu conteúdo devem ser codificados. Os esquemas sem regras de sintaxe prescritas são chamados de esquemas de sintaxe independente.

Marcondes (2006, p. 96) fala que um dos maiores objetivos do uso de metadados na Web é permitir não só a descrição de recursos de informação eletrônicos, possibilitando assim sua avaliação de relevância por usuários humanos, “mas também permitir agenciar computadores e programas especiais, robôs e agentes de software, para que eles compreendam os metadados associados aos documentos e possam então recuperá-los, avaliar sua relevância e manipulá-los com mais eficiência”.

O autor também cita que neste ponto entra a linguagem XML, como “a forma que tem se firmado como um padrão na Web para codificar metadados”. A respeito disso, Zeng (op. cit., p. 178, grifo nosso) cita:

Durante a evolução da nossa era digital, a XML (Extensible Markup Language - Linguagem de Marcação Extensível) – desenvolvida por um grupo de trabalho em XML, anteriormente formado sob os auspícios da World Wide Web Consortium (W3C) em 1996 (continuamente atualizado) – se transformou na *metalinguagem oficial de muitas comunidades*. Simplificando, a XML “descreve uma classe de objetos de dados chamados documentos XML e, parcialmente, descreve o comportamento de programas de computador que os processam”. As marcações codificam uma descrição do layout de armazenagem do documento e a estrutura lógica. A XML fornece um mecanismo para impor restrições no layout de armazenamento e na estrutura lógica. O W3C não criou uma

linguagem de marcação baseada no domínio, ao contrário, *deixou para os criadores específicos de cada domínio fazerem isso*. Assim, as linguagens de marcação específicas por domínio também foram desenvolvidas por várias comunidades, particularmente em áreas científicas.

Marcondes (op. cit., p. 102) ainda acrescenta que a linguagem XML tem como proposta ser um padrão universal:

Voltada para a descrição da estrutura de documentos eletrônicos em que o conteúdo e sua descrição compõem um único arquivo, de tipo texto puro – portátil, portanto sem problemas para qualquer tipo de sistema de computador [...] Ao contrário de outros formatos de arquivos, os chamados formatos proprietários [...] é explicitada, além dos conteúdos dos documentos, a sua estrutura, sob a forma de marcações, identificadas por um nome significativo, envolvido nos caracteres “<” e “>”; a cada marcação indicando o início de um determinado conteúdo, por exemplo, <autor>, corresponde a uma marcação de fim, </autor>; os conteúdos ficam entre marcações, <autor> João Silva </autor>.

A descentralização da criação de padrões de metadados e de linguagens de marcação por domínio, conforme comunidades determinadas ou finalidades específicas fica evidente na seguinte colocação de Campos (2007, p. 35):

À medida que a utilização da informação em ambientes digitais sofisticada-se, é esperado que haja uma ampliação no escopo das funcionalidades dos metadados. A direção que seu desenvolvimento tomará está fortemente relacionada às características da comunidade que os utiliza e às tarefas a serem desempenhadas.

Outro problema que se coloca aos metadados está associado à tempestividade das informações. A respeito disso, Marcondes¹¹ (2001 apud MARCONDES, op. cit., p. 96) diz “a informação relevante para um dado problema tem que estar disponível no tempo certo. De nada adianta a informação existir se quem dela necessita não sabe da sua existência ou se ela não puder ser encontrada”.

O contexto envolto aos metadados foi descrito pela OCLC (2004, p. 16) em seu relatório sobre análise de cenários para a identificação de padrões sobre o gerenciamento de conhecimento e informação. Neste relatório, a OCLC concluiu que:

¹¹ MARCONDES, C. H. Representação e economia da informação. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 61-70, 2001.

Três padrões se destacam entre muitos [...]. Um deles é uma redução no acesso guiado ao conteúdo. O segundo é uma tendência à desagregação, não só de conteúdo, mas também de serviços, da tecnologia, da economia e das instituições. O terceiro padrão é o da colaboração: jogos, *software* de código aberto, conferências na Web, *blogs*, mensagens instantâneas, objetos de aprendizado [...].

Portanto, são muitos os desafios de se criar ou escolher um conjunto ou conjuntos de metadados e devem ser levados em consideração o contexto dos dados e informações, as questões dos usuários no referente ao desenvolvimento de suas atividades, as próprias finalidades dos metadados e os padrões de metadados disponíveis.

2.2 METADADOS DE ASSUNTO E PADRÕES DE METADADOS

Metadados de assunto para efeitos deste estudo são aqueles referentes a termo ou conjunto de termos que representam o conteúdo temático do documento, o que é conhecido, na Biblioteconomia, como "representação temática" ou indexação de assuntos de documentos/objetos/recursos, não importando se esses termos provêm de vocabulário controlado ou não.

A indexação de assuntos, de acordo com Lancaster (2004, p. 1 et seq.), consiste na preparação da representação do conteúdo temático dos documentos por meio de termos. Esses termos atribuídos pelo indexador, ou, em outras situações, pelo produtor da informação, servem como pontos de acesso, mediante os quais um item é localizado e recuperado, durante uma busca por assunto. Quando os termos utilizados realmente aparecem no documento, a indexação é chamada de *indexação por extração*, quando os termos julgados relevantes para indexar o documento são extraídos de outras fontes, é *indexação por atribuição*.

Ainda segundo o autor, os termos podem ser simples ou compostos e o ideal é que os itens indexados possam ser recuperados tanto por meio dos termos isolados, na sua forma simples ou composta, como sob qualquer combinação deles. Ele também afirma que os termos utilizados na indexação podem ser designados como 'termos de indexação', apesar de ser também empregada a palavra 'descritores', em especial no caso dos termos extraídos

de um tesouro. Em geral, os termos provêm de algum tipo de vocabulário controlado, mas também podem ser termos 'livres', selecionados a partir do próprio documento.

No momento, a Câmara dos Deputados desenvolve seu próprio vocabulário controlado, o Tesouro da Câmara dos Deputados (TECAD), o qual já é utilizado. Contudo, ele não será objeto deste trabalho, ainda que tenha uma importante relação com os metadados de assunto disponíveis no Legin Web, em princípio, porque esse alinhamento entre metadados de assunto e TECAD é um outro estudo, o qual exige, por sua própria natureza, a criação de uma equipe multidisciplinar.

Outro motivo para a opção por essa linha de discussão trata da funcionalidade dos metadados denominada por Campos (2007, p. 31) "flexibilidade", citada anteriormente. Segundo o autor metadados são muito explorados com objetivo de padronizar, sem perder a flexibilidade, o que cria condições para a inovação dos sistemas. Assim, é interessante fazer uma análise, dos recursos propiciados por metadados de assunto, no caso da legislação federal, com todas as suas peculiaridades, independente dos recursos oferecidos pelo TECAD, que, como vocabulário controlado, necessita seguir padrões mais rigorosos.

Isso não quer dizer que não ocorram equivalências entre tipos de metadados de assunto relacionados à legislação federal e tipos de categorias de assuntos adotadas pelo TECAD. De fato, essas equivalências precisam ocorrer, mas isso não será uma preocupação deste estudo. Casos que justificam a manutenção dessa independência entre vocabulário controlado e metadados de assunto é o dos "termos candidatos a descritores" e mesmo dos "termos livres", que constituem muitas vezes a solução de metadados encontrada para o rigor dos vocabulários controlados, que têm sua elaboração apoiada "na Lingüística Estrutural, Semiótica, Terminologia e Lógica Formal". (Tálamo e Lara¹², 2006 apud VOGEL, 2007, p. 2).

No que refere a "termos candidatos a descritores" é necessário observar que se diferenciam dos livres, pois podem vir a se tornar um termo

¹² TÁLAMO, M.F.G.M; LARA, M.L.G. O campo da lingüística documentária. **TransInformação**, Campinas, v. 18, n. 3, p.203-211, set./dez. 2006.

autorizado do vocabulário controlado. Já os termos livres, em princípio, não preenchem os requisitos para se tornarem termos controlados.

Segundo Cunha e Cavalcanti (2008, p. 360), “termo candidato a descritor” é uma palavra, ou grupo de palavras, propostas para inclusão eventual em um tesouro”. Os autores definem também “termo controlado”, como “palavra ou expressão (locução) que indica um conceito em uma linguagem de indexação”. Assim, “termo livre” ou “não controlado” também pode ser entendido como aquele usado para indexar, mas que não faz parte de uma linguagem de indexação.

Existem outros tipos de metadados que podem ser considerados de assunto ou temáticos, principalmente sob a ótica de pessoas que não são da área de informação, mas não são considerados assim aqui. Um exemplo é o metadado de resumo do conteúdo de um objeto informacional, que pode ser entendido como uma forma de representação temática do conteúdo, porém não constitui uma representação baseada em termos, mas em construções sintáticas ou sintagmas¹³.

A indexação por meio de sintagmas seria complexa no caso de atos normativos topicalizados em dispositivos (artigos, parágrafos, alíneas, etc.), nos quais o conteúdo está muito disperso ao longo de atos normativos que chegam a ter mais de 1000 artigos. Isso sem considerar que o texto de uma norma deve ser extremamente criterioso e exato¹⁴, o que levanta a questão de como transpor essas características para um resumo, sem tornar esse processo oneroso em termos de custos humanos e operacionais.

E também no caso do resumo, é preciso dizer que não só a forma é outra, como os recursos disponíveis para tratar o assunto também são outros. Uma segunda razão para essa separação entre esses tipos de metadados de tratamento de conteúdo é a complexidade agregada ao processo de indexar e recuperar objetos informacionais se os metadados de assunto e resumo fossem considerados, hipoteticamente, como formas complementares de tratamento de assunto. Lancaster (2004, p. 7) discorre sobre isso:

¹³ Segundo Dicionário Houaiss Eletrônico (2009), sintagma é uma “unidade linguística composta de um núcleo (p.ex., um verbo, um nome, um adjetivo etc.) e de outros termos que a ele se unem, formando uma locução que entrará na formação da oração”.

¹⁴ Segundo Dickerson (1954 apud CARVALHO, 2010, p. 3) “redigir leis é a forma de expressão mais rigorosa depois da matemática”. – DICKERSON, R. **Legislative drafting**. Boston: Little, Brown and Company, 1954.

Se a indexação e a redação de resumos fossem consideradas como atividades inteiramente complementares, [...] o indexador se concentraria na atribuição de termos que complementassem os pontos de acesso existentes no resumo. Tal complementaridade, porém, deve ser inteiramente reconhecida e compreendida pelo usuário da base de dados. Do contrário, um conjunto de termos de indexação isolados daria uma imagem bastante equivocada do conteúdo de um item.

Também podemos compreender melhor a acepção diferenciada entre metadados de assunto e metadados de resumo olhando para muitos padrões de metadados existentes, os quais também fazem essa distinção.

Dessa forma, com base na definição de metadados de assunto proposta no estudo, em que metadados de assunto são somente aqueles que contêm como dados termos de indexação controlados ou livres, foram identificados alguns padrões de metadados que partiam desse princípio. Também foram analisados padrões de metadados reconhecidos por tratarem, em nível granular maior, metadados de assunto, ou ainda por estarem diretamente relacionados ao tratamento de assunto de informações governamentais, entre as quais se encontra a legislação federal.

Um padrão que remonta ao início dos estudos de metadados na Ciência da Informação é o *Machine Readable Cataloging Record* (MARC), mais precisamente, década de 60. O MARC sofreu alterações ao longo do tempo e até mesmo alguns países adotaram suas próprias versões, por exemplo USMARC (Estados Unidos) e CAN/MARC (Canadá). Mas no início de 1998 as diferenças são compatibilizadas e desse alinhamento surge o MARC 21. (LIBRARY OF CONGRESS, 1998)

O MARC 21¹⁵ foi originalmente concebido para tratar basicamente informação bibliográfica, bem como necessidades decorrentes desse tipo de informação e dividiu-se em cinco formatos: **bibliográfico, de autoridade, de coleção (*holdings*), de classificação e de comunidade**. Os formatos MARC 21 são amplamente utilizados para representação e também para o intercâmbio de registros via sistemas automatizados, o qual segue as rotinas estabelecidas pelo *Format for Information Exchange* (ISO 2709).

¹⁵ O MARC 21 designa códigos e convenções, *tags* ou parágrafos, dois indicadores, subcampos e valores de códigos que identificam os elementos de dados no registro bibliográfico.

Todavia, é preciso dizer que o MARC 21 é criticado por ser um formato de metadados bastante especialista, que necessita para um aproveitamento eficiente de seus recursos que o usuário tenha conhecimentos a respeito não só do próprio formato, mas também de outros padrões como o *International Standard Bibliographic Description* (ISBD), o *Anglo-American Cataloging Rules* - segunda edição (AACR2), o qual está em revisão para ser substituído pelo *Resource Description and Access* (RDA), o *Library of Congress Subject Headings* (LCSH) e outros.

Entretanto, o MARC 21 por sua complexidade e própria história de desenvolvimento tem um alto poder de previsibilidade de metadados. De fato, a complexidade agregada ao formato não descaracteriza seu poder de previsibilidade, não são coisas antagônicas. Inclusive, nota-se que o MARC 21 tem caminhado no sentido de tornar o formato menos complexo, sem perder sua riqueza de detalhes, em artigo que trata da conversão do MARC21 para um padrão de descrição bibliográfica semântica, o *MarcOnt Initiative*. A respeito disso, Castro e Santos (2009) colocam:

Um dos motivos desta pesquisa, e na busca por subsídios a fim de caracterizar o MarcOnt, está pautado na tese que Synak (2005) procura defender com tanta propriedade, ao mencionar que o MARC21, mesmo sendo um padrão mundial e internacionalmente aceito e adotado pelas bibliotecas e pela comunidade da biblioteconomia, possuindo uma estrutura completa e complexa, não tem um padrão de descrição semântica que contemple a profundidade da descrição, ou ainda, que cubra a extensão requerida ou suportada por ele. Pensando nisso, pode-se considerar que o padrão de descrição bibliográfica semântica MarcOnt pode possibilitar uma melhora na explicitação das informações, permitindo sua compreensão por máquinas nos atendimentos das solicitações de busca realizadas pelo usuário final.

A previsibilidade ou completude apresentada pelo MARC 21 é uma característica importante de um padrão de metadados. Pode-se dizer que um padrão com maior grau de previsibilidade, porém que tente ser mais simples, diminuindo características que agreguem complexidade desnecessária, garante a contemplação de necessidades mais esporádicas de metadados, porém não menos importantes, sem grandes modificações de sistemas.

Isso ainda remete ao tratamento mais granular da informação, o que proporciona maior interoperabilidade entre esquemas de metadados, se considerarmos a citação da NISO (2004, p. 11) de que “o mapeamento de

esquemas com poucos elementos (menor granularidade) para aqueles com mais elementos (maior granularidade) é problemático”. Assim, conclui-se que o tratamento mais granular da informação facilitará a interoperabilidade entre esquemas de metadados.

Outra questão são os próprios metadados de assunto, que são vistos, segundo ZENG (2010, p.181), como “um elemento universal de quase todos os padrões de metadados”. Portanto, apesar da legislação possuir características muito peculiares, inclusive em relação ao jargão jurídico, os metadados de assunto da legislação podem ser espelhados de metadados de assuntos já efetivamente utilizados em outros padrões, desde que correspondam a assuntos também presentes na legislação federal brasileira.

No caso do MARC 21, existe dentro de alguns metadados de assunto uma série de subdivisões de assunto, geralmente para uso das instituições que utilizam “cabeçalhos de assunto”, um tipo de vocabulário controlado pré-coordenado¹⁶. O Legin, assim como outros sistemas que tratam a legislação federal brasileira, entre eles, o Sicon do Senado Federal e a base de legislação do Planalto, utilizam linguagens de natureza essencialmente pós-coordenada, como tesauro.

Apesar da norma ISO 2788, que trata da construção de tesouros monolíngües, ser clara ao dizer que a utilização de tesauro não implica necessariamente o uso de pós-coordenação. Conforme a ISO 2788 (1986, p. 2, tradução de Bianca Amaro de Melo) “dentro do possível, as técnicas descritas nessa norma não estão limitadas a um método particular de indexação, seja pós-coordenado ou pré-coordenado”. A norma também exemplifica e define os métodos conforme segue.

Um indexador que trabalha com um documento sobre “Computadores em bancos de Amsterdam” pode atribuir três termos ao documento: “Banco”, “Computadores” e “Amsterdam”. Em um sistema pós-coordenado, a relação entre esses termos não é explícita, e o documento seria recuperado se um ou todos esses termos fossem utilizados como chave de recuperação. Em um índice pré-coordenado, as relações entre os termos podem ser indicadas de várias maneiras; por exemplo, através de símbolos que expressam relações específicas, as posições dos termos na entrada, sua tipografia e/ou pontuação. Os termos deste exemplo não se associam de acordo com as estruturas de referência comuns, e, portanto, suas relações podem ser consideradas como dependentes do documento.

¹⁶ Conceitos de pós-coordenação e pré-coordenação serão explicados logo abaixo.

As colocações acima sobre o tipo de indexação adotada têm um impacto na definição dos tipos de metadados de assuntos utilizados, principalmente no caso da adoção de subdivisões de assunto, que em um esquema de metadados podem ser entendidas tanto como novos metadados de assunto, como um metadado dentro de outro metadado de assunto (ou elemento dentro de outro elemento). A última situação constitui uma forma de pré-coordenação.

Na sequência estão exemplos hipotéticos de parágrafos ou campos MARC, no caso o 650 e 651, que correspondem a assunto tópico e geográfico. São mostradas também subdivisões de campo (assunto pré-coordenado ou cabeçalho de assunto) em contraponto à adoção de tesouro, sem subdivisões (formação de assunto pós-coordenada).

*Exemplo de linguagem de indexação pré-coordenada em Campo MARC 21
extraído da Rede Virtual de Bibliotecas (RVBI)*

650 14 \a Direitos do menor \x Legislação \z Brasil.

Indicadores e marcadores de subcampo podem variar em significado conforme o campo. Os indicadores 1 e 4, referentes ao campo 650 acima, apontam respectivamente que o cabeçalho é do nível primário ou cobre com maior foco o assunto da obra, e o segundo “fonte do vocabulário controlado não especificada”. O indicador 7 no lugar do 4, indica fonte do vocabulário especificada no subcampo \2, conforme dados do *site* da OCLC abaixo (2012).

1st Indicator Level of subject - The level of the subject entry. Use the 1st indicator to distinguish primary and secondary descriptors.

- 1 Primary. A subject covers the main focus or subject content of the material.
- 2 Secondary. A subject covers a less important aspect of the content of the material.

2nd Indicator Thesaurus

- 0 Library of Congress subject heading
- 1 LC subject heading for children's literature
- 2 Medical subject heading
- 3 National Agricultural Library subject authority file
- 4 Source not specified
- 5 Canadian subject heading
- 6 Répertoire de vedettes-matière
- 7 Source is specified in subfield \2
- 8 Sears subject heading

Resumindo, quando o indicador 1 é igual a 1 (nível primário), significa que o assunto informado é principal no documento tratado. Já indicador 1 igual a 2 (nível secundário) mostra que o assunto é periférico ou cobre um aspecto menos importante. No indicador 2, estão os tesouros ou vocabulários controlados explicitamente nomeados, e a opção 4, que significa fonte do vocabulário não especificada. Já indicador 2 igual a 7 significa que a fonte do vocabulário será especificada no subcampo \2 do campo 650.

As subdivisões indicadas pelos marcadores de subcampo \a, \x e \z representam respectivamente, assunto tópico, divisão genérica ou geral e geográfica. Indicadores e marcadores de subcampo além de indicarem como ocorreu a formação do assunto, também facilitam a incorporação de regras e controles específicos para partes do assunto.

*Exemplo hipotético de linguagem de indexação pós-coordenada
em Campo MARC 21*

```
650 17 \a Direitos do menor. \2 Tesouro jurídico.
650 27 \a Legislação. \2 Tesouro jurídico.
651 #7 \a Brasil. \2 Tesouro jurídico.
```

O símbolo '#', no campo 651, significa "indicador indefinido ou não especificado". Esse é um exemplo hipotético, baseado na ideia de que uma linguagem pós-coordenada não precisa seguir nenhuma forma de ordenação ou encadeamento pré-concebido de assuntos, o que não significa que não possa, já que a norma ISO 2788 não estabelece que tesouros, considerados geralmente linguagens fundamentalmente pós-coordenadas, precisem usar somente pós-coordenação.

Outro exemplo da polêmica pré *versus* pós trata da "não naturalidade" da associação entre termos (caso da pré-coordenação), o que remete a um maior distanciamento da linguagem natural. Isso para a legislação federal brasileira, que é um instrumento de interesse e consulta pública, pode ser visto por muitos como mais uma barreira para o tratamento e a recuperação da legislação federal brasileira.

Por outro lado, em relatório que discute as vantagens e desvantagens de linguagens pré-coordenadas e pós-coordenadas, a *Library of Congress* (2007, p. 4, tradução nossa) afirma que Elaine Svenonius, especialista da área, no final do seu livro, “Fundamento intelectual da organização da informação”, “salienta que a pré-coordenação é necessária para a desambiguação, precisão e sugestibilidade”. No mesmo relatório, são elencadas oito vantagens da pré-coordenação:

- 1) Flexibilidade em termos de concepção do sistema, já que a indexação pré-coordenada pode ser desmontada para uso em sistemas pós-coordenados;
- 2) Recomendação do *National Coordinated Cataloging Program* (NCCP) para a criação de nova subdivisão de assunto;
- 3) A pré-coordenação oferece uma característica de proximidade entre as palavras-chave que pode ser útil para busca por operadores de proximidade;
- 4) Obras genéricas ou introdutórias são indicadas por meio da navegação em cabeçalhos não subdivididos.
- 5) Expressa mais claramente a relação entre assuntos do ponto de vista do documento;
- 6) Exibição hierárquica que propicia a navegabilidade;
- 7) A ordem padrão das subdivisões de assunto dá sentido às palavras usadas.
- 8) Fornece critérios para melhorar a classificação de relevância de resultados de pesquisa de sistemas como o Google.

Entre as desvantagens da pré-coordenação, o relatório aponta: é muito complicada e ainda requer indexação humana, não é facilmente compreendida pelos usuários finais, é cara, apresenta hierarquia rígida e pouco valor quando a navegação não é uma opção, pois os usuários normalmente não pesquisam por assuntos ordenados ou conceitos pré-estabelecidos, mas por palavras-chave.

E na atualidade, com o avanço exacerbado da Web e dos motores de busca, é impossível não pensar em metadados de assunto densos em termos de conteúdo, mais simples em termos de estrutura, próximos da realidade de quem trata e pesquisa informações, e tendo em mente que esses não são muitas vezes especialistas das áreas de informação, como acontece na própria Câmara dos Deputados.

Com o intuito de verificar metadados de assuntos mais simples para a legislação federal brasileira, e mesmo pelo fato de que a indexação da legislação na Câmara usa tesouro, uma linguagem geralmente de natureza pós-coordenada, optou-se pela discussão de metadados de assunto genéricos, sem adentrar em possíveis sub-metadados ou sub-elementos de metadados de assunto. Apesar da existência de categorias de assuntos na legislação que apresentam características de subdivisões ou de subcampos ou sub-metadados, como é o caso de assuntos que representam ações ou verbos, geralmente conhecidos como “modificadores” ou “diferença” de assuntos tópicos na indexação (ISO 2788, 1986, p. 3, tradução de Bianca Amaro de Melo).

Com o propósito de exemplificar o uso de modificadores na indexação de legislação no Legin, são apresentados os dados de indexação de lei ordinária com os modificadores “composição” e “criação”. A sigla TRT pode ser entendida como uma espécie de qualificador do nome da instituição.

LEI Nº 12.721, DE 2 DE OUTUBRO DE 2012

Ementa: Dispõe sobre a criação de cargos de provimento efetivo no Quadro de Pessoal da Secretaria do Tribunal Regional do Trabalho da 6ª Região.

Indexação:

TRIBUNAL REGIONAL DO TRABALHO (TRT) - Sexta Região - Recife, PE -
Composição - Secretaria - Quadro de Pessoal - Analista Judiciário - Área -
Tecnologia da Informação - Cargo efetivo - Cargo público - Criação

Fonte: Legin Web, 2012.

Então, devido à abordagem de metadados de assunto genéricos preferida neste estudo, no caso do *MARC 21 bibliográfico* não houve uma preocupação excessiva com as subdivisões de assunto assumidas no formato. Mesmo porque cada tipo de metadado de assunto apresenta subcampos ou “sub-metadados” peculiares. O formato *MARC 21 de autoridade* não foi utilizado por tratar mais especificamente do desenvolvimento de vocabulários controlados. Os campos ou metadados de assunto indicados para uso local (correspondentes aos 69X) também não foram usados por causa de seu escopo específico.

Já as regras/controles atribuídos aos metadados no MARC 21 precisam ser considerados tanto quanto a delimitação dos próprios metadados de assunto. Essas regras e controles definem as características daquele metadado (no formato MARC 21 também entendido como campo ou subcampo) e dizem, geralmente, se o campo ou subcampo é obrigatório ou opcional, se aceita determinado tipo de informação, por exemplo, uma URL, ou qual o formato que deve ser assumido pelo conteúdo, como no caso do dado “data”, ou até mesmo limites absolutos em número de caracteres que o campo aceita.

O campo 653, por exemplo, mesmo referindo-se a assunto não extraído de vocabulário controlado, possui opções para os dois indicadores, sendo que o primeiro indica o nível do termo, ou se o *assunto tem caráter principal* no documento, e o segundo, o *tipo de termo ou nome* (não informado, tópico, pessoal, entidade coletiva, evento, cronológico, geográfico, de forma/gênero). É um campo repetitivo (R), com alguns subcampos.

Seguem os metadados de assunto do *MARC 21 bibliográfico* selecionados, bem como suas respectivas definições, conforme dados extraídos da Online Computer Library Center (2012) e da Library of Congress (2012).

QUADRO 1		
Campos MARC que configuram metadados de assunto (6XX)		
Campos MARC de assunto (6XX)	Definição e características	
	Definição	Exemplos/Regras/Controles
6XX	Contêm entradas e termos para acesso ao assunto.	A maioria destes campos contem assuntos baseados em listas de autoridade ou vocabulários controlados definidos no segundo indicador e no subcampo \2 (Fonte do vocabulário controlado). Usam para controle, indicadores e subcampo, e são repetitivos (R).
600 – Nome pessoal (R)	Entrada de assunto em que o elemento é um nome pessoal.	600 10 \a Shakespeare, William, \d 1564-1616.
610 – Entidade coletiva (R)	Entrada de assunto em que o elemento é o nome de uma entidade coletiva.	610 24 \a Sociedade Brasileira de Agricultura. 610 14 \a Brasil. \t [Constituição (1946)]
611 – Evento (conferências, seminários, etc.) (R)	Entrada de assunto em que o elemento é um evento.	611 24 \a Conferência das Nações Unidas sobre o Direito dos Tratados \n (6. : \d 1968-1969 : \c Viena, Áustria)

Continuação		
630 – Título uniforme (R)	Entrada de assunto em que o elemento é o nome de um título uniforme (títulos padronizados, que visam a reunião de todas as expressões (edições, traduções, etc.) e manifestações de uma obra (diferentes formatos/suportes).	630 04 \a Bíblia. \l Português \x Comentários. 630 00 \a New York Times \v Índices. 630 04 \a Estatuto da Criança e do Adolescente (ECA).
648 – Assunto cronológico (R)	Entrada em que o elemento é um termo cronológico.	648 #7 \a 1900-1999. \2 fast 648 #4 \a Desde 1990. \2 fast
650 – Assunto tópico (R)	Entrada em que o elemento é um termo tópico, que consiste em um assunto genérico ou geral.	650 #4 \a Indexação (Economia) 650 #4 \a Arquitetura moderna \z França \y Século XIX.
651 – Assunto geográfico (R)	Entrada em que o elemento é um nome geográfico.	651 #4 \a Rio Amazonas. 651 #4 \a Rússia \x História \v Mapas.
653 – Assunto não-controlado ou Termo livre (R)	Índice de termos não controlados, que não são construídos segundo convenções das linguagens documentárias.	653 14 \a Joyce. 653 14 \a Dublin.
654 – Assunto tópico facetado (R)	Assunto tópico construído a partir de um vocabulário facetado, por exemplo, o Art and Architecture Thesaurus (AAT). O termo é associado à codificação da faceta no vocabulário.	654 ## \c r \a Paisagem \y Século XVIII \z Inglaterra. \2 aat
655 – Assunto gênero/forma (R)	Termos que indicam o gênero, a forma e/ou características dos materiais descritos.	655 #4 \a Diário. 655 #4 \a Dicionário.
656 – Assunto ocupação (R)	Índice de termos que designam ocupações (inclusive passatempos) presentes nos conteúdos dos materiais.	656 #7 \a Antropólogos. \2 [Código Tesouro] 656 #7 \a Educadores. \2 [Código Tesouro]
657 – Assunto função (R)	Índice de termos que descrevem a atividade ou função que gerou os materiais descritos.	657 #7 \a Gestão de benefícios pessoais \x Acidentes de trabalho \x Doenças \x Bioestatística. \2 New York State Management Functions Index.
658 – Assunto objetivos de cursos ou currículos (R)	Índice de termos que denotam os objetivos	658 ## \a Conscientização sobre o abuso de drogas \b pressão dos pares \b

Continuação	de currículo ou de cursos aplicáveis ao conteúdo dos materiais descritos.	compreender a lei. \2 local
662 – Assunto geográfico hierárquico (R)	Forma hierárquica de um assunto nome geográfico.	662 ## \a Canadá \d Toronto. 662 ## \a África \g Rio Nilo.

Fonte: Autor.

Notas: (R) = campo repetitivo, que pode possuir mais de uma ocorrência no conjunto de metadados.

A seguir serão discutidos, resumidamente, alguns outros padrões de metadados por serem muito conhecidos ou por fazerem parte de estudos voltados para o setor público, responsável pela criação e publicidade da legislação.

O padrão de metadados *Dublin Core Metadata Element Set* (DCMES) surgiu de discussões em um seminário realizado em 1995 em Dublin, Ohio, patrocinado pela OCLC e pelo *National Center for Supercomputing Applications* (NCSA). O conjunto de elementos foi nomeado de Dublin Core devido ao local de realização do evento. O desenvolvimento contínuo do Dublin Core e as especificações relacionadas são geridos pela *Dublin Core Metadata Initiative* (DCMI).

Segundo Marcondes (2006, p. 100), o Dublin Core é composto por “15 elementos descritivos, simples e auto-explicativos”, que são:

- Subject (assunto);
- Title (título);
- Creator (autor, responsável pelo documento);
- Publisher (publicador, quem torna o documento disponível na Internet);
- Contributor (outros colaboradores – editores, tradutores e outros);
- Description (descrição, resumo, sumário);
- Date (data de publicação);
- Type (tipo de recurso – *homepage*, romance, poesia, *software*, dicionário);
- Format (formato do arquivo que contém o documento eletrônico [...]);
- Identifier (geralmente o URL de um documento eletrônico);
- Relation (relacionamentos com outros documentos. Ex.: versões);
- Source (fonte ou origem);
- Language (idioma do documento);
- Coverage (cobertura – espacial ou temporal, sobre que lugar, ou sobre a que época o texto se refere);
- Rights (texto livre especificando qualquer restrição referente a direitos autorais).

O Dublin Core foi desenvolvido para ser utilizado pelos próprios produtores da informação. Todos os elementos Dublin Core são opcionais e repetitivos. De acordo com a NISO (2004, p. 3), o objetivo original do Dublin Core foi definir poucos elementos e algumas regras simples para descrever documentos Web. Existem partidários de uma visão minimalista do padrão, mantendo menos elementos, com semântica e sintática simples, e de uma visão estruturalista, que defendem uma semântica mais refinada, com mais distinções e extensibilidade para comunidades específicas.

Essas distinções criaram o Dublin Core qualificado e não qualificado. Segundo a NISO “os qualificadores podem ser utilizados para refinar (estreitar o escopo de) um elemento ou para identificar o esquema de codificação utilizado na representação do valor ou do conteúdo do elemento”. Marcondes (op. cit., p. 101) exemplifica essa situação com os metadados de assunto abaixo, sendo o qualificador, a explicitação do vocabulário controlado utilizado, no caso, Descritores em Ciências da Saúde (DECS) ou a tabela de áreas do conhecimento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

```
<meta name=>"dc.Subject scheme="DECS" content="SIDA"> ou  
<meta name="dc.Subject scheme="CNPq" content="Ciências Sociais  
Aplicadas">.
```

A NISO (op. cit., p. 3) afirma que enquanto o Dublin Core deixa regras de conteúdo para a implementação local ou particular, a DCMI encoraja a adoção de perfis de aplicação (regras de domínios específicos) para domínios ou áreas como educação e governo.

No Brasil, existem iniciativas recentes na área pública para a geração de metadados como o Padrão de metadados do Governo Eletrônico (e-PMG) (2010, p. 11). Sua última versão aponta os seguintes metadados de assunto:

a) *Assunto.categoria* que deve representar o assunto principal do recurso (o termo deve ser extraído do Vocabulário Controlado do Governo Eletrônico - VCGE). “Permite aos usuários navegar nos diretórios dos portais de categorias mais amplas (ex.: Saúde e Bem-Estar) para mais específicas (ex.: Vacinação,

calendário do idoso)”. Exemplo: assunto.categoria=”Agricultura, extrativismo e pesca. Agricultura orgânica”;

b) Assunto.palavra-chave usado para descrever o assunto principal do recurso. Recomenda-se o uso de um vocabulário controlado. Exemplo: assunto.palavra-chave=”Servidor público” (VCB – Senado Federal);

c) Assunto.pessoa utilizado quando um recurso é sobre uma pessoa (física ou jurídica). Exemplo: assunto.pessoa=”Fundação Getúlio Vargas”.

No e-ARQ Brasil, modelo de requisitos para sistemas informatizados de gestão arquivística de documentos, há a indicação de metadado de assunto genérico, mas nenhuma orientação quanto a possíveis tipos de metadados de assunto. Existe apenas uma nota de aplicação a qual informa “as instituições devem definir sua política de indexação”, o que, de forma mais ampla, também pode envolver o desenvolvimento granular de metadados de assunto. (CONARQ, 2011, p. 102)

2.3 CRIAÇÃO E QUALIDADE DE METADADOS

Desenvolver uma metodologia para a criação de metadados parece ter se tornado fator determinante do sucesso dos próprios metadados no alcance das finalidades a que se propõem. Assim, na literatura técnica e científica é possível encontrar orientações e até metodologias para a criação de metadados, como no caso do estudo de Almeida (1999, p. 29), sobre a disseminação da informação geográfica e os seus metadados.

Nesse estudo, o autor apresenta uma metodologia para a criação de metadados por, em tese, qualquer instituição ou organização, que consiste de *duas premissas e cinco passos ou etapas* ordenadamente dispostas. Essa metodologia pode ser generalizada para outros tipos de informação, que não geográficas. As *duas premissas* para a geração de metadados dizem:

- os dados devem ser priorizados – os dados que devem ser documentados primeiro são aqueles conjuntos de dados que tenham uso corrente ou futuro, que sejam básicos para a geração de outros conjuntos de dados ou aqueles que representam um alto investimento, em termos de esforço e custo;
- deve-se documentar a um nível que preserve o valor dos dados dentro da instituição ou organização.

Resumidamente, seguem as *cinco etapas* da metodologia proposta pelo autor, as quais são apresentadas de forma generalizada, com o objetivo de abranger outros tipos de informação, que não sejam geográficas:

1) Deve-se começar com um padrão de metadados já existente. Pode-se procurar na Internet *sites* onde são listados padrões de metadados de organizações reconhecidas, privadas ou públicas, e identificar conjuntos de dados semelhantes aos da instituição que pretende adotar os metadados.

2) A decisão de se estender ou não o padrão de metadados escolhido deve ser feita por um grupo composto por pessoas que representem o conhecimento da organização.

3) Deve-se ter em mente que os elementos estendidos terão caráter específico e não se beneficiaram de um padrão internacional de pesquisa a bases de dados distribuídas, contudo, se o objetivo é a utilização por um grupo de instituições ou organizações, os novos elementos poderiam ser padronizados para possibilitar uma pesquisa uniforme;

4) Bases de dados e não somente de metadados (as chamadas metabases) também podem ser usadas como insumo para a criação de metadados e de interfaces de busca comuns;

5) A cultura de preenchimento de dados deve ser analisada. Nesse caso existem duas possibilidades:

a) Alocar um profissional responsável pelo preenchimento do conteúdo dos metadados. Nesse caso será necessária “uma boa comunicação entre o responsável pelos metadados e o produtor dos dados, o primeiro terá, obrigatoriamente, que fazer perguntas ao último”. E vice-versa, porque quem cria os metadados tem outro olhar, que nem sempre é compartilhado por quem alimenta os dados;

b) Deixar que os próprios produtores dos dados preencham os metadados. Nesse caso, serão necessários *softwares* que auxiliem tanto na compreensão como no preenchimento dos metadados. O autor finaliza dizendo “como é necessário conhecer detalhadamente o documento para o qual estão sendo criados os metadados, esta visão é a que mais se adéqua a um metadado consistente”. Isso é questionável, se considerarmos que o produtor dos dados nem sempre compreende os metadados.

A *segunda premissa* para a criação de metadados de Almeida (1999, p. 29), “deve se documentar a um nível que preserve o valor dos dados dentro da instituição”, parece tratar, em princípio, da granularidade da informação. Isso porque, em geral, informações prioritárias ou de grande valor devem ser mais detalhadas, a fim de facilitar sua recuperação futura, o que implica em um maior refinamento de metadados. A granularidade está sendo muito discutida. A respeito, Zeng (2010, p. 179 et seq.) cita:

Se olharmos para os recursos de informação propriamente ditos, o que há para ser descoberto está em diferentes níveis atômicos. [...] Com o intuito de revelar a riqueza de conteúdos de um recurso, geralmente, a primeira tentativa é descrever a estrutura dos componentes em nível granular maior. [...] Outra tentativa é expor o(s) tópico(s) que um recurso de informação discute ou coisas que ele representa por meio da decomposição do conteúdo de um recurso. [...] Os esquemas de metadados ou perfis de aplicação apontados para áreas científicas estenderam seus elementos além da descrição de um *container* e enfatizam a descrição específica orientada por domínio de assunto.

As questões da granularidade, decomposição de estrutura e conteúdo são centrais neste estudo, porque a ideia é decompor em nível granular maior o único metadado de assunto existente na base Legin, no que refere à legislação federal brasileira, para criar novas possibilidades tanto de recuperação, como de estruturação dos dados de assunto, baseadas em novos metadados de assunto.

E é preciso dizer que na legislação são inúmeras as possibilidades de decomposição de estrutura e conteúdo para a criação de novos metadados de assunto, pois os atos normativos além da própria estrutura interna têm uma estrutura de relações e temporal paralela, no caso, por exemplo, dos textos atualizados dos atos normativos.

A NISO (2004, p. 10, tradução nossa) relata que a criação de metadados quase sempre é resultado de esforços cooperativos, apesar de isso ser variável conforme a área ou disciplina, o recurso a ser descrito, as ferramentas disponíveis e os resultados esperados. Ela ainda acrescenta:

Muitos metadados estruturais e administrativos básicos são fornecidos pela equipe técnica, que inicialmente digitaliza ou cria de outra forma o objeto digital, ou são gerados através de um processo automatizado.

Entretanto, não basta criar os metadados, é preciso assegurar a qualidade dos mesmos em dois momentos: no *processo de desenvolvimento* e na *alimentação dos conjuntos de metadados*. E garantir essa qualidade inclui a adoção de metodologias e processos constantes de validação. Assim como a natureza das informações pode mudar, os metadados que as descrevem também. É necessário checar se conseguimos ou não responder às questões que se colocam à organização com os metadados que dispomos constantemente.

Mas é claro que essas mudanças devem ser avaliadas conforme o custo-benefício da reutilização, implantação e disponibilização de novos metadados. Por exemplo, o acréscimo de um novo conjunto de metadados em um conjunto pré-existente pode implicar em novas alimentações de dados, antes não tratados. Porém, se esses metadados são prioritários, devem ser desenvolvidas ações que assegurem a inserção e disponibilização à organização dos mesmos.

Essa ideia da evolução dos metadados, assim como da incorporação ou combinação de novos conjuntos de metadados, está presente no estudo de Zeng (2010, p. 188) sobre a incorporação, por exemplo, de metadados de assunto originários de linguagens de marcação específicas, por domínio ou área, em conjuntos de metadados descritivos de documentos. O autor faz a seguinte colocação sobre metadados recombinantes e qualidade de metadados:

O emprego de um [...] esquema como fundação para agregar características de descrição de metadados e de registros (ou documentos) de marcação aumenta consideravelmente a consistência, e, portanto, garante também a interoperabilidade. [...] Isso pode ser usado para criar documentos que indicam todas as relações de composição, decomposição, combinação e recombinação para recursos originais ou novos. Somando-se aos resultados da combinação, essa abordagem permite, também, integrar descrições geradas por máquinas e por pessoas. As descrições existentes podem ser reusadas para outros projetos apropriados. Simultaneamente, a qualidade dos metadados pode ser melhorada por meio de metadados recombinantes.

Essa proposta de metadados recombinantes faz lembrar as relações entre documentos ou informação ou conjuntos de metadados descritivos tão presentes na legislação, só para citar a relação necessária entre a proposição originária e o ato normativo resultante dela.

Já a qualidade dos metadados no processo de criação, propriamente dita, de metadados para a legislação, está relacionada, principalmente, com a adoção de metodologia, alinhada com o propósito da reutilização de dados e orientada tanto para a teoria de metadados, como para o texto legal e seu respectivo ordenamento jurídico.

A qualidade dos metadados também está ligada com a própria alimentação destes metadados e com a validação dessa alimentação. A respeito disso, a NISO (op. cit., p. 10) coloca:

Para metadados descritivos, é melhor em algumas situações, que o autor do recurso forneça as informações. Isto é particularmente verdadeiro nos conjuntos de dados da documentação científica, onde o autor tem compreensão significativa da justificativa para o conjunto de dados e dos usos que poderiam ser feitos dos dados, e para os quais existe pouca ou nenhuma informação textual a partir da qual um indexador possa trabalhar. No entanto, muitos projetos têm verificado que é mais eficiente dispor de indexadores ou de outros profissionais de informação para criar metadados descritivos, porque os autores ou criadores dos dados não têm tempo ou habilidades. Em outros casos, uma combinação de profissional de informação e pesquisador é usada. O pesquisador pode criar um esqueleto, completando os elementos que podem ser fornecidos mais facilmente. Em seguida, os resultados podem ser complementados ou revistos por um especialista na informação, para verificar consistência e conformidade com a sintaxe do esquema e diretrizes locais.

Ainda sobre a qualidade, a NISO diz que muitas iniciativas de projetos de metadados têm desenvolvido ferramentas e as disponibilizado, às vezes, gratuitamente. E que tais ferramentas podem ser divididas em várias categorias:

- a) *Templates* (Modelos) permitem que um usuário insira os valores de metadados para campos pré-definidos ou elementos utilizados;
- b) *Mark-up tools* (Ferramentas de marcação) estruturam os valores e atributos de metadados na língua do esquema especificado;
- c) *Extraction tools* (Ferramentas de extração) criam metadados automaticamente a partir da análise de recursos digitais. Estas ferramentas são geralmente limitadas a recursos textuais. Os metadados resultantes devem ser sempre manualmente revisados e editados;
- d) *Conversion tools* (Ferramentas de conversão) traduzem um formato de metadados para outro.

No mesmo estudo, há uma indicação para o *Framework of Guidance for Building Good Digital Collections*, disponibilizado no site da NISO, o qual articula seis princípios aplicados a bons metadados:

- 1) Devem ser adequados aos materiais e usuários da coleção e preverem o uso atual e potencial do objeto digital;
- 2) Suportar interoperabilidade;
- 3) Usar vocabulários controlados para refletir o “que”, “onde”, “quando” e o “quem” presentes no conteúdo;
- 4) Incluir definição clara sobre as condições e termos de uso do objeto digital;
- 5) Bons registros de metadados são eles próprios objetos e, portanto, devem ter as qualidades de arquivamento persistente, identificação única, etc. Devem ser confiáveis e verificáveis;
- 6) Suportar a gestão, em longo prazo, de objetos em coleções.

Por fim, a NISO apresenta alguns esforços em andamento para lidar com o desafio da qualidade dos metadados: melhorias na validação e no processo de inserção de dados das ferramentas de criação, desenvolvimento e aperfeiçoamento de programas interoperáveis para a troca (*crosswalk*) entre diferentes esquemas, treinamento de criadores de conteúdo na compreensão de conceitos relacionados aos metadados, uso mais amplo e consciente dos vocabulários controlados existentes e, por fim, o desenvolvimento e aperfeiçoamento de esquemas de metadados específicos.

2.4 LEGISLAÇÃO FEDERAL: CONCEITOS E ESTRUTURA

De acordo com Mourão & Mastro¹⁷ (1988, p. 13-15 apud PEREIRA, 2004, p. 154) e em sentido mais amplo, legislação é “o conjunto de leis e atos normativos emanados de autoridades competentes (do Executivo ou do Legislativo).” No caso, a legislação federal brasileira constitui no conjunto de leis e atos normativos emanados dos Poderes Legislativo e Executivo da esfera federal de governo do Brasil. Sendo que a esfera ou validade espacial é um dos critérios de classificação das normas jurídicas segundo Poletti¹⁸ (1991

¹⁷ MOURÃO, L. de A.; MASTRO, B. T. **Manual de implantação**: serviço de documentação jurídica. São Paulo: Fund. Pref. Faria Lima, CEPAM, 1988.

¹⁸ POLETTI, R. **Introdução ao direito**. São Paulo: Saraiva, 1991. 334 p.

apud CARNEIRO, SANTOS e NÓBREGA NETTO, 2011, p. 38), que podem ser classificadas em:

- a) do *sistema* a que pertencem: nacionais (normas de um determinado país), estrangeiras (pertencentes a outros países), de direito uniforme (normas de direito internacional);
- b) da *fonte*: legislativa (emanadas do processo legislativo), consuetudinária (oriundas dos costumes da sociedade), jurisprudencial (provenientes das sentenças judiciais e dos repositórios jurisprudenciais) e contratual (fruto do ajuste de vontades decorrentes de contrato);
- c) da *validade espacial*: federais, estaduais, municipais;
- d) da *temporalidade*: vigência determinada e vigência indeterminada (ou seja, tempo de vida das normas);
- e) da *materialidade*: de direito público (direito constitucional, administrativo, penal, processuais, internacional público, tributário), de direito privado (direito civil, comercial, internacional privado);
- f) de *hierarquia*: constitucionais, complementares, ordinárias, regulamentares, individualizadas privadas e públicas;
- g) de *sua qualidade*: positivas ou permissivas, proibitivas ou negativas.

Os autores ainda afirmam que “a lei é um ato normativo de caráter geral, abstrato e obrigatório, que tem por objetivo ordenar a vida coletiva, trazendo certeza, precisão e garantia às relações jurídicas” e que são leis em sentido amplo (*lato sensu*):

As emendas à Constituição, as leis complementares, as leis ordinárias, as leis delegadas, as medidas provisórias, os decretos legislativos e as resoluções (CF [Constituição Federal de 1988], art. 59). Em sentido estrito (*stricto sensu*), são consideradas leis apenas as leis ordinárias, leis complementares e leis delegadas.

Outra espécie normativa muito representativa, em termos quantitativos da legislação federal brasileira é o Decreto, que segundo Andreta (2004, p. 94) é:

Ato administrativo emanado do Poder Executivo, com o fim de regulamentar a lei propriamente dita, ou de ensejar a tal Poder, a realização de atos inerentes a sua competência. [...] é, hierarquicamente, inferior à lei, [...] pode ser autônomo ou regulamentar.

Com a denominação de “Decreto”, existem também atos emanados do Poder Legislativo, encontrados no início da República. Porém, esses decretos, emitidos pelo Legislativo, não foram utilizados neste estudo por serem esporádicos, e, numericamente, menos representativos.

A legislação federal brasileira apresenta geralmente alterações, revogações, inclusões e retificações, que dão origem ao texto modificado ou atualizado do ato normativo em questão. Segundo Scheweighofer e Lachmayer¹⁹ (1997 apud LIMA, 2006, p. 178) “as idéias e as realizações são altamente interligadas (...) o pensamento jurídico é altamente dominado por estruturas de *links*”. O autor exemplifica:

[...] ao se promulgar uma nova norma jurídica, ela se insere em um ordenamento jurídico já existente, alterando-o [segundo o autor a única exceção é a promulgação de uma Constituição, que institui um novo ordenamento jurídico] [...] ao se considerar o ciclo documentário da norma jurídica, é importante considerar também a proposição legislativa que originou a norma jurídica.

O ordenamento jurídico brasileiro é bastante complexo. Isso é agravado por problemas relacionados à falta ou insuficiência da técnica legislativa, a qual segundo Carvalho (2010, p. 4 e p. 105) é “instrumento indispensável à redução e simplificação do excesso de normas jurídicas” e consiste “no modo correto de elaborar as leis, de forma a torná-las exequíveis e eficazes. Envolve um conjunto de regras e de normas técnicas que vão desde a necessidade de legislar até a publicação da lei”.

De acordo com Soares (2007, p. 125) a técnica legislativa evoluiu, basicamente, do seguinte: “a necessidade de uma legislação mais eficaz [...], o questionamento da lei como o instrumento exclusivo para a consecução de mudanças sociais, a necessidade de democratizar o acesso aos textos legais em todos os níveis”.

Os estudos de técnica legislativa discutem, entre outras coisas, critérios que garantam clareza aos atos normativos. Um deles trata da economia temática do texto legal, o qual deve conter só o necessário, sem repetições. A respeito disso, Carvalho (op.cit., p. 132) afirma que o texto legal “não deve enunciar a mesma regra em mais de um lugar”.

A Lei complementar nº 95, de 26 de fevereiro de 1998, alterada pela Lei complementar nº 107, de 26 de abril de 2001 veio regulamentar o parágrafo único do art. 59 da Constituição Federal de 1988, o qual diz “Lei complementar disporá sobre a elaboração, redação, alteração e consolidação das leis”.

¹⁹ SCHEWEIGHOFER, E.; LACHMAYER, F. **Ideas, visualisations and ontologies. International Workshop on Legal Ontologies.** Melbourne, Austrália: [s.n.], 1997, p. 7-13.

Carvalho (ibid., p. 4) trata dos benefícios propiciados por essa lei ao ordenamento jurídico brasileiro que:

[...] concorre para o melhor conhecimento do ordenamento jurídico brasileiro, que se tem apresentado de forma não harmônica, repleto de incongruências, comandos repetitivos, a necessitar, por isso, de consolidação capaz de torná-lo coerente e claro.

Entre outras coisas, a Lei complementar nº 95, de 1998, trata da estrutura e do teor dos atos normativos. O parágrafo único do art. 12 da referida Lei, acrescido pela Lei complementar nº 107, de 2001, diz:

Parágrafo único. O termo 'dispositivo' mencionado nesta Lei refere-se a artigos, parágrafos, incisos, alíneas ou itens.

A Lei complementar nº 95, de 1998 ainda diz que a unidade básica de articulação dos textos legais será o artigo, os quais se desdobram em parágrafos ou incisos, os parágrafos em incisos, os incisos em alíneas e as alíneas em itens. O agrupamento de artigos poderá constituir em Subseções, o de Subseções, em Seção, o de Seções, em Capítulo, o de Capítulos, em Título, o de Títulos, em Livro e o de Livros, em Parte, a qual poderá desdobrar-se em Geral e Especial ou subdivididas em numeração. Os agrupamentos previstos também poderão compreender Disposições preliminares, Gerais, Finais e Transitórias, conforme necessário. A referida Lei complementar também diz:

Art. 3º A lei será estruturada em três partes básicas:

I - parte preliminar, compreendendo a epígrafe, a ementa, o preâmbulo, o enunciado do objeto e a indicação do âmbito de aplicação das disposições normativas;

II - parte normativa, compreendendo o texto das normas de conteúdo substantivo relacionadas com a matéria regulada;

III - parte final, compreendendo as disposições pertinentes às medidas necessárias à implementação das normas de conteúdo substantivo, às disposições transitórias, se for o caso, a cláusula de vigência e a cláusula de revogação, quando couber.

Mas antes da Lei complementar nº 95, de 1998, os textos legais não seguiam essa forma clara e mais rígida de organização, o que deve ser considerado porque a legislação federal apresenta outras formas de estruturação diferentes desta, que podem ser importantes, no caso da decomposição de estrutura para a criação de metadados de assunto. A título

de exemplo, a própria grafia “Parágrafo único”, encontrava o semelhante “Artigo único”.

Em função disso, no portal LexML Brasil (2008), um metabuscador de legislação, jurisprudência, proposições legislativas e doutrina, foi encontrado, a respeito da criação de um esquema de metadados para a marcação de textos legais o seguinte:

Devido à falta de padronização da técnica legislativa em relação às três esferas e à existência de normas federais com articulação incompatível com a Lei Complementar nº 95/1998 (principalmente normas anteriores a 1998), foram criados dois esquemas, a exemplo da estratégia do Projeto NIR²⁰, conforme abaixo:

- Esquema Rígido, de acordo com a Lei Complementar nº 95/1998 – Normas e projetos de normas que seguem as regras de articulação definidas pela Lei Complementar nº 95/1998.
- Esquema Flexível – Permite mais combinações na articulação dos dispositivos. Aplica-se às normas e proposições legislativas que não seguem a Lei Complementar nº 95/1998 além de servir como um denominador comum entre os esquemas rígidos definidos para cada técnica legislativa.

Também no Portal LexML Brasil (2008) existe a seguinte colocação sobre a vinculação de normas por assunto:

Normas jurídicas tais como leis, decretos ou instruções normativas não são documentos isolados uns dos outros. Nosso ordenamento jurídico é composto pelo conjunto de normas emanadas por vários órgãos e publicadas oficialmente. Alguns elementos permitem caracterizar o conjunto de normas como um sistema coeso: [...] Vinculação por assunto: apesar de ser um dos aspectos mais difíceis de se estabelecer pela via da automação, o agrupamento de normas em função do assunto que tratam é de grande importância. Metodologias da Biblioteconomia e da Jurisprudência garantem a confiabilidade de procedimentos computacionais assistidos por pessoas especializadas.

Essa dificuldade de ligar normas por assunto é acentuada devido aos textos legais nem sempre também estarem de acordo com as seguintes disposições da Lei complementar nº 95, de 1998:

Art. 7º O primeiro artigo do texto indicará o objeto da lei e o respectivo âmbito de aplicação, observados os seguintes princípios:
I - excetuadas as codificações, cada lei tratará de um único objeto;

²⁰ “Um trabalho desenvolvido pela Itália, conhecido como Projeto NIR, visando à simplificação, agilização e unificação do sistema judiciário do país através da informática, será estendido a partir deste ano para todos os países da Comunidade Européia. Sua concepção e montagem, com detalhamento técnico, foi o tema da palestra proferida pelo professor Pierluigi Spinoza, um dos idealizadores e membro do Instituto de Teoria e Técnica de Informação Jurídica da Itália”. – PROJETO italiano simplifica a justiça (2007).

II - a lei não conterà matéria estranha a seu objeto ou a este não vinculada por afinidade, pertinência ou conexão;

III - o âmbito de aplicação da lei será estabelecido de forma tão específica quanto o possibilite o conhecimento técnico ou científico da área respectiva;

IV - o mesmo assunto não poderá ser disciplinado por mais de uma lei, exceto quando a subseqüente se destine a complementar lei considerada básica, vinculando-se a esta por remissão expressa.

Art. 11
 III -

b) restringir o conteúdo de cada artigo da lei a um único assunto ou princípio.

Em relação à linguagem e conteúdo do ato normativo, Carvalho (2007, p.8) afirma que para atender ao princípio democrático, as normas jurídicas deverão ser dotadas de determinados atributos:

[...] precisão ou determinabilidade, clareza e densidade, [...] A clareza das leis envolve a necessidade de se dosar a linguagem normativa de modo a propiciar o equilíbrio entre linguagem normativa e técnica, o que se mostra fundamental para que a lei seja conhecida do povo a fim de que possa ser cumprida. Deve-se, contudo, levar em consideração a vantagem da utilização, em certos casos, da linguagem técnica, para que se permita ao especialista do direito raciocinar com mais facilidade, em razão da área de conhecimento que reclama uma adequada terminologia.

A declaração acima traz à tona o caráter universal do conhecimento abordado pela legislação devido a mesma ser um instrumento para a ordem pública. A respeito do conteúdo da legislação, Guimarães (2004, p. 27) diz:

Considerando que o texto do ato normativo reflete, tanto jurídica quanto diplomaticamente, uma ação ou conjunto de ações (como a criação, a modificação ou a extinção de um direito), verifica-se a aplicabilidade da concepção de Kaiser na análise e representação temática [...] da legislação uma vez que os direitos em si, enquanto entidades abstratas (via de regra institutos jurídicos ou instituições, não necessariamente expressos, na legislação, por meio de termos técnicos jurídicos, dado o uso genérico da mesma) correspondem ao Concreto, núcleo temático do enunciado, ao passo que o Processo estaria exatamente nesse espectro de ações aplicáveis ao direito.

Portanto, o conteúdo ou assunto da legislação pode ser expresso por meio de Concretos ou núcleos conceituais dos enunciados e Processos ou ações.

2.5 BUSCAS DE INFORMAÇÃO NA WEB E BUSCAS DE LEGISLAÇÃO

Em pesquisa de 2003, citada anteriormente neste estudo e que será detalhada agora, a OCLC entrevistou bibliotecários, arquivistas e mesmo pessoas comuns, que buscavam informações no seu dia a dia, e organizou grupos de estudo com pessoas idosas, professores e alunos do ensino fundamental e médio. Coletou dados de gastos com educação, bibliotecas e tecnologia em 29 países, que representam 60% da população mundial e 85% do PIB. A pesquisa apontou como um dos padrões reconhecidos uma redução no acesso guiado ao conteúdo, ou, a maioria das pessoas que buscavam informações recorria diretamente à Internet e utilizavam, na maioria das vezes, motores de busca como o Google, não recorrendo a bibliotecas ou qualquer outro tipo de serviço de informação.

Assim, devido à importância da Web como fonte de pesquisas, existem estudos que tratam do comportamento de busca de informação na Web, os quais segundo Pinto (2010, p. 17) delimitam alguns tipos de busca possíveis:

O tipo de busca é definido pela natureza da necessidade de informação, que abrange uma grande variedade de motivações, que vão desde a necessidade de encontrar um fato até a procura exploratória (SUTCLIFFE; ENNIS, 1998 apud HEARST, 2009²¹). A natureza da necessidade, por sua vez, depende do momento e do contexto na qual ela surge. Diversos autores apresentam propostas de classificar os diferentes tipos de busca advindos das diferentes necessidades de informação. Após análise de *logs* de usuários do AltaVista, Broder²² (2002) propôs uma classificação com três tipos de busca, que veio influenciar autores posteriores: **Busca de navegação**: quando o objetivo imediato é alcançar um site particular; **Busca informacional**: quando o objetivo é adquirir informação que se assume presente em um ou mais *sites*; **Busca transacional**: quando o objetivo é realizar alguma atividade mediada pela Web, por exemplo, compras ou *download*.

Ao traçar um paralelo com os tipos de busca de informação possíveis da legislação na Web, pode-se dizer que tanto a busca de navegação, quanto a informacional devem acontecer frequentemente no caso da pesquisa na legislação federal brasileira, e a transacional pode ser mais rara, se pensarmos, por exemplo, na emissão de relatórios. No caso, a busca de navegação corresponderia à busca de um ato normativo específico, sobre o

²¹ HEARST, M. **Search user interfaces**. Berkeley: Cambridge University Press, 2009.

²² BRODER, A. A taxonomy of web search. **ACM Sigir Forum**, Nova York, v. 36, n. 2, 2002.

qual o pesquisador possua dados identificadores, como o número e o ano do ato normativo.

A busca informacional já envolveria pesquisas mais extensas e transversais sobre um determinado tema, a fim de determinar quais atos normativos tratam de um assunto específico. Costa²³ (1994 apud CARVALHO, 2010, p. 115, grifo nosso) ilustra bem esse tipo de pesquisa quando trata do processo legislativo, ou, mais especificamente das etapas da redação de uma proposição legislativa.

1. Verificação da relação social a ser normatizada.
2. Verificação da possibilidade jurídica da matéria [o autor cita constitucionalidade e legalidade]
3. *Pesquisa da legislação em vigor.*
4. Pesquisa da jurisprudência.
5. Estudo da matéria.
6. Elaboração do anteprojeto.
7. Revisão do anteprojeto.
8. Redação final da proposição.

No que refere à pesquisa na legislação o autor diz que ela contribui para a harmonização da futura lei com o conjunto de leis que tratam da mesma matéria e com o próprio ordenamento jurídico como um todo. Ainda segundo o autor, “nesta etapa o técnico legislativo irá verificar se a nova lei não viola o texto constitucional ou afronta leis hierarquicamente superiores a ela”. É recomendável inclusive o exame da legislação estrangeira, com as necessárias ressalvas referentes às peculiaridades nacionais.

Ainda sobre a busca informacional na legislação, Pereira (2004, p. 154) relata que a recuperação da informação jurídica é diferente da de outros tipos de documentação. Para ele, uma recuperação mais eficiente da legislação envolve:

Uma diversidade de índices. Quando solicitada uma pesquisa no sistema de informação jurídica todos os atos legais referentes ao assunto devem ser fornecidos (o pedido pode resultar em um conjunto de documentos), não tem validade para o usuário o fornecimento de parte da legislação requerida. Como a recuperação se faz diferente, a indexação também.

²³ COSTA, J. J. da. **Técnica legislativa**: procedimentos e normas. Rio de Janeiro: Destaque, 1994.

De modo geral, atualmente, muito se tem feito para melhorar as possibilidades de recuperação de informações em todas as áreas. Geralmente, as iniciativas estão ligadas ao desenvolvimento de sistemas mais direcionados ao apoio às tarefas e ao atendimento dos desejos de um público determinado.

Em obra sobre a norma *Resource Description and Access* (RDA) [Recursos: Descrição e Acesso] voltada para o tratamento e recuperação de informações bibliográficas e desenvolvida para ser utilizada com o modelo conceitual *Functional Requirements for Bibliographic Records* (FRBR) [Requisitos Funcionais para Registros Bibliográficos], Oliver (2011, p. 2) diz que “cada instrução da RDA reporta-se ao usuário e às tarefas que ele deseja executar. Essas tarefas do usuário têm sua origem nos modelos FRBR e FRAD [Requisitos Funcionais para Dados de Autoridade]”, sendo que dados de autoridade significam dados autorizados, que seguem modelos pré-estabelecidos, por exemplo, para assuntos. O autor complementa:

Tarefas que usam dados bibliográficos	Tarefas que usam dados de autoridade
encontrar	encontrar
identificar	identificar
selecionar	esclarecer
obter	compreender

O autor diz também que os dados RDA foram projetados para serem usados em ambiente de Rede, com novas estruturas de bases de dados, diferentes das usadas tradicionalmente por bibliotecas, e podem ser empregados como a base de um conjunto de elementos de metadados que permitam a visibilidade e utilização de dados, não necessariamente bibliográficos. Por fim, Oliver (ibid., p. 4) menciona:

As fronteiras entre comunidades de metadados não tem significado algum para o usuário que faz buscas num ambiente de rede em linha. [...] Quanto maior for a compatibilidade de dados entre comunidades de metadados, maiores serão os benefícios para os usuários.

O relatório do Grupo de trabalho da *International Federation of Library Associations and Institutions* (IFLA) (2010, p. 19, tradução nossa), órgão responsável pelo FRBR, sobre o modelo conceitual para *Functional Requirements for Subject Authority Data* (FRSAD) [Requisitos funcionais para dados de autoridade de assuntos] constata que não devem ser definidos os

atributos ou tipos de “tema”, que “constituem a categoria a qual o tema pertence no contexto de um sistema de organização do conhecimento particular”.

O Grupo afirma que, em uma implementação, temas podem ser organizados com base na categoria, espécie ou tipo. Assim, o relatório não sugere categorias\espécies\tipos específicos porque eles diferem dependendo da implementação. O grupo exemplifica a questão conforme segue.

Em algumas implementações, as entidades originais do FRBR [nesse caso, entidade no modelo entidade-relacionamento para estruturação de bases de dados] obra, manifestação, item, pessoa, família, entidade coletiva, conceito, objeto, evento e local podem ser usadas como tipos [de assunto], talvez até tempo. [...] Em outras implementações, um conjunto diferente de tipos pode ser definido. [...] [como no] exemplo de implementação Universal Medical Language System (UMLS).

Tipos semânticos UMLS

Entidades

- Objeto Físico
 - Organismo
 - Estrutura anatômica
 - Objeto manufaturado
 - Substância
- Entidade conceitual
 - Ideia ou conceito
 - Descoberta (Finding)
 - Atributo Organismo
 - Produto intelectual
 - Linguagem
 - Ocupação ou Disciplina
 - Organização
 - Atributo Grupo
 - Grupo

Evento

- Atividade
- Fenômeno ou processo

Assim, o relatório do Grupo FRASAD da IFLA conclui que não é possível determinar as categorias, espécies ou tipos de temas, os quais podem ser entendidos como os metadados de assunto de um recurso, em um modelo conceitual genérico como o FRASAD. Isso só torna maior a responsabilidade de quem vai desenvolver um sistema de busca por assunto em uma área específica, como no caso da legislação federal brasileira.

3 MÉTODO DE PESQUISA

O apontamento de aspectos a serem considerados na identificação de metadados de assunto é uma pesquisa que pode ser classificada como descritiva, porque tem como “objetivo primordial a descrição das características de uma determinada população” (GIL, 2007, p. 42), no caso, a legislação federal brasileira disponível no Legin e seus metadados de assunto. E, por focar na base Legin da Câmara, a pesquisa também pode ser classificada como estudo de caso.

A triangulação, segundo Vergara (2010, p. 243), “pode ser definida como uma estratégia de pesquisa baseada na utilização de diversos métodos para investigar um mesmo fenômeno”. No caso, a triangulação de dados, a qual também “refere-se ao uso de diferentes fontes de dados” foi o método de pesquisa adotado por contemplar tanto parâmetros quantitativos, como qualitativos, e, assim, possibilitar que os pressupostos do estudo pudessem ser verificados.

A ideia da necessidade também de dados qualitativos é reforçada por Almeida (1999, p. 29), em sua “metodologia para a criação de metadados”, a qual aponta como uma das premissas para a geração de metadados o seguinte: “deve-se documentar a um nível que preserve o valor dos dados dentro da instituição ou organização”. Isso deixa evidente a importância de se compreender melhor os dados da perspectiva de uso organizacional, ou, mais especificamente, das atividades fins desenvolvidas por meio da busca por assunto na legislação federal, disponível no Portal da Câmara, e do nível de granularidade da informação buscada.

Assim, esta pesquisa possui duas linhas de coleta de dados: uma que busca dados qualitativos do ponto de vista de quem realiza pesquisas por assunto na legislação federal brasileira via Legin Web, representada pelos questionários enviados para servidores em áreas específicas da Câmara; e uma que se apóia em dados quantitativos, a qual, por sua vez, utiliza duas fontes.

A primeira fonte constitui de levantamento de padrões de repetição de assuntos no campo de indexação da legislação federal brasileira disponível na Base Legin, portanto a perspectiva de quem indexa. A segunda é um

relatório emitido pelo Centro de Informática (Cenin) da Câmara dos Deputados, com os *logs* de acesso (URLs dinâmicas formadas por meio de consultas) do Legin Web de consultas realizadas, por usuários internos e externos da Câmara dos Deputados, a qual corrobora, assim como no caso dos dados qualitativos, o ponto de vista de quem faz pesquisa por assunto na legislação federal.

Devido ao número elevado de consultas, foi solicitado ao Cenin que o relatório se restringisse aos 30 dias anteriores à coleta, ou de 1 de setembro a 30 de setembro de 2012, o que correspondeu ao todo a 506 *logs* de acesso, que somaram 676 buscas, pois cada *log* podia implicar em mais de um tipo de busca.

A segunda fonte de dados foi composta de amostra dos atos normativos federais da Base Legin Web. A definição da amostra começou com levantamento dos atos normativos disponíveis no Legin, realizado em 5 de setembro de 2012, e mostrou que havia uma composição de normas federais no Legin com maior representação quantitativa de alguns tipos, que somavam 94,77% da base, conforme a disposição na TAB. 1 abaixo. Devido a isso, optou-se por trabalhar com amostra estratificada composta pelos tipos normativos mais representativos do Legin Web, e, também, excepcionalmente, em função da sua posição de destaque no ordenamento jurídico, com as emendas constitucionais.

TABELA 1
Tipos de normas federais mais recorrentes e Emendas Constitucionais em relação ao total de registros do Legin Web (Total de registros = 225.843)

Tipos	Quantidade em relação ao total de registros do Legin	Porcentagem em relação ao total de registros do Legin	Amostra (Margem de erro de \pm 10%)
Decreto	151.831	67,25%	100
Decreto Sem Número	16.318	7,23%	27
Lei Ordinária	13.973	6,19%	23
Decreto Legislativo	12.657	5,61%	21
Decreto-Lei	12.375	5,48%	21
Medida Provisória	6.677	2,96%	11
Emenda Constitucional	119	0,05%	1
Total	213.950	94,77%	204

Fonte: Autor.

A amostra de cada tipo de norma usou número ideal estimado para um nível de confiança de 95% com margem de erro de \pm 10%, com base na

tabela de Arkin e Colton, citada por Tagliacarne²⁴ (1976 apud GIL, 2007, p. 124). Se a quantidade de atos normativos referente a determinado tipo (p) for maior que 50% do Legin, usa-se o valor apontado na referida tabela. Esse é o caso do Decreto, que constitui 67,25% da Base, o que corresponde na tabela a uma amostra de 100 atos normativos para a margem de erro de $\pm 10\%$. O cálculo da amostra para tipos de normas com quantidade (p) inferior a 50% dos registros do Legin usou a fórmula “Fator de Tagliacarne (FT) = $4\{p(1-p)\}$ ”.

Os elementos a serem analisados de cada tipo normativo foram selecionados conforme os anos (um de cada ano, geralmente o primeiro ato exibido na lista de resultados por tipo de norma), de forma decrescente, de 2012 (geralmente) até a década de 1930, quando foram detectados atos normativos ainda sem o campo de indexação. Nesses casos, para a composição da amostra necessária, voltou-se a 2012, selecionando novamente de forma decrescente um ato normativo por ano até atingir a amostra desejada.

Com o objetivo de realizar a parte qualitativa da pesquisa, foram selecionados alguns servidores representantes de áreas que têm a legislação federal como base para a execução de suas atividades, onde se consulta o Legin com certa regularidade. Também é importante dizer que no contexto qualitativo, o critério de seleção dos sujeitos e da amostra foi, assim como relata Vergara (2010, p. 242), intencional e por acessibilidade.

Intencional porque foram escolhidas pessoas que desempenhavam atividades que, em tese, deveriam envolver a consulta à legislação. E, por acessibilidade, porque, a princípio, foram selecionadas pessoas do círculo de relacionamentos da pesquisadora, que teriam disponibilidade de tempo para responder o questionário e lidavam com a legislação.

Entretanto, a pesquisa não ficou focada nesse círculo. Inclusive, algumas pessoas envolvidas com atividades que poderiam implicar na consulta por assunto à legislação foram contatadas, aleatoriamente, conforme suas áreas de atuação, por meio do telefone, e consultadas sobre sua disponibilidade para responder ao questionário.

O questionário foi enviado por e-mail. Nele foi informado que os dados seriam anonimizados e a monografia resultante da pesquisa

²⁴ TAGLIACARNE, G. **Pesquisa de mercado**. São Paulo: Atlas, 1976.

disponibilizada na Biblioteca Digital da Câmara dos Deputados. Dados anonimizados significam, neste estudo, que não só os nomes, como os cargos dos respondentes não seriam identificados nos resultados da pesquisa. Mesmo porque a primeira questão do questionário aplicado tratava da descrição da atividade, o que era suficiente para a definição da utilização da pesquisa na legislação federal via Legin Web.

O quantitativo de servidores foi limitado conforme o método *snowball* (bola de neve). Segundo Baldin e Munhoz (2011, p. 332), *snowball* é uma forma de amostra não probabilística utilizada em pesquisas sociais em que:

Os participantes iniciais de um estudo indicam novos participantes que por sua vez indicam novos participantes e assim sucessivamente, até que seja alcançado o objetivo proposto (o “ponto de saturação”). O “ponto de saturação” é atingido quando os novos entrevistados passam a repetir os conteúdos já obtidos em entrevistas anteriores, sem acrescentar novas informações relevantes à pesquisa (WHA, 1994). Portanto, a *snowball* (“Bola de Neve”) é uma técnica de amostragem que utiliza cadeias de referência, uma espécie de rede.

Os servidores selecionados e indicados receberam um questionário semi-estruturado (corresponde ao APÊNDICE A – Questionário e respostas) que procurou averiguar, assim como na proposta quantitativa, tipos de assuntos recorrentes, porém agora nas buscas por assunto na legislação. Além disso, também foram checadas as necessidades de busca, ou se eram comuns buscas por assuntos mais genéricos, ou específicos e se ocorriam pesquisas que necessitavam restringir a busca a determinadas partes do texto do ato. Os dados coletados tinham a função de apontar as necessidades informacionais da organização e estão alinhados com as decomposições de estrutura e de conteúdo do ato normativo, adotadas aqui com o propósito da geração de metadados de assunto pertinentes às demandas da Câmara.

O levantamento da literatura sobre metadados e também sobre a legislação federal brasileira considerou como critério para a delimitação do período o surgimento do Dublin Core em 1995 e uma importante conferência sobre o tema em 2001, a *International Conference on Dublin Core and Metadata Applications*, onde foram discutidas iniciativas para a criação de esquemas de metadados para diferentes áreas, com o fim de evitar o surgimento caótico de esquemas de metadados.

Assim, o período do levantamento abrangeu principalmente os últimos quinze anos, de 1997 a 2012, no que refere à literatura em língua portuguesa e inglesa. No caso da língua inglesa, contudo, foi realizada uma varredura genérica, focando somente em padrões de metadados reconhecidos pela Ciência da Informação, ou segundo a literatura, bastante utilizados por profissionais da referida área.

A análise dos resultados da pesquisa considerou os metadados de assunto comumente citados na literatura que são basicamente: nome pessoal, entidade coletiva, evento, título uniforme, cronológico, tópico, geográfico, gênero/forma, ocupação ou atividade, função ou propósito de criação do documento, adaptado na análise para função/ação.

Tal adaptação corresponde a ideia de que, no caso dos atos normativos analisados, a ação ou “processo” presentes no texto legal, conforme cita Guimarães (2004, p. 27 et seq.) em seu trabalho sobre a construção de ementas de atos normativos, tem caráter relevante e estão vinculados ao propósito do documento.

Excepcionalmente, a análise foi realizada com base em tipos de metadados de assunto não tão citados na literatura, mas ligados ao tratamento de informações governamentais.

Quando necessário, para o clareamento dos dados da pesquisa, possíveis subdivisões dos metadados de assunto foram analisadas nos resultados do estudo. Isso ocorreu extraordinariamente, pois subdivisões constituem um nível granular maior de metadados de assunto, que foge ao escopo do trabalho.

Por fim, também foram realizadas análises com base em tendências de busca de informação, apoiadas na triangulação da pesquisa.

4 DISCUSSÃO DOS RESULTADOS

Nos resultados são apontados os principais aspectos relacionados com a identificação de metadados de assunto para a legislação federal brasileira, com base tanto na pesquisa da literatura, como na parte quantitativa e qualitativa do estudo focada no sistema Legin Web e na Câmara dos Deputados.

A apresentação dos resultados das partes quantitativa e qualitativa da pesquisa tem como fundamento de sua organização a decomposição da estrutura e conteúdo de atos normativos federais, com o objetivo de gerar tipos de metadados de assunto mais específicos, portanto um tratamento mais granular dos assuntos abordados na legislação federal brasileira.

4.1 ANÁLISE DOS DADOS QUANTITATIVOS

A decomposição da estrutura de um ato normativo pode se dar tanto por meio de relacionamentos, como é o caso da publicação original de um ato normativo com o seu texto atualizado, caso exista, como pela decomposição do ato em partes menores, a exemplo das partes de atos normativos estipuladas pela Lei complementar nº 95, de 1998 (1998, art. 10). No caso de relacionamentos, um metadado de assunto de relacionamento pode ser criado, a exemplo dos metadados de relacionamento utilizados pelo Dublin Core. A TAB. 2 a seguir trata dessa decomposição da estrutura dos atos normativos.

TABELA 2
Decomposição da estrutura de atos normativos federais

Tipos de atos normativos	Amostra	Somatória dos artigos dos atos originais (texto topicalizado)	Com textos atualizados ou Convertidos em Lei ¹	Textos atualizados ou Convertidos em Lei com assuntos novos ²
Decreto	100	1952	22	10
Decreto Sem Número	27	106	2	0
Lei Ordinária	23	895	20	14
Decreto Legislativo	21	42	0	0
Decreto-Lei	21	122	7	1
Medida Provisória	11	59	7	0
Emenda Constitucional	1	3	0	0
Total	204	3179	58	25
%	100%	100%	28,4%	12,3%

Fonte: Autor.

Notas: ¹ A conversão em Lei, aplica-se à Medida Provisória, que poderá também apresentar alterações, no caso de ser convertida em Lei.
² O ato normativo foi considerado com assunto novo quando ocorreram acréscimos de artigos ao texto legal original.

O que se pode observar, no que refere à decomposição de atos normativos, é que eles são bastante topicalizados se considerarmos que 204 normas geraram 3179 artigos. Isso reafirma a consideração de Pereira (2004, p. 155) quando cita como uma das características dos atos legais “textos com tópicos (artigos, incisos, alíneas e parágrafos) cada um com uma função”. Além disso, a Lei complementar nº 95, de 1998 (1998, art. 11, inciso III, alínea b) diz que cada artigo da lei deve se restringir a um único assunto ou princípio.

Outra característica confirmada dos atos normativos é o número elevado de relacionamentos entre os atos, já que 58 dos 204 analisados possuem texto atualizado, ou seja, aproximadamente 28,4% do total de normas analisadas.

Ainda há o acréscimo de artigos nos textos atualizados, o que implica em novos assuntos, com base na restrição de único assunto ou princípio a cada artigo do ato legal, conforme mencionado na Lei complementar nº 95, de 1998. Os textos legais anteriores a 1998, apesar de não seguirem obrigatoriamente tal disposição, também eram regidos pelas normas linguísticas, que pedem para a obtenção de clareza e coesão, a concatenação de ideias em unidades ou parágrafos menores, o que se supõe aplicar também aos atos normativos, mesmo por causa de sua natureza de instrumento público.

Portanto, a TAB. 2 sugere que metadados de assunto delimitadores da localização do assunto dentro do ato normativo, assim como delimitadores de texto atualizado, que inserem novos assuntos no texto legal original, podem ser de grande importância, não só para a eliminação de resultados não desejados em pesquisas na legislação federal brasileira, como para uma melhor qualificação do assunto tratado por meio de metadado de assunto, delimitado por estrutura e/ou relacionamento do ato normativo.

A TAB. 3, a seguir, apresenta os dados de decomposição da estrutura do campo de indexação dos atos normativos federais brasileiros amostrados. A última coluna trata da coincidência dos dados desses campos com os dados do campo de indexação da proposição legislativa, ou Medida

Provisória em apreciação pelo Congresso Nacional, que deram origem aos atos normativos analisados. As proposições legislativas e Medidas Provisórias, as últimas por serem apreciadas pelo Congresso, são registradas em outro banco de dados, no caso o Sistema de Informação Legislativa (Sileg). As Medidas Provisórias também são inseridas no Legin Web.

TABELA 3
Decomposição da estrutura do campo de indexação de atos normativos federais

Tipos de atos normativos	Amostra	Número total de assuntos exclusivos ¹	Cadeias/Estruturas de assuntos exclusivos ²	Assuntos periféricos ³	Assuntos modificadores ⁴	Assuntos coincidentes com a proposição/ Medida Provisória em apreciação no CN ⁵
Decreto	100	726	251	475	173	Não procede
Decreto Sem Número	27	239	44	195	21	Não procede
Lei Ordinária	23	1104	257	847	163	344
Decreto Legislativo	21	120	33	87	14	21
Decreto-Lei ⁶	21	297	55	242	48	Não procede
Medida Provisória	11	201	32	169	29	85
Emenda Constitucional	1	33	8	25	3	6
Total	204	2720	680	2040	451	456
%	100%	100%	100%	75%	16,6%	16,8%

Fonte: Autor.

Notas: ¹ Assuntos são chamados de exclusivos porque as duplicações de assunto dentro do campo de indexação de um mesmo ato normativo foram eliminadas. Porém assuntos duplicados entre os atos foram mantidos.

² Cadeias/estruturas de assunto correspondem, no campo de indexação dos atos normativos do Legin, a àqueles conjuntos ou linhas de assuntos iniciadas por um assunto grafado em caixa alta. Esse assunto, em caixa alta, foi utilizado para a contagem das cadeias. Ex.: NEOPLASIA MALIGNA - Tratamento médico - Paciente - Comprovação - Diagnóstico de saúde - Laudo médico - Prazo - Início - Sistema Único de Saúde (SUS) - Quimioterapia - Radioterapia - Cirurgia - Direitos – Gratuidade (assunto referente à Lei nº 12.732, de 22/11/2012). Os assuntos em caixa alta das cadeias/estruturas também foram selecionados de forma exclusiva, para manter a linearidade da análise, mas a contagem desses tipos de assunto, com base no campo de indexação original, pode ser significativamente maior.

³ No caso da legislação, os assuntos periféricos são todos os assuntos que vierem após o assunto em caixa alta, na cadeia de assunto. Entre os assuntos periféricos temos os definidos pela ISO 2788 (1986, p. 3-7) como modificadores. Entretanto, alguns desses assuntos podem não corresponder exatamente à categoria de modificadores, mas têm o papel de modificação do assunto em destaque.

⁴ Assuntos modificadores, de acordo com a ISO 2788 (1986, p. 3-7), constituem essencialmente em um ou mais componentes posteriores ao assunto foco do documento, que restringem a extensão desse assunto, eliminando possíveis ambiguidades.

⁵ As Medidas Provisórias, quando ainda em fase de apreciação no Congresso, são indexadas no Legin e também no Sistema de Informações Legislativas (Sileg) que contem, basicamente, as proposições legislativas. Isso possibilitou a análise comparativa da indexação disponível no Sileg com a do Legin. No caso dos Decretos e Decretos Sem Número, que configuram atos

administrativos emanados do Poder Executivo, não existem proposições originárias e, consequentemente, nem indexação, motivo pelo qual a análise comparativa dos assuntos não procede.

⁶ Os Decretos-Lei precedem as Medidas provisórias, originárias da Constituição Federal de 1988, sendo atos emanados do Poder Executivo, e submetidos ao Poder Legislativo, contudo os Decretos-Lei não possuem proposições relacionadas e não são inseridos no Sileg, o que inviabilizou a análise comparativa da indexação.

A coincidência de 456 assuntos ou 16,8% dos 2720 assuntos analisados, entre a indexação do ato normativo e da proposição legislativa, vem ratificar o cenário apresentado por vários autores, entre eles Zeng (2010, p. 187) e Marcondes (2006, p. 98), sobre o tratamento de conjuntos de metadados, que podem ser intercambiados entre provedores de dados diferentes, como uma biblioteca digital, um repositório de dados abertos ou mesmo entre bases diferentes. Isso eliminaria a duplicidade de trabalho e aumentaria o poder de localização de recursos informacionais se for levado em conta que, em muitos casos, áreas especializadas poderão fornecer metadados específicos da área, inclusive os de assunto.

A primeira coluna (número total de assuntos exclusivos) mostra a incidência de assuntos tratados por cada ato normativo, o que dá uma média geral de 13 assuntos aproximadamente para cada ato normativo, sendo que alguns tipos normativos, como o Decreto Legislativo, têm uma média menor, igual a aproximadamente 5,7 assuntos por ato.

Assim, essa particularidade de grande número de assuntos presentes na indexação da legislação federal gera uma maior revocação, ou muitos itens irrelevantes e resultados de pesquisa que, muitas vezes, precisam ser avaliados um a um, manualmente. Tal observação ratifica a importância do desenvolvimento de metadados de assunto específicos, que possam propiciar novas formas e filtros de recuperação para a legislação federal brasileira.

A coluna 4 (cadeias/estruturas de assunto) tornou possível ter uma ideia do que teve mais relevo na leitura técnica, para a indexação do ato normativo, os assuntos marcados em caixa alta, e mesmo a própria ordem deles dentro do campo de indexação. No entanto, essa é uma informação perdida para o sistema de recuperação do Legin Web. Em princípio, porque só existe um campo para se colocar todos os assuntos, dificultando a atribuição de regras de valoração aos assuntos, conforme sua importância no documento. Assim, seria oneroso, em termos de sistemas e desenvolvimento, atribuir pesos diferenciados para assuntos que estejam somente grafados diferentemente.

Por outro lado, se houvessem campos repetitivos ou metadados de assuntos repetitivos, seriam criadas novas possibilidades, pois a repetição significa que a cada assunto novo, seriam criados novos metadados de assunto ou campos, aos quais poderiam ser adicionados controles e regras diferenciadoras como: definir se o metadado tem *peso prioritário* ou *secundário*, se é *obrigatório* ou *opcional*, entre outras coisas.

Aliás, controles e regras adicionados aos metadados são tão fundamentais quanto definir os próprios metadados. São eles que garantirão no futuro, a integração entre sistemas e a troca de metadados ou conjuntos de metadados. Outra característica importante dos controles e regras é permitir o rastreamento dos metadados tanto por máquinas como pelos seres humanos, a fim de identificar sua origem no caso de intercâmbios.

Os controles e regras ainda contribuem para a exibição hierárquica que propicia a navegabilidade, como no caso da organização por *clusters* ou agrupamentos de resultados de buscas, conforme relatório da *Library of Congress* (2007, p. 4 et seq., tradução nossa) e também outros autores como Marcondes (2006, p. 96).

Assim como recomenda o MARC 21, até mesmo para campos de assuntos livres, como o 653 são necessários controles. Nesse campo, o indicador 1 refere-se ao nível primário (opção 1 – indica que o assunto cobre o principal foco temático do material ou constitui um assunto principal) e ao secundário (opção 2 – o assunto cobre aspectos menos importantes do material).

Os assuntos periféricos da coluna 5 da TAB. 3 foram obtidos pela subtração dos valores da coluna 4 (Cadeias/Estruturas de assuntos exclusivos) dos valores da coluna 3 (Número total de assuntos exclusivos) e representam 75% dos assuntos analisados. Essa coluna de periféricos também mostra a importância de se aumentar a granularidade dos metadados de assunto dos atos normativos.

Apesar dos assuntos modificadores também não deixarem de ser assuntos periféricos, foram tratados à parte. Isso porque representam, neste estudo, basicamente, os processos ou ações ligados ao caráter imperativo do ato normativo, os quais, segundo Guimarães (2004, p. 29), merecem um tratamento diferenciado porque visam “atingir o modo pelo qual o pesquisador

solicita a informação”. Assim, são fortes candidatos à formação de novo metadado de assunto para a legislação federal. Os assuntos modificadores representaram 16,6 % dos assuntos analisados e foram extraídos da linha “Função/Ação” da TAB. 4, na sequência.

Os metadados de assunto reincidentes identificados na literatura (nome pessoal, entidade coletiva, evento, título uniforme, cronológico, tópico, geográfico, gênero/forma, ocupação ou atividade, função/ação adaptado) foram utilizados como categorias de análise referentes aos tipos de assuntos encontrados no campo de indexação dos atos legais analisados.

Conforme pode ser visto na TAB. 4, os tipos de assuntos mais representativos foram: tópico, função/ação, entidade e geográfico, respectivamente e aproximadamente, 66,1%, 16,6%, 9,6% e 4,5%. Os assuntos pessoal, cronológico, evento, título uniforme, gênero/forma e o assunto agrupador (que surgiu ao longo da própria análise) não foram significativos.

Entretanto, o único que realmente zerou foi o assunto pessoal, o qual segundo Andreta (2004, p. 95) é tema, de modo geral, dos Decretos Sem Número, que são “atos do Poder Executivo, de caráter administrativo, mas interessando apenas a um indivíduo ou grupo de indivíduos, como nomeações, [...] exonerações, aposentadorias e reformas de funcionários, ministros ou secretários, embaixadores, delegados e militares”, etc. Isso remete à ideia de que metadados devem ter previsibilidade, portanto atender ao máximo de situações, mesmo que não sejam tão reincidentes em um ou outro momento.

TABELA 4

Decomposição do conteúdo do campo de indexação de atos normativos federais (assuntos sem repetição dentro do ato ou exclusivos)

Tipos de ato normativo	Decreto	Decreto Sem Número	Lei Ordinária	Decreto Legislativo	Decreto -Lei	Medida Provisória	Emenda Constitucional	Total	%
Amostra	100	27	23	21	21	11	1	204	100%
Tópico	408	118	779	82	221	161	29	1798	66,1%

Continuação	Pessoal	0	0	0	0	0	0	0	0	0	-
Tipos de Assunto	Entidade	88	56	82	5	22	8	0	261	9,6%	
	Geográfico	33	43	25	16	2	2	1	122	4,5%	
	Cronológico	3	0	7	2	0	0	0	12	0,4%	
	Evento	2	1	0	0	0	0	0	3	0,1%	
	Título uniforme	0	0	1	0	0	0	0	1	0,04%	
	Gênero/Forma	3	0	0	0	0	0	0	3	0,1%	
	Ocupação/Atividade	16	0	45	1	4	1	0	67	2,5%	
	Função/Ação	173	21	163	14	48	29	3	451	16,6%	
	Assunto Agrupador	0	0	2	0	0	0	0	2	0,1%	
	Total	726	239	1104	120	297	201	33	2720	100%	
%	26,7%	8,8%	40,6%	4,4%	10,9%	7,4%	1,2%	100%	-		

Fonte: Autor.

O assunto tópico foi utilizado para tipos de assuntos que não se enquadravam em nenhuma das outras categorias. Contudo, essas outras categorias também podem assumir o papel de assunto tópico, quando representarem a temática do documento de forma preponderante.

Assim, as categorias não são necessariamente excludentes, como acontece inclusive com o MARC 21, no qual alguns assuntos podem ser tanto tópicos, como subdivisões de outros assuntos. Um exemplo disto é o caso do assunto “Aposentadoria”, que também pode ser uma subdivisão sob a classe de pessoas e grupos étnicos, conforme está na lista de autoridades da Fundação Biblioteca Nacional (2012), originada de tradução do *Library of Congress Subject Headings*.

No entanto, isso não interfere na definição dos metadados de assunto, pois o ideal é que existam os metadados, os quais serão preenchidos conforme o contexto do assunto no documento indexado. De fato, esse estudo terminológico cabe ao tesouro.

O título de um ato normativo corresponde a sua epígrafe, a qual segundo a Lei complementar nº 95, de 26/12/1998 (Art. 4º) “propiciará identificação singular à lei e será formada pelo título designativo da espécie normativa, pelo número respectivo e pelo ano de promulgação.” A categoria de assunto “título uniforme” foi utilizada, a princípio, para apelidos de normas, apesar de que existe no Legin Web um campo específico para “Apelido”. O apelido aqui foi tratado como uma forma variante de título, de maneira semelhante ao que ocorre no MARC 21, como um “título assunto”, no campo 630.

O metadado de assunto agrupador foi pouco detectado na análise do campo de indexação, mas tem um papel de “fio condutor”, porque perpassa vários documentos, permitindo que sejam recuperados de uma única vez, sem grandes esforços. Entre exemplos de assuntos agrupadores temos: legislação eleitoral, legislação tributária federal, leis orçamentárias, ato internacional, programas governamentais, etc.

Há no Legin Web um campo chamado “tema”, que remete aos macrotemas do tesouro, os quais, por sua vez, derivam das 20 (vinte) comissões temáticas da Câmara dos Deputados. No entanto, ele apresenta assuntos bem mais genéricos que os considerados nos assuntos agrupadores.

Na sequência, a TAB. 5 se refere às buscas no Legin Web, tanto por assunto, como por número/ano do ato normativo e tipo de norma, realizadas por usuários internos e externos da Câmara dos Deputados, de 01/09/2012 a 30/09/2012.

TABELA 5
Usuários internos e externos da Câmara dos Deputados:
tipos de buscas realizadas no Legin Web de 01/09/2012 a 30/09/2012

Tipos de buscas realizadas no Legin Web	Total	%
Número/Ano	361	53,4%
Tipo de norma	179	26,5%
Tópico	94	13,9%
Pessoal	0	-
Entidade coletiva	13	1,9%
Geográfico	3	0,4%

Continuação

Cronológico (tempo como assunto, não como parte da epígrafe do ato normativo)	0	–
Evento	0	–
Título uniforme	25	3,7%
Gênero/Forma	0	–
Ocupação/Atividade	0	–
Função/Ação dependente de outro assunto necessariamente	0	–
Assunto Agrupador	1	0,1%
Total	676	100%

Fonte: Autor.

A análise dos dados dos tipos de buscas por assunto realizadas no Legin Web usou como categorias de análise os assuntos: tópico, pessoal, entidade, geográfico, cronológico, evento, título uniforme, gênero/forma, ocupação/atividade, função/ação, agrupador. Foram os mesmos tipos de assunto utilizados na análise dos assuntos do campo de indexação dos atos normativos disponíveis no Legin Web.

Para se ter um panorama mais genérico de quais tipos de buscas ocorrem no Legin Web, houve ainda um acréscimo de duas outras categorias. No caso, não de tipos de buscas por assunto, mas de outros tipos de busca, por demonstrarem uma tendência de busca apontada na literatura, que também se confirmou na análise dos dados qualitativos, as categorias de buscas por “Número/Ano” e “Tipo de norma”. Essas duas categorias somaram 79,9% (540) das 676 pesquisas realizadas.

Isso indica que quando os usuários chegaram a realizar a pesquisa no Legin Web já tinham conhecimento do ato normativo exato que estavam procurando, ou seja, outra fonte deve ter sido utilizada, caso a pesquisa tenha se iniciado por assunto.

Tal situação vem ao encontro de algumas tendências apontadas no relatório da OCLC (2004, p. 16) que são a redução no acesso guiado ao conteúdo, ou a maioria das pessoas que buscavam informações recorria diretamente à Internet, especialmente a motores de busca como o Google, não se dirigindo a bibliotecas ou qualquer outro tipo de serviço de informação. Outra tendência apontada foi a desagregação não só de conteúdo, mas de serviços, da tecnologia e das instituições. Tal discussão está intimamente

ligada com a forma como buscamos e encontramos informação, seja ela de que natureza for.

Isso afeta diretamente os recursos de busca por assunto. Até pode levantar questões a respeito da validade deles, mas também em relatório, agora do *Cataloging Policy and Support Office* da *Library of Congress* (2007, p. 8, tradução nossa), são apontadas vantagens dos vocabulários controlados. O relatório é baseado em estudos de Lois Mai Chan, Elaine Svenonius, e Arlene Taylor/Danny Joudrey, e apresenta uma declaração postada por David Bade, da Universidade de Chicago, a qual se refere aos vocabulários controlados, mas pode ser estendida para os metadados de assunto. Assim, segundo ele, uma das vantagens dos metadados de assuntos e dos vocabulários controlados é contribuir para:

Relevância dos resultados de pesquisa recuperados. Cabeçalhos de assunto pré-coordenados podem ser usados para melhorar a classificação de relevância de sistemas como o Google. [...] A razão principal para isso é que os cabeçalhos de assunto são o produto de decisões humanas que apresentam a mesma natureza dos julgamentos de relevância feitos pelos pesquisadores, enquanto uma classificação de relevância a *la* Google e Endeca nunca poderá ser outra coisa além de análises estatísticas de probabilidades com base em substitutos.

Não deixa de ser um ponto válido. Mas é inegável que o Google incrementou muito seus recursos. A busca por partes de texto quando se tem um conhecimento maior do item procurado é extremamente útil, e parece ter sido bastante utilizada para substituir buscas por assunto, a rigor, na legislação. Talvez, esse seja um dos motivos porque as buscas por assunto referentes à TAB. 5 somaram apenas 20,1% (136) das buscas. No que se refere ao escalonamento dos tipos de assunto consultados, o resultado ficou bem parecido com o da análise do campo de indexação, com poucas variações.

A ordenação dos tipos de assunto mais procurados ficou assim (todos com valores aproximados): tópico 13,9% (94), título uniforme 3,7% (25), entidade coletiva 1,9% (13), geográfico 0,4% (3) e assunto agrupador 0,1% (1). Os outros, assunto pessoal, cronológico, evento, gênero/forma, ocupação/atividade, função/ação zeraram. O título uniforme teve um destaque porque foram muitas consultas por apelidos ou títulos variantes dos atos

normativos, como Código Civil, Código penal, Código de processo penal, Código florestal, Código eleitoral, Código de trânsito, Código ambiental, etc.

A outra variação que ressalta aos olhos foi assunto função/ação ter zerado. Parece incomum, já que a literatura afirma que os processos/ações são formas importantes de pesquisa, por outro lado, pode demonstrar um desconhecimento por parte dos que recorrem ao assunto, ou seja, pode sugerir que são usuários mais inexperientes sobre a legislação federal brasileira, enquanto os mais experientes consultam diretamente pela norma, não recorrendo ao assunto. Essas são possíveis interpretações.

E tal provável inexperiência suscita a ideia da adoção não só de metadados para termos livres, mais próximos da realidade dos usuários, mas a abertura do sistema para a inserção de termos pelo próprio usuário, como sugere o relatório do *Cataloging Policy and Support Office da Library of Congress* (ibid., p. 10, tradução nossa).

Também permitir aos usuários adicionar seus próprios termos de acesso ou "folksonomias" em registros bibliográficos para aumentar o acesso descontrolado. Bibliotecários de referência, especialmente, deveriam ser encorajados, se não convocados, a contribuir para este esforço.

Enfim, definir quais aspectos devem ser considerados na identificação de metadados de assunto envolve inúmeras ações, como o estudo constante da evolução tanto das tecnologias de busca de informação, como das funcionalidades atribuídas aos metadados e da própria busca por assunto. Em seguida será feita uma breve análise dos dados qualitativos do trabalho, destacando pontos não tratados pelos quantitativos e reforçando os recorrentes.

4.2 ANÁLISE DOS DADOS QUALITATIVOS

Foram enviados 28 questionários, dos quais 20 foram respondidos. As questões e respostas estão no Apêndice A, QUADRO 2. Os dados foram anonimizados conforme informado aos respondentes no próprio questionário.

A questão 1a) pretendia identificar as atividades desenvolvidas pelos respondentes que implicavam na busca por assunto na legislação federal. Isso devido à atividade ser determinante para o desenvolvimento de metadados de

assunto orientados para o atendimento, no caso dos usuários internos da Câmara.

Detectou-se que dos 20 respondentes, 45% (9) usam a busca por assunto na legislação federal para prestar informações, 40% (8) para elaborar parecer ou relatório a respeito da legislação, 50% (10) para assessorar na elaboração ou elaborar proposição legislativa, 40% (8) para acompanhar a evolução da discussão de proposições legislativas e 55% (11) responderam “Outras”. Notou-se que alguns respondentes realizam mais de uma das atividades elencadas.

No caso do respondente optar por outras atividades na 1a), era solicitado que essas “Outras” fossem indicadas na questão 1b). Assim, foram enumeradas basicamente: resumos de proposições legislativas para pautas de Comissões, elaboração de projeto de lei de compilação, estudos, notas técnicas, artigos, livros técnicos, anexação de legislação citada a proposições, atualização de normas jurídicas, compilação da legislação para publicação temática, revisão/correção de dispositivos regulamentados ou passíveis de regulamentação da Constituição de 1988, confirmação dos dados da norma para indexação de discurso que a citou, análise de indexação de proposições e busca de informações sobre leis votadas na Câmara.

A questão 2a) perguntava se o respondente se lembrava ou não das últimas três pesquisas por assunto realizadas na base de legislação federal brasileira disponível no Portal da Câmara. A maioria dos respondentes (55%, 11) respondeu “Sim”.

A questão 2b) pedia para citar quais foram os assuntos pesquisados. As respostas apresentaram os seguintes: legislação trabalhista, idoso, teorias do direito penal, crimes virtuais, anteprojeto da nova lei de direitos autorais, anistia de condenado pela justiça federal brasileira, concessões, energia elétrica, energia solar, Lei Maria da Penha, Lei de Tóxicos, Lei Anticorrupção, Comissão de Ética Pública, isenção de impostos sobre produtos da cesta básica e Medidas Provisórias, terceirização no serviço público, Código florestal, equipamentos obrigatórios aos motociclistas (antena corta fio), licença-gestante, Programa Bolsa Família, Corregedoria da Câmara dos Deputados, procedimentos investigatórios, informações sigilosas, lei de regulamentação da

profissão de motorista. Todos os assuntos citados podem ser encaixados nas categorias de análise utilizadas na TAB. 4 deste estudo.

A questão 3a) solicitou aos respondentes que dissessem se já precisaram realizar alguma pesquisa que levantasse toda a legislação federal sobre um determinado tema. A resposta massificadora foi “Sim”, somando 75% (15) dos respondentes. Essa questão está ligada ao metadado de assunto agrupador, ou pesquisas que precisam rastrear temas mais genéricos, e confirma tendências de pesquisa da legislação levantadas pela literatura.

A questão 3b) deu continuidade a 3a) pedindo para que fossem citados os temas genéricos pesquisados. Os temas mencionados foram: matéria trabalhista, datas comemorativas desde o Império, punição para a condução de veículos automotores sobre efeitos de drogas ou álcool, direito tributário (geralmente de determinado setor ou atividade), energia elétrica, legislação anticorrupção, mensagens presidenciais que encaminharam atos internacionais de 1988 a 2012, isenção de IPI, processo administrativo disciplinar de servidor público, tomada de contas especial, meio ambiente, legislação indígena, legislação eleitoral e sobre a mulher.

As questões 2c), 2d), 3c) e 3d) tratavam das dificuldades encontradas na pesquisa por assunto na legislação federal. Nas questões 2c) e 3c), respectivamente, 30% (6) e 60% (12) responderam “Sim”, ou que tiveram dificuldades. As questões 2d) e 3d) solicitavam que as mesmas fossem mencionadas.

Assim, entre as dificuldades apontadas estavam: excesso de dados não pertinentes ou grande revocação, desatualização, ordenação por relevância não muito boa, ausência de legislação infraconstitucional, de portarias e de normas técnicas, incompletude, problemas de *status* de vigência, *layout* inapropriado, falta de *links* para normas referenciadas, insuficiência de filtros, falta de combinadores booleanos e problemas de indexação. Sendo que no que se refere à ordenação por relevância deficiente, o respondente citou “o retorno pode não corresponder à realidade (da importância das normas) tornando a pesquisa trabalhosa”.

Entre os problemas de indexação foram apontados pelos respondentes: indexação mal feita ou ausente e falta de coerência na

indexação. Inclusive os Respondentes 5 e 10 (cf. QUADRO 2 no Apêndice 1) exemplificam, respectivamente:

As leis do setor de energia elétrica utilizam ambos os termos “energia elétrica” e “eletricidade”. Quando se faz a procura de legislação empregando conjuntamente esses termos a resposta é bem menor do que quando se faz a pesquisa isolada de cada um deles. Portanto, o resultado da pesquisa com termos conjuntos é do tipo “termo 1” E “termo 2”. Eventualmente, seria interessante poder optar pela pesquisa ser do tipo “termo 1” OU “termo 2”.

Consegui efetuar o levantamento exaustivo apenas quando não coloquei filtro algum em “Mensagem”, fazendo a seleção daquelas referentes a atos internacionais manualmente. Verifiquei que nem sempre os pactos internacionais estão catalogados como “atos internacionais”. Outros, só encontrei por acaso, ao pesquisar algum outro dado, mas, cruzando as ferramentas do sistema foi possível fazer um levantamento exaustivo de tudo que o Executivo encaminhou ao Congresso no período pesquisado.

Todos os problemas acima têm impacto na geração de metadados de assunto. De fato, os dados qualitativos ratificam o apontado pelos dados quantitativos, e mostram, mais uma vez, a necessidade de mais opções para navegação e ranqueamento de resultados. As questões de interface e de falta de completude e coerência de dados também comprometem tanto a confiabilidade da base como um todo, como dos metadados. Ou não basta ter uma base com metadados estruturados, é preciso que sejam claras as orientações sobre sua alimentação e utilização.

A questão 4a) tentou checar o domínio que os respondentes tinham do sistema de busca por assunto, para entender a origem das dificuldades, e, ao mesmo tempo, ver se algum novo metadado de assunto surgia. A referida questão perguntava se em alguma pesquisa não foi encontrado nenhum resultado. Apenas 30% dos respondentes disseram “Sim”.

A questão 4b) apontou os assuntos não encontrados: crimes virtuais, Dia das Artes, assuntos que são regulamentados pelos estados, municípios ou normas técnicas, como construção e manutenção de calçadas públicas. Alguns respondentes disseram que não se lembravam e que em situações assim recorrem a outros sites como o do Planalto, do Senado Federal e ao próprio Google.

As questões 4a) e 4b) mostram uma necessidade de ampliar os escopos, e, pelo menos, criar mais vínculos entre legislação federal e

infralegal, mesmo que isso não seja missão primeira do Legin Web. Os metadados de assuntos podem ser uma ponte neste sentido, pois fogem dos vínculos ou relacionamentos legais e baseiam-se nas ligações temáticas.

Um metadado de assunto com propósito estabelecido de agrupamento temático é apontado pelo Padrão de metadados do Governo Eletrônico (e-PMG), o metadado *Assunto.categoria*, o qual representa o assunto principal do recurso (o termo deve ser extraído do Vocabulário Controlado do Governo Eletrônico - VCGE) e permite aos usuários navegarem nos diretórios dos portais, de categorias mais amplas para mais específicas.

A extração de metadados de outras fontes, como o e-PMG, desde que devidamente identificados, é um ganho do ponto de vista da economia de recursos e da própria interoperabilidade, bem como do alcance da informação. E a possibilidade de navegação temática ou taxonômica, ou mesmo na forma de *clusters* ou agrupamentos de resultados, dá direção a pesquisas com resultados negativos, pois o usuário tem mais recursos para se certificar de que o que não encontrou, realmente não existe. Isso vem ao encontro da explanação do respondente 13.

Não encontrar é um problema? Pode não existir a legislação procurada. Em minha opinião, a questão relevante é como garantir se a informação/conteúdo procurada não existe ou se simplesmente não foi encontrada.

O item 5a) do questionário tratava da decomposição da estrutura de um ato normativo, como meio de restrição para a pesquisa de metadados de assunto. A maioria dos respondentes 60% (12) respondeu “Sim” ou que já havia pesquisado em partes específicas de atos normativos. A questão 5b) listava partes específicas de atos normativos. Todas as partes foram marcadas como pontos de pesquisa, porém as mais representativas foram Artigo (55%, 11), Parágrafo (50%, 10), Inciso (50%, 10), Alínea (45%, 9), Item (45%, 9) e Ementa (45%, 9).

A questão 5c) pedia que fossem citadas outras partes do ato normativo não elencadas na 5b). As respostas da 5c) constituíram mais em comentários sobre a própria questão, os quais podem ser vistos no QUADRO 2 (APÊNDICE A).

No caso da 5d) e 5e), questionou-se, respectivamente, se foi difícil a busca em partes específicas do ato normativo e quais foram as dificuldades encontradas. Houve empate na 5d), 35% (7) dos respondentes responderam que “Sim”, que tiveram dificuldades e os outros 35% (7) “Não” e 6 não responderam à questão. Entre as dificuldades citadas, os respondentes 6, 17 e 18 disseram, nessa ordem:

Lembro apenas que não havia a possibilidade de escolher itens e compara-los, por exemplo.

A pesquisa não permite detalhamento específico de item.

Falta de indexação ou indexação com termos livres.

Sucintamente, o que os usuários mencionaram é que a única opção de busca em partes do texto legal, fora a ementa, é a pesquisa em texto completo, o que aumenta bastante a possibilidade de resultados, principalmente em assuntos muito recorrentes na legislação federal brasileira. A busca por partes poderia ser tratada por um metadado de gênero/forma, pois um “artigo” de um ato normativo não deixa de ser a forma de uma parte do documento, assim como o “Anexo”, etc.

A questão 6) pedia que os respondentes apontassem novas formas de filtrar ou refinar as buscas por assunto na legislação disponível no Portal da Câmara. O intuito dessa pergunta era identificar, a partir da ideia do filtro, outros possíveis metadados de assunto, porque, afinal, os filtros também são uma forma de apresentação de metadados, ou uma maneira de evidenciar a funcionalidade dos metadados designada de “visibilidade”, conforme Campos (2007, p. 17).

Alguns preferiram não opinar na 6) e um respondente afirmou que acreditava serem suficientes os filtros existentes. Outros já apresentaram várias sugestões como: uso de combinadores booleanos “and” e “or”, exclusão de textos que contenham determinada palavra (not), seleção de normas que contenham termos ou expressão somente na ementa e promulgadas em um intervalo de tempo, filtro para normas em vigor e revogadas, critérios de pesquisa avançada e uso de aspas.

Alguns dos filtros e recursos solicitados já estão disponíveis. Contudo, essa discussão de filtros, de forma genérica, não cabe aqui, porque

muitos filtros não estão relacionados com metadados de assunto, e sim com metadados descritivos, de outra natureza. E mesmo os metadados de assunto não foram abordados aqui sob a perspectiva única de filtros, os quais compõem uma parte muito específica de um sistema de recuperação da informação.

Também surgiram propostas que parecem destoar dos objetivos do estudo, mas estão diretamente relacionadas, como as que falam da indexação e tradução de textos legais. Isso porque a indexação também estabelece de que formas se pode filtrar o assunto, e a tradução facilita ou promove a recuperação no caso de assuntos de documentos em outras línguas. Os respondentes 10, 17 e 18 disseram nesta ordem:

Sugeriria que, nos casos dos pactos internacionais celebrados pelo Brasil, quer bilaterais, multilaterais etc., independentemente de seu formato (tratado, convenção, convênio, acordo, protocolo etc., sejam indexados como "ato internacional", que é o nome genérico no qual cabem todos os tipos e feitios de avenças internacionais.

Gostaria de otimizar os produtos: apresentação de listagens, navegação pelos anos e tipos de legislação (como no site do Planalto).

Indexar cada parte do texto legal, mas sobretudo a matéria ali contida, não necessariamente as palavras.

Também tentaria fazer um serviço de tradução de textos legais relevantes sobre temas em discussão no Brasil e em outros países.

A questão 7) discorria sobre o uso de termos compostos ou palavras isoladas na recuperação de assuntos porque isso também é uma forma de decomposição, no caso do conteúdo. Foi utilizado o exemplo "Reforma tributária" ou "reforma", "tributo", "imposto", isoladamente. Muitos respondentes disseram que usam ambas as formas. A forma como essa decomposição se dá foi bem exemplificada pelos respondentes 1, 4, 10 e 13.

Se eu souber exatamente a expressão que consta da lei, procuro diretamente pelo termo composto. Se a pesquisa for mais ampla, procurando normas que tratem da matéria, uso palavras isoladas.

Depende da situação, no caso acima colocaria Reforma Tributária porque é um termo consagrado. Em outras situações poderia colocar, por exemplo: imposto de renda, dedução, juros sobre capital próprio.

Tento ambas as formas: ato internacional, por exemplo. Mas já descobri que, para exaurir todas as mensagens pertinentes recebidas pelo Congresso, tenho de verificar uma a uma de todas as mensagens encaminhadas pela Presidência da República, o que acaba incluindo assuntos que nada têm a ver com o tema da pesquisa, tais como afastamentos da Presidente; vetos; promulgações, concessões de rádio e televisão etc.

Reforma tributária, entretanto a aspas só é possível quando se sabe exatamente a expressão de busca. Quando os termos buscados não estão ligados diretamente, existem numa relação de proximidade, não utilizo aspas.

Essa questão sugere, então, que ambas as formas são importantes, já que os metadados de assunto não devem ser nivelados pelos que conhecem bem a matéria, mas por todos, inclusive os que não têm esse conhecimento. Portanto, é interessante ter metadados de assunto que apresentem possibilidades para possíveis subdivisões ou sub-metadados dependendo da situação, no caso de decomposição de assuntos compostos em partes isoladas. Contudo, esse estudo não se propôs a adentrar em subdivisões.

5 CONSIDERAÇÕES FINAIS

Ao longo do estudo, foram elencados alguns aspectos a serem considerados na identificação de metadados de assunto, no que refere à legislação federal brasileira disponível no Legin Web, via Portal da Câmara dos Deputados. Alguns desses aspectos foram extraídos e adaptados das premissas e metodologia para criação de metadados para a disseminação de informações geográficas de Almeida (1999, p. 29), a qual pode ser aplicada, em princípio, a qualquer conjunto de metadados.

A esses aspectos, que constituem os itens 1 a 10, foram acrescentados outros vindos da literatura e dos dados coletados na pesquisa. Entre os aspectos que devem ser observados estão:

- 1) Os dados devem ser priorizados. Essa é uma das premissas do referido autor e indica que os dados ou metadados, que devem ser documentados primeiro, são aqueles dos quais dependem outros metadados/dados, ou representam um alto investimento em termos de esforço e custo. Nesse caso, o custo para resolver problemas por falta de metadados também deve ser considerado;
- 2) A granularidade dos metadados deve estar de acordo com o valor dos seus dados dentro da organização;
- 3) Pesquisa de padrões de metadados já existentes dentro e fora da organização, assim como iniciativas no mesmo sentido e correlatas;
- 4) A decisão de se estender ou não o padrão de metadados escolhido deve ser feita por um grupo multidisciplinar, que represente o conhecimento da organização. Se possível, esse é um processo que também deve ser compartilhado com a organização como um todo, isso evitará lacunas nos sistemas futuros;
- 5) A extensão do padrão terá implicações como: falta de um padrão internacional de pesquisa a bases de dados distribuídas;
- 6) Bases de dados tanto como de metadados podem ser usadas como insumo para a criação de um padrão de metadados.
- 7) A cultura de preenchimento de dados existente, se a inserção nos sistemas é realizada pelos responsáveis pelos metadados ou pelos produtores dos dados. Cada hipótese deve ser checada ao ponto de verificar e testar

vantagens e desvantagens. Mesmo porque, hoje, existem opções, como *softwares* que auxiliam o preenchimento de metadados.

8) Inventário das principais funcionalidades atribuídas aos metadados para ampliar o escopo de soluções que eles propiciam.

9) Desenvolvimento de procedimentos claros e sucintos de implantação e alimentação dos metadados. E estipulação de metas e prazos a cumprir, para que sejam de conhecimento de todos que usam a base, independentemente de sua atividade. Isso, além de contribuir para a qualidade dos metadados, gera empatia por parte da comunidade usuária que se sente, assim, parte do processo.

10) Clareza na identificação, para os usuários, de metadados não obrigatórios, ou que ainda não estejam devidamente preenchidos, em função do impacto direto da falta dessa informação nos resultados de pesquisas.

Os aspectos listados acima têm caráter genérico ou servem para a criação de qualquer tipo de metadado, todavia também existe um aspecto relacionado especificamente a metadados de assunto para a legislação federal brasileira.

Esse aspecto consiste na verificação dos tipos de atos e suas principais funções e conteúdos, conforme o ordenamento jurídico brasileiro e a literatura. Em relação ao ordenamento jurídico, é preciso dizer que ele muda com uma nova Constituição, ou surgem novas espécies normativas, ou mesmo mudam-se as propriedades das existentes até aquele momento. E, devido a isso, é preciso prever características passadas das espécies normativas que, talvez, no ordenamento jurídico atual não sejam tão evidentes ou mesmo existam, com o objetivo de criar metadados de assunto condizentes com as variadas situações.

Ainda em relação aos metadados de assunto, com base na decomposição de estrutura e conteúdo da legislação federal brasileira, pode-se dizer, em resumo, que os dados levantados confirmaram a necessidade da maior granularidade dos metadados de assunto. O motivo é garantir resultados de pesquisa mais precisos, tanto no caso de pesquisas genéricas, como bem específicas.

Barité e Guimarães (1999, p. 49, tradução nossa) reforçam a constatação do estudo, quando discorrem sobre o tratamento temático de

documentos jurídicos (entre eles a legislação), com argumentos fundamentados no estudo de Julius Kaiser, bibliotecário americano que publicou, em 1911, obra intitulada “Systematic Indexing”. Tal obra partiu dos assuntos compostos para criar um sistema que sugeria que todos os assuntos podiam ser divididos em duas categorias, Concretos (Nomes) e Processos (Verbos). De acordo com os autores, Kaiser previu seis regras para a formação de enunciados de indexação:

a) Selecione o que é realmente importante para seu objetivo sem considerar forma ou extensão; b) concentre-se na informação relativamente específica; c) trate cada item independentemente; d) não altere o nome dos concretos; e) evite inversões, proposições e plural sempre que possível; f) teste a exatidão de cada enunciado [de indexação] do ponto de vista tanto do indexador como do usuário. Foskett (1986, p. 42) ressalta pontos que considera avançados na concepção de Kaiser: a) a necessidade de referências do assunto específico para o geral [...]; b) a possibilidade de decompor termos concretos no binômio Concreto/Processo; c) a necessidade da representação de ideias complexas em palavras simples.

Dessa forma, a decomposição de conteúdo está fundamentada tanto no conteúdo em si, como nos pontos de vista do indexador e do usuário, lembrando que o indexador também não deixa de ser um usuário. A decomposição de conteúdo, aliada ao princípio da decomposição da estrutura de atos normativos, mais as concepções extraídas de padrões de metadados já bastante reconhecidos e utilizados, balizaram as recomendações dos possíveis metadados de assunto.

Dessa forma, foram identificados e adaptados metadados de assunto, conforme a TAB. 6, apresentada na sequência, para a legislação federal brasileira, com algumas regras/controles atribuídos a cada metadado, respectivamente. Cabe observar que apesar do Legin Web não ter como base o XML, os nomes atribuídos aos metadados de assunto se basearam no conceito de “elemento”, utilizado por linguagens como o XML, mas isso é meramente ilustrativo, apenas para passar a ideia do metadado de assunto. Eles deverão ser revistos pela equipe responsável pelo desenvolvimento do Legin Web. Esse raciocínio também se aplica a outros exemplos usados na conclusão.

TABELA 6
Metadados de assunto para a legislação federal brasileira
com base em estudo do Legin Web

Metadados de assunto para a legislação federal brasileira (Origem Legin)	Regras/Controles
legin.Subject.Topico	Repetitivo/Opcional
legin.Subject.Titulo.uniforme	Repetitivo/Opcional
legin.Subject.Entidade	Repetitivo/Opcional
legin.Subject.Geografico	Repetitivo/Opcional
legin.Subject.Agrupador	Repetitivo/Opcional
legin.Subject.Pessoal	Repetitivo/Opcional
legin.Subject.Cronologico (tempo como assunto, não como parte da epígrafe do ato normativo)	Repetitivo/Opcional
legin.Subject.Evento	Repetitivo/Opcional
legin.Subject.Genero.forma (usado para identificação de partes do ato normativo)	Repetitivo/Opcional
legin.Subject.Ocupacao.atividade	Repetitivo/Opcional
legin.Subject.Acao	Repetitivo/Opcional
legin.Subject.Termo.livre	Repetitivo/Opcional
legin.Subject.Termo.candidato	Repetitivo/Opcional
legin.Subject.Relation.Texto.atualizado	Repetitivo/Opcional
legin.Subject.Relation.Governo (Vocabulário Controlado do Governo Eletrônico - VCGE)	Repetitivo/Obrigatório/Subdividido
sileg.Subject.Relation.Proposicao	Repetitivo/Opcional

Fonte: Autor.

É preciso dizer que esses seriam metadados de assunto elementares, mas foram selecionados devido a serem utilizados por padrões bastante conhecidos e/ou porque foram muito representativos na análise da pesquisa.

Apesar de o *default* de metadados de assunto geralmente ser opcional, o metadado que usará o VCGE e seus níveis, sendo que tais níveis constituem em subdivisões do metadado `legin.Subject.Relation.Governo` (Nível 1 até o nível necessário), é obrigatório. Excepcionalmente, trataremos de subdivisão aqui, pois como dito anteriormente esse não é o foco desse estudo.

O referido metadado foi classificado como obrigatório para criar um vínculo tecnológico de organização da informação necessário entre instituições governamentais e públicas. Outra razão é que esse metadado de assunto está voltado para o desenvolvimento da navegação em *sites* governamentais, ou estreitamente relacionados, como é o caso da Câmara dos Deputados, como Poder Legislativo.

Abaixo, está um exemplo de preenchimento do metadado de assunto `legin.Subject.Relation.Governo`, com suas subdivisões de nível. Também é ideal que exista um elemento indicando o esquema ou vocabulário utilizado, o VCGE. É importante lembrar que o Padrão de Metadados do Governo Eletrônico (e-PMG) do Brasil usa o VCGE para o preenchimento do metadado *Assunto.categoria*. Os níveis como subdivisões possíveis, sendo que apenas o “Nível.1” seria obrigatório, devem ser utilizados conforme necessidade do indexador. O ideal é que os campos para a inserção de dados de outros níveis pudessem ser gerados na planilha de entrada de dados, conforme o indexador os habilitasse.

```
<legin.Subject.Relation.Governo>
    <Nível.1> Educação </Nível.1>
    <Nível.2> Educação básica </Nível.2>
    <Nível.3> Educação infantil</Nível.3>
</legin.Subject.Relation.Governo>
```

No relacionado à atribuição de obrigatoriedade a outros metadados de assunto, isso é praticamente impossível. Pois, em alguns casos não seria necessário nem mesmo o assunto tópico na indexação, só um assunto “entidade coletiva” (`legin.Subject.Entidade`) poderia ser o principal, acompanhado de algum assunto modificador, por exemplo, e não existir nenhum tópico.

O último metadado analisado teria origem no próprio Sileg, nos metadados de assunto da proposição. Logo, um conjunto de metadados de assunto de origem externa ao Legin, em uma perspectiva ideal.

Algumas observações ainda devem ser feitas no que no trata da criação de metadados de assunto para o Legin. Essa base comporta ambas a legislação federal brasileira e a interna, e mesmo em função de no futuro poderem ser incluídos atos normativos estrangeiros, é necessário que sejam feitos estudos voltados especificamente para tais tipos de atos normativos.

No que refere à decomposição da estrutura de atos normativos para a geração de metadados de assunto, o que se mostrou pertinente é criar dois mecanismos diferentes para atribuição de metadado de assunto conforme estrutura. O primeiro seria usar o próprio metadado de assunto `legin.Subject.Genero.forma` como subdivisão de outro metadado de assunto, e usá-lo para indicar a parte onde está o assunto indexado.

Outra maneira seria quando forem criadas as marcações XML de atos normativos, conforme proposta do próprio Grupo LexML, na forma dos esquemas rígido (atos normativos cf. Lei complementar nº 95, de 1998) e flexível (atos que fogem ao proposto pela referida Lei), incluir a possibilidade de relacionar cada marcação de parte do documento (metadado descritivo e/ou estrutural) ao metadado de assunto respectivo.

Assim, criar um conjunto de metadados recombinantes, onde apareceriam tanto as partes do documento, como seus respectivos assuntos. Para conseguir isso seria necessário usar o metadado de assunto relativo à forma. Essa concepção, de conjuntos de metadados recombinantes é apresentada por Zeng (2010, p. 188).

O emprego de um terceiro esquema como fundação para agregar características de descrição de metadados e de registros (ou documentos) de marcação aumenta consideravelmente a consistência, e, portanto, garante também a interoperabilidade. [...] Somando-se aos resultados da combinação, essa abordagem permite, também, integrar descrições geradas por máquinas e por pessoas. As descrições existentes podem ser reusadas para outros projetos apropriados. Simultaneamente, a qualidade dos metadados pode ser melhorada por meio de metadados recombinantes. De modo geral, os registros integrados podem ser gerados para melhorar o acesso e o compartilhamento. Essa abordagem requererá um esforço adicional para planejar, testar e organizar o fluxo de trabalho e, portanto, é um processo mais complicado e envolverá mais etapas [...].

Por fim, a granularidade dos metadados de assunto ainda pode ser bastante aumentada se olharmos para o número, por exemplo, de assuntos tópicos gerados pelos 204 atos normativos analisados, somando 1798. Todavia, isso exigiria estudos complementares que tratassem da decomposição inclusive desses metadados de assunto em outros. E isso também teria implicações em futuros intercâmbios de metadados.

REFERÊNCIAS

ALMEIDA, L. F. B. de (autor); SILVA, J. X. (orient.). **A metodologia de disseminação de informação geográfica e os metadados**, 1999. Tese (doutorado) - Universidade Federal do Rio de Janeiro, Centro de Ciências Matemáticas e da Natureza, Instituto de Geociências, Rio de Janeiro, RJ. Disponível em:

<http://www.cprm.gov.br/publique/media/dou_fernandobarbosa.pdf>. Acesso em: 04 mar. 2012.

ALVES, R. C. V.; SIMIONATO, A. C.; SANTOS, P. L. V. A. da C. Aspectos de granularidade na representação da informação no universo bibliográfico. In: ENCONTRO NACIONAL DE CATALOGADORES (ENACAT), 1., ENCONTRO DE ESTUDOS E PESQUISAS EM CATALOGAÇÃO (EEPC) , 3., 2012, Rio de Janeiro. **Anais...** Rio de Janeiro, 2012. p. 1-10. Disponível em:

<<http://pt.scribd.com/doc/109274547/Aspectos-de-granularidade-na-representacao-da-informacao-no-universo-bibliografico>>. Acesso em: 29 nov. 2012.

ANDRETA, C. A. Legislação como assunto: uma proposta de extensão para a Classificação Decimal de Direito. In: PASSOS, E. (Org.). **Informação jurídica: teoria e prática**. Brasília: Thesaurus, 2004. p. 79-124.

AUSTIN, D.; DALE, P. **Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngües**. Brasília: IBICT/Senai, 1993. 86p. (Unesco, PGI/81/WS/15). Tradução de Bianca Amaro de Melo e revisão de Lígia Maria Café de Miranda. Este documento foi uma base para a ISO 2788.

BALDIN, N.; MUNHOZ, E. M. B. Snowball (bola de neve): uma técnica metodológica para pesquisa em educação ambiental comunitária. In: CONGRESSO NACIONAL DE EDUCAÇÃO (EDUCERE), 10., SEMINÁRIO INTERNACIONAL DE REPRESENTAÇÕES SOCIAIS, SUBJETIVIDADE E EDUCAÇÃO (SIRSSE) , 1., 2011, Curitiba. **Anais...** Curitiba: PUC, 2011. p. 329-341. Disponível em:

<http://educere.bruc.com.br/CD2011/trabalhos_4.html>. Acesso em: 24 out. 2012.

BARITÉ, M.; GUIMARÃES, J. A. C. **Guia metodológica para el acceso, el análisis y la organización temática de documentos jurídicos**. Montevideo: Universidad de La República, Comisión Central de Educación Permanente, 1999.

BRASIL. Comitê Executivo de Governo Eletrônico. Grupo de Trabalho Organização e Intercâmbio de Informações. Subgrupo Padrão de Metadados do Governo Eletrônico. **Padrão de metadados do governo eletrônico: e-PMG**. Versão 1.0. Brasília: Comitê Executivo de Governo Eletrônico, 2010. Disponível em:

<<http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padroes-de-interoperabilidade/padrao-de-metadados-do-governo-eletronico-e-pmg>>. Acesso em: 24 jul. 2012. Coordenador do Subgrupo Padrão de Metadados do Governo Eletrônico: Eloi Juniti Yamaoka (SERPRO).

BRASIL. Lei Complementar nº 95, de 26 de Fevereiro de 1998. **Diário Oficial [da] República Federativa do Brasil**, Poder Executivo, Brasília, DF, 27 fev. 1998. Seção 1, p. 1. Disponível em: <<http://www2.camara.gov.br/legin/fed/leicom/1998/leicomplementar-95-26-fevereiro-1998-363948-norma-pl.html>>. Acesso em: 18 fev. 2012.

CAMPOS, L. F. de B. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 12, n. 23, p. 16-46, jan./jun. 2007. Disponível em: <<http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2007v12n23p16>>. Acesso em: 30 jul. 2012.

CARNEIRO, A. C. de S.; SANTOS, L. C. A. dos; NÓBREGA NETTO, M. G. da. **Curso de Regimento Interno**. Brasília: Câmara dos Deputados, Edições Câmara, 2011. 446 p. Acesso em: <<http://bd.camara.gov.br/bd/handle/bdcamara/7586>>. Acesso em: 16 out. 2012.

CARVALHO, K. G. **Técnica legislativa**. 5. ed. rev., atual. e ampl. Belo Horizonte: Del Rey, 2010. ISBN 9788538401131.

CASTRO, F. F. de.; SANTOS, P. L. V. A. da C. Uso das tecnologias na Representação descritiva: o padrão de descrição bibliográfica semântica MarcOnt *Initiative* nos ambientes informacionais digitais. **Ciência da Informação**, Brasília, v. 38, n. 1, p. 74-85, jan./abr. 2009. Disponível em: <<http://dx.doi.org/10.1590/S0100-19652009000100005>>. Acesso em: 30 jul. 2012.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de Documentos Eletrônicos. **e-ARQ** : modelo de requisitos para sistemas informatizados de gestão arquivística de documentos: Brasil. Versão 1.1. Rio de Janeiro: Conarq, 2011. 136 p.

CUNHA, M. B. da; CAVALCANTI, C. R. de O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília, DF: Brique de Lemos, 2008. 451 p.

FUNDAÇÃO BIBLIOTECA NACIONAL. **Catálogo de terminologia de assuntos**. 2012. Disponível em: <http://www.bn.br/portal/?nu_pagina=1>. Acesso em: 30 out. 2012.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2007. 175 p.

GUIMARÃES, J. A. C. Elaboração de ementas de atos normativos: elementos de análise documentária com subsídio teórico à técnica legislativa. In: PASSOS, E. (Org.). **Informação jurídica**: teoria e prática. Brasília: Thesaurus, 2004. p. 9-32.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS (IFLA). **Functional Requirements for Subject Authority Data (FRSAD)**: a conceptual model: Functional Requirements for Subject Authority Data (FRSAD). 2010. (Editores: Marcia Lei Zeng, Maja Zumer, Athena Salaba; aprovado pelo Standing Committee of the IFLA Section on Classification and Indexing). Disponível em: <<http://www.ifla.org/node/1297>>. Acesso em: 01 ago. 2012.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). **Documentation – Guidelines for the establishment and development of monolingual thesauri (ISO 2788)**. 2nd ed. 15 nov. 1986. 32 p.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. rev. atual. Brasília, DF: Briquet de Lemos, 2004.

LEXML BRASIL. **LexML**: rede de informação legislativa e jurídica. Versão 1.0 (RC1). Brasília, 2008. Disponível em: <<http://www.lexml.gov.br/>>. Acesso em: 19 out. 2012.

LIBRARY OF CONGRESS. Cataloging Policy and Support Office (prepared by the). **Library of Congress Subject Headings Pre- vs. Post-Coordination and Related Issues**: report for Beacher Wiggins, Director, Acquisitions & Bibliographic Access Directorate, Library Services, Library of Congress. 2007. Disponível em: <http://www.loc.gov/catdir/cpso/pre_vs_post.pdf>. Acesso em: 20 set. 2012. 49 p.

LIBRARY OF CONGRESS. **MARC 21**: harmonized USMARC and CAN/MARC. out. 1998. Disponível em: <<http://www.loc.gov/marc/annmarc21.html>>. Acesso em: 01 ago. 2012. Announcement.

LIBRARY OF CONGRESS. **MARC Standards**. 2012. Disponível em: <<http://www.loc.gov/marc/>>. Acesso em: 24 jul. 2012.

LIMA, J. A. de O. A importância dos relacionamentos entre unidades de informação no tratamento da informação jurídica. In: MIRANDA, A.; SIMEÃO. E. (Org.). **Alfabetização digital e acesso ao conhecimento**. Brasília: UNB, Departamento de Ciência da Informação e Documentação, 2006. p. 177-198.

MARCONDES, C. H. Metadados: descrição e recuperação de informações na Web. In: MARCONDES, C. H (Org.) et al. **Bibliotecas digitais: saberes e práticas**. Salvador: UFBA, 2006. p. 95-111.

MEGREY, B. A. (Ed.); et. al. **Metadata Federation of PICES Member Countries**. Sidney, CA: Secretariat/Publisher North Pacific Marine Science Organization (PICES), 2007. Disponível em: <<http://www.tinro.ru/pices/tcode/fr.pdf>>. Acesso em: 15 abr. 2012. (PICES Technical Report No. 1, 2007)

MORESI, E. A. D.; RAMOS, R. G. C.; PRADO, H. A. Mapeamento de informações organizacionais: um estudo na Embrapa. **Transinformação**, v. 22, n. 2, p. 101-110, mai./ago., 2010. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=000009763&dd1=1bfdc>>. Acesso em: 14 nov. 2012.

NATIONAL INFORMATION STANDARDS ORGANIZATION (NISO). **Understanding metadata**. Bethesda, 2004. Disponível em: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. Acesso em: 29 jun. 2012.

OLIVER, C. **Introdução à RDA**: um guia básico. Brasília, DF: Briquet de Lemos, 2011.

ONLINE LIBRARY COMPUTER CENTER (OCLC). **Análise do cenário da OCLC em 2003**: reconhecimento de padrões, resumo executivo do reporte The 2003 environmental scan: pattern recognition. Dublin: Alane Wilso, c2004. Principais colaboradores: Cathy De Rosa, Lorcan Dempsey, Alane Wilson.

ONLINE LIBRARY COMPUTER CENTER (OCLC). **Bibliographic Formats and Standards**. 4th ed. Dublin, 2012. Disponível em: <<http://www.oclc.org/bibformats/en/default.shtml>>. Acesso em: 24 jul. 2012. Inclui Formato MARC (Machine-Readable Cataloging).

PEREIRA, A. A. de C. Representação documentária de informação legislativa. In: PASSOS, E. (org.). **Informação jurídica**: teoria e prática. Brasília: Thesaurus, 2004. P. 153-172.

PINTO, A. R. F. S. **Princípios de design de interfaces aplicados à busca**. Rio de Janeiro, 2010. Trabalho de conclusão de curso – Universidade Federal do Rio de Janeiro, Curso de Biblioteconomia e Gestão de Unidades de Informação, Rio de Janeiro, RJ. Disponível em: <<http://www.slideshare.net/arfserra/principios-do-design-de-interfaces-aplicados-a-busca>>. Acesso em: 01 abr. 2012.

PROJETO italiano simplifica a justiça. **Notícias do 1º Seminário de Gestão de Informação Jurídica em Espaços Digitais, realizado pelo Supremo Tribunal Federal, de 12 a 14 de fevereiro de 2007**. 2007. Disponível em: <<http://www.stf.jus.br/sijed/noticia.asp?seqNoticia=222766>>. Acesso em: 18 jan. 2013.

REIS, G. **Arquitetura de informação**. 2007. Disponível em: <<http://www.quilhermo.com/apresentacoes/ArquiteturaInformacao-DiaDoBibliotecario-GuilhermoReis.pdf>>. Acesso em: 01 abr. 2012.

SANTOS, L. C. A. dos; NÓBREGA NETTO, M. G. da; CARNEIRO, A. C. de SÁ. **Curso de Regimento Interno da Câmara dos Deputados**. 2. ed. rev., atual. e ampl. Brasília: Vestcon, 2007.

SOARES, F. de M. Legística e desenvolvimento: a qualidade da lei no quadro da otimização de uma melhor legislação. **Revista da Faculdade de Direito da UFMG**, Belo Horizonte, n. 50, p. 124-142, jan./jul. 2007. Disponível em: <<http://www.direito.ufmg.br/revista/articles/31.pdf>>. Acesso em: 15 fev. 2012.

VERGARA, S. C. **Métodos de pesquisa em administração**. 4. ed. São Paulo: Atlas, 2010. 277 p.

VOGEL, M. J. M. A evolução do conceito de linguagem documentária: as linhas francesa e brasileira. In: ENANCIB – ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 7., 2007, Salvador. **Anais...** Salvador: UFBA, 2007. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--146.pdf>>. Acesso em: 15 fev. 2012.

ZENG, M. L. Linguagens de marcação específicas por domínio e metadados descritivos: funções para a descoberta de recursos científicos. **Encontros Bibli:** revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, edição especial, p. 177-191, jul./dez. 2010. Disponível em: <www.sumarios.org/sites/default/files/pdfs/55363_6389.PDF>. Acesso em: 20 abr. 2012. (Tradução de Gleisy R. B. Fachin.)

APÊNDICE A

Questionário e respostas

QUADRO 2

Respostas dos 20 (vinte) questionários sobre buscas por assunto na legislação federal brasileira disponível no Portal da Câmara dos Deputados

1) a) Você poderia dizer em qual(is) situação(ões) no trabalho precisa ou precisou consultar a legislação federal brasileira disponível no Portal da Câmara dos Deputados? **Essa questão permite a escolha de mais de uma opção. Assinale a(s) opção(ões) desejada(s).**

Opções	<p>1 - Levantar quais normas federais tratam de determinado tema para prestar informação ao cidadão ou ao público interno da Câmara dos Deputados.</p> <p>2 - Elaborar relatório ou parecer a respeito da legislação sobre o assunto pesquisado.</p> <p>3 - Assessorar na elaboração ou elaborar proposição de lei.</p> <p>4 - Acompanhar a evolução da discussão de proposições legislativas.</p> <p>5 - Outras.</p>
Respostas	<p>1 – 45% (9) – Respondentes 1, 4, 5, 6, 9, 10, 14, 17 e 18</p> <p>2 – 40% (8) – Respondentes 1, 3, 4, 5, 6, 10, 13 e 15</p> <p>3 – 50% (10) – Respondentes 1, 3, 4, 5, 6, 8, 10, 12, 13 e 15</p> <p>4 – 40% (8) – Respondentes 1, 3, 4, 5, 6, 10, 12 e 17</p> <p>5 – 55% (11) – Respondentes 2, 4, 7, 9, 11, 13, 14, 16, 17, 19 e 20</p>

1) b) **No caso da resposta à questão anterior ser “Outras”,** você poderia citar em qual(is) situação(ões) no seu trabalho precisa ou precisou consultar a legislação federal?

Respostas	<p>Respondente 2 Faço resumos de PL's para a pauta da Comissão de Trabalho, Administração e Serviço Público. Pareceres e relatórios, e pesquisa legislativa, geralmente utilizando os operadores disponíveis no sistema.</p> <p>Respondente 4 Elaborar estudos ou Notas Técnicas sobre determinado tema. Porém minhas pesquisas no portal se resumem a Medidas Provisórias, Projetos de Lei e atos internos. Quando necessito consultar a legislação em vigor busco no site do Planalto, pois é mais atualizado.</p> <p>Respondente 7 Para anexar o texto em vigor das normas citadas em proposição apresentadas na Câmara dos Deputados. Obs: Conhecida como Legislação Citada.</p> <p>Respondente 9 Para elaborar artigos ou livros técnicos.</p> <p>Respondente 11 Em geral pesquisamos pelo nº da norma, do artigo, do parágrafo ou do inciso. No trabalho de indexação de proposições, comparamos o texto vigente da legislação citada com o texto proposto. Eventualmente, para efeito de análise/avaliação/revisão do nosso trabalho, comparamos os termos utilizados na indexação da legislação federal com os termos utilizados na indexação das proposições.</p> <p>Respondente 13 Conhecer a legislação em vigor sobre determinado assunto.</p> <p>Respondente 14 Consulto a legislação federal para atualizar as normas jurídicas, com as alterações publicadas no Diário Oficial da União.</p> <p>Respondente 16 Acessar texto atualizado das normas; Identificar a proposição originária.</p>
------------------	---

Continuação **Respondente 17**
Compilar legislação para publicação temática.

Respondente 19
Trabalho de revisão/correção dos dispositivos da Constituição de 1988 regulamentados ou passíveis de regulamentação.

Respondente 20
1 - Vertente indexação: Quando o Deputado se refere a alguma lei em seu discurso e se torna necessária a confirmação dos dados apresentados para a indexação do discurso; 2 - Vertente pesquisa. Quando são solicitadas votações que deram origem a uma determinada lei.

2) a) Você se lembra das três últimas pesquisas por assunto que fez na legislação federal brasileira disponível no Portal da Câmara? **Assinale a opção desejada.**

Opções e Respostas **Não – 45% (9) – Respondentes 4, 7, 8, 9, 11, 14, 15, 16 e 19**
Sim – 55% (11) – Respondentes 1, 2, 3, 5, 6, 10, 12, 13, 17, 18 e 20

2) b) **No caso da resposta à questão anterior ser “Sim”,** quais foram os assuntos pesquisados na legislação federal brasileira?

Respondente 1
Legislação trabalhista

Respondente 2
Idoso, Teorias do Direito Penal, crimes virtuais, anteprojeto da nova Lei de Direitos Autorais.

Respondente 3
Pesquisa para levantar se já existiu precedente de, Projeto de Lei, aprovado pela Câmara dos Deputados, que anistia condenado pela justiça estadual/federal Brasileira. Pesquisa para perceber a tramitação/evolução dos projetos de Lei dos cinco mandatos do Deputado. Pesquisa pra apurar pareceres em comissões.

Respostas **Respondente 4**
Lembro apenas do último, Projetos de Lei que revoguem a dedução do pagamento de juros sobre capital próprio (Direito Tributário). Geralmente minhas pesquisas não são feitas por assunto, mas pelo número da lei ou proposição.

Respondente 5
Concessões, Energia elétrica, Energia solar

Respondente 6
Lei Maria da Penha, Lei de Tóxicos e Convenção Anticorrupção.

Respondente 10
Mensagens (MSC); Projetos de Decreto Legislativo (PDC); Decretos legislativos e Decretos do Presidente da República.

Respondente 12
Legislação referente à Comissão de Ética Pública, veto à isenção de impostos sobre produtos de cesta básica e Medidas Provisórias.

Respondente 13
Terceirização no serviço público; Código florestal; Equipamentos obrigatórios para as motocicletas (antena corta fio).

Respondente 17
Licença-gestante.

Respondente 18
Programa Bolsa Família; Corregedoria da CD e Procedimentos investigatórios; Legislação sobre informações sigilosas.

Respondente 20

Continuação 1- Lei Maria da Penha; 2 - Lei de regulamentação da profissão de motorista; 3 - Código florestal.

2) c) Encontrou dificuldades para realizar a pesquisa por assunto na legislação federal brasileira? **Assinale a opção desejada.**

Opções e Respostas Não – 55% (11) – Respondentes 1, 2, 3, 4, 5, 6, 7, 8, 9, 18 e 19
Sim – 30% (6) – Respondentes 10, 12, 13, 15, 17 e 20

Observação: Não responderam a questão 2c) os respondentes 11, 14 e 16.

2) d) **No caso da resposta à questão anterior ser “Sim”, qual(is) foram as dificuldade(s) encontrada(s) na pesquisa por assunto na legislação federal brasileira?**

Respostas

Respondente 10
Passei esta manhã tentando encontrar um acordo de cooperação judiciária gratuita entre Brasil e Japão e encontrei apenas um sobre cooperação judiciária muito antigo. Havia um ícone referente a atos internacionais na página da Comissão de Relações Exteriores que era bastante útil e mais rápido que o link correspondente do Executivo, mas não está mais no ar.

Respondente 12
Demora na atualização, não mostra legislação correlata e vinculada (por exemplo, Decreto não mostra lei original).

Respondente 13
O conceito "antena corta fio" não é uniforme. Nos projetos, na legislação e nas normas do CONTRAN encontramos, "antena corta fio", "antena apara pipa", "dispositivo corta fio". Em geral as pesquisas apresentam uma grande revocação.

Respondente 15
O portal de legislação da CD é interessante quando faço uma busca exploratória sobre legislação temática. No entanto, a hierarquização não é muito boa. Quando ponho o critério relevância, para apresentação dos dados, o retorno pode não corresponder à realidade (da importância das normas). Daí que a pesquisa fica mais trabalhosa. No dia a dia, para achar legislação que não conheço, uso mais o Google com mais frequência. Para ir direto à norma que já conheço, vou ao site do Planalto.

Respondente 17
Ausência da legislação infraconstitucional, portarias, normas técnicas, etc.

Respondente 20
Excesso de dados não pertinentes.

3) a) Você já fez alguma pesquisa que precisou levantar toda a legislação federal brasileira que trate de um determinado assunto? **Assinale a opção desejada.**

Opções e Respostas Não – 25% (5) – Respondentes 7, 9, 11, 16 e 19
Sim – 75% (15) – Respondentes 1, 2, 3, 4, 5, 6, 8, 10, 12, 13, 14, 15, 17, 18, 20.

3) b) Sobre qual(is) assunto(s) pretendia fazer o levantamento de normas federais?

Respostas

Respondente 1
Matéria trabalhista, para subsidia a elaboração do parecer ao PL-1987/07, que consolida essa legislação.

Respondente 2
Datas comemorativas, desde o Império.

Respondente 3
Sim, o Deputado pretendia punir severamente os motoristas que provocarem mortes na condução de veículos automotores sobre o efeito de drogas ou álcool, então foi necessário estudar toda a legislação Brasileira.

Continuação	Respondente 4 Diversas matérias no âmbito do Direito Tributário. Geralmente a forma de tributação de determinado setor ou atividade.
	Respondente 5 Energia elétrica (para elaborar Projeto de Lei de consolidação).
	Respondente 6 Legislação anticorrupção.
	Respondente 7 Diversos, mais geralmente já temos os números dos atos normativos pesquisados.
	Respondente 8 Não lembro.
	Respondente 9 Não costumo pesquisar leis por assunto. Em geral, pesquisa uma lei cujo número e ano já sei.
	Respondente 10 Fiz um levantamento completo das mensagens presidenciais que encaminharam atos internacionais ao Congresso Nacional entre 1/1/1988 e 31/12/2011, que, no momento, estou atualizando com os dados referentes a 2012.
	Respondente 12 Isenção IPI
	Respondente 13 Matéria de tributária.
	Respondente 14 Sobre processo administrativo disciplinar de servidor público e sobre tomada de contas especial.
Respondente 15 Sobre diversos assuntos na área de meio ambiente, na qual trabalho.	
Respondente 17 Legislação indígena	
Respondente 18 Legislação eleitoral	
Respondente 20 Sobre a mulher.	
Observação: Os respondentes 7 e 9 responderam a questão 3b, apesar de selecionarem “Não” na questão 3a.	

3) c) Encontrou dificuldades para realizar esse levantamento? **Assinale a opção desejada.**

Opções e Respostas	Não – 25% (5) – Respondentes 3, 8, 9, 14 e 15
	Sim – 60% (12) – Respondentes 1, 2, 4, 5, 6, 7, 10, 12, 13, 17, 18 e 20
	Observação: Não responderam a questão 3c) os respondentes 11, 16 e 19.

3) d) **No caso da resposta à questão anterior ser “Sim”, qual(is) foram as dificuldade(s) encontradas no levantamento?**

Respostas	Respondente 1 Realizei a pesquisa simultaneamente nos sites da Câmara, do Senado e da Presidência da República, pois muitas vezes as informações estavam incompletas.
	Respondente 2

Continuação As que já estavam revogadas, derogadas e em vigor.

Respondente 4

A legislação tributária é muito disseminada, incluindo normas que possuem décadas de existência. Muitas dessas normas, apesar de vigentes, não aparecem na pesquisa e outras, apesar de aparecerem, já foram revogadas tacitamente por textos mais recentes.

Respondente 5

As leis do setor de energia elétrica utilizam ambos os termos “energia elétrica” e “eletricidade”. Quando se faz a procura de legislação empregando conjuntamente esses termos a resposta é bem menor do que quando se faz a pesquisa isolada de cada um deles. Portanto, o resultado da pesquisa com termos conjuntos é do tipo “termo 1” E “termo 2”. Eventualmente, seria interessante poder optar pela pesquisa ser do tipo “termo 1” OU “termo 2”.

Respondente 6

Há muito material relativo a Direito Comparado, ou seja, não somente Direito Internacional, mas legislação estrangeira, de outros países (Estados Unidos, reio Unido) ou blocos (União Européia). Assim, o problema com a tradução dos textos legais estrangeiros é sério.

Respondente 7

O Layout da pagina de pesquisa não ajuda.

Respondente 10

Consegui efetuar o levantamento exaustivo apenas quando não coloquei filtro algum em “Mensagem”, fazendo a seleção daquelas referentes a atos internacionais manualmente. Verifiquei que nem sempre os pactos internacionais estão catalogados como “atos internacionais”. Outros, só encontrei por acaso, ao pesquisar algum outro dado, mas, cruzando as ferramentas do sistema foi possível fazer um levantamento exaustivo de tudo que o Executivo encaminhou ao Congresso no período pesquisado.

Respondente 12

Falta de links no texto, por exemplo, lei que altera outra lei não tem link para lei alterada.

Respondente 13

A pesquisa em si não é problema, porque uso não apenas o site da CD, mas também o Google e links oficiais.

Respondente 17

Insegurança da atualização das leis e das informações sobre revogação e alteração.

Respondente 18

Fontes desatualizadas quanto à validade da norma, indexação mal feita ou ausente.

Respondente 20

Gostaria de filtrar por aspectos. Porém só são oferecidos filtros por ano, pessoas, etc. Você perde o foco do total da legislação.

4) a) Você já procurou algum assunto na legislação federal brasileira disponível no Portal da Câmara dos Deputados e não encontrou nada a respeito? **Assinale a opção desejada.**

Opções e Respostas

Não – 65% (13) – Respondentes 1, 3, 4, 5, 6, 7, 9, 10, 13, 14, 15, 16 e 19
Sim – 30% (6) – Respondentes 2, 8, 12, 17, 18 e 20

Observação: Não respondeu a questão 4a) o respondente 11.

4) b) Qual(is) o(s) assunto(s) não encontrado(s)?

Respostas

Respondente 2

Crimes virtuais. Somente em outras fontes legislativas e pelo próprio Google.

Continuação	<p>Respondente 8 Não lembro.</p> <p>Respondente 12 Não recorde, mas várias vezes tive que recorrer ao Portal da Presidência da República ou do Senado Federal.</p> <p>Respondente 13 Não encontrar é um problema? Pode não existir a legislação procurada. Em minha opinião, a questão relevante é como garantir se a informação/conteúdo procurada não existe ou se simplesmente não foi encontrada.</p> <p>Respondente 17 Dia das Artes</p> <p>Respondente 18 De maneira geral, assuntos que são regulamentados pelos estados ou municípios, ou ainda por norma técnica, por exemplo, construção e manutenção de calçadas públicas.</p> <p>Respondente 20 Não me lembro, porém quando tenho dificuldades pesquisei na legislação do portal da presidência da República ou no Google.</p> <p>Observação: O respondente 13 respondeu a questão 4b, apesar de selecionar “Não” na questão 4a.</p>
--------------------	---

5) a) Você já precisou realizar pesquisas de assunto em partes específicas do texto de uma norma federal? **Assinale a opção desejada.**

Opções e Respostas	<p>Não – 35% (7) – Respondentes 1, 5, 8, 11, 12, 16 e 19</p> <p>Sim – 60% (12) – Respondentes 2, 3, 4, 6, 7, 9, 10, 14, 15, 17, 18 e 20</p> <p>Observação: Não respondeu a questão 5a) o respondente 13.</p>
---------------------------	---

5) b) **No caso da resposta à questão anterior ser “Sim”, por gentileza assinale abaixo em qual(is) parte(s) da norma federal realizou ou necessitou realizar a pesquisa por assunto?**

Opções e Respostas	<p>1 – Ementa – 45% (9) – Respondentes 2, 3, 6, 7, 9, 10, 14, 15 e 18</p> <p>2 – Preâmbulo – 20% (4) – Respondentes 3, 6, 7 e 10</p> <p>3 – Artigo – 55% (11) – Respondentes 2, 3, 4, 6, 7, 9, 10, 15, 17, 18 e 20</p> <p>4 – Parágrafo – 50% (10) – Respondentes 2, 3, 4, 6, 7, 9, 10, 15, 17 e 18</p> <p>5 – Inciso – 50% (10) – Respondentes 2, 3, 4, 6, 7, 9, 10, 15, 17 e 18</p> <p>6 – Alínea – 45% (9) – Respondentes 2, 3, 4, 6, 7, 9, 10, 15 e 18</p> <p>7 – Item – 45% (9) – Respondentes 2, 3, 4, 6, 7, 9, 10, 15 e 18</p> <p>8 - Cláusula de vigência – 30% (6) – Respondentes 4, 6, 7, 10, 14 e 15</p> <p>9 - Cláusula de revogação – 25% (5) – Respondentes 4, 6, 7, 10 e 15</p> <p>10 - Subseções ou Seções – 20% (4) – Respondentes 3, 7, 9 e 10</p> <p>11 – Capítulo – 30% (6) – Respondentes 3, 6, 7, 9, 10 e 15</p> <p>12 – Título – 30% (6) – Respondentes 3, 6, 7, 9, 10 e 15</p> <p>13 – Livro – 25% (5) – Respondentes 3, 6, 7, 9 e 10</p> <p>14 – Parte – 25% (5) – Respondentes 3, 6, 7, 9 e 10</p>
---------------------------	--

Continuação	<p>15 - Agrupamentos em disposições (preliminares, gerais, finais ou transitórias) – 25% (5) – Respondentes 3, 6, 9, 10 e 15</p> <p>16 – Anexo –35% (7) – Respondentes 3, 4, 6, 7, 10, 15 e 18</p> <p>17 – Outras – 10% (2) – Respondentes 4 e 10</p>
5) c) No caso da resposta à questão anterior ser “Outras”, por gentileza cite qual(is) foram a(s) outra(s) parte(s) que desejou consultar ou consultou?	<p>Respondente 4 Exposição de Motivos (Medidas Provisórias) e Justificação (Projetos de Lei).</p> <p>Respondente 10 Na verdade, em todos os itens acima: muitas vezes busca-se um determinado dado sem se saber exatamente o dispositivo em que está e, do resultado da pesquisa, obtêm-se as fontes possíveis.</p> <p>Respondente 12 A maioria das pesquisas que faço é a respeito de legislação sobre determinado assunto ou sanção e promulgação de leis, e a parte específica do texto é irrelevante.</p> <p>Respondente 13 Não entendi o objetivo dessa pergunta. Normalmente faço a pesquisa no inteiro teor da norma, que de certa forma incide sobre todo o conteúdo acima listado.</p> <p>Observação: Os respondentes 12 e 13 fizeram apenas comentários referentes às questões 5a e 5b.</p>
5) d) Encontrou dificuldades para realizar a pesquisa por assunto em partes específicas da legislação federal brasileira? Assinale a opção desejada.	<p>Não – 35% (7) – Respondentes 2, 3, 4, 7, 8, 15 e 19</p> <p>Sim – 35% (7) – Respondentes 6, 9, 10, 13, 14, 17 e 18</p> <p>Observação: Não responderam a questão 5d) os respondentes 1, 5, 11, 12, 16 e 20.</p>
5) e) No caso da resposta à questão anterior ser “Sim”, qual(is) foram as dificuldade(s) encontrada(s) na pesquisa por assunto em partes específicas da legislação federal brasileira?	<p>Respondente 6 Não lembro, no momento. Lembro apenas que não havia a possibilidade de escolher itens e compara-los, por exemplo.</p> <p>Respondente 10 Por vezes, tem-se a ideia de que determinado dispositivo está em um lugar "X" e vai-se encontrá-lo em "y": considerando a vastidão das normas legais, nem sempre se tem em mente o "locus" preciso da norma.</p> <p>Respondente 14 Na realidade, não é que eu tenha encontrado dificuldades para pesquisar a legislação, mas em pesquisas em que o ponto de partida é um trecho de norma ou de dispositivo de norma utilizo o Google como ferramenta para identificar a norma e só depois da identificação consulto a norma na base de dados de legislação da Câmara, que considero mais confiável. Considero o Google um instrumento tão poderoso e de fácil utilização que não tenho utilizado nem tenho me dedicado a aprender ou desenvolver as pesquisas com as ferramentas disponíveis nos bancos de dados de legislação.</p> <p>Respondente 15 O objetivo da pergunta não está claro para mim. Quando faço a pesquisa, entendo que o sistema faz a varredura por um dos dois critérios: ementa ou todo o corpo do texto. O sistema não procura partes isoladas da norma, no corpo do texto. Coloco a palavra-chave e o sistema responde quais leis tratam do assunto. Cabe a mim abrir cada texto e verificar se tem partes que me interessam. Todas as partes do texto de</p>

Continuação uma norma são importantes no conjunto das pesquisas.

Respondente 17

A pesquisa não permite detalhamento específico de item.

Respondente 18

Falta de indexação ou indexação com termos livres.

Observação: Não respondeu a questão 5e) o respondente 9, apesar de ter respondido "Sim" à questão 5d). O respondente 15 comentou as questões 5a, 5b, 5c e 5d.

6) Se, hipoteticamente, pudesse apontar novas formas de filtrar ou refinar o resultado das buscas por assunto na legislação federal brasileira disponível no Portal da Câmara, quais seriam?

Respondente 1

No momento, não posso apontar nenhuma.

Respondente 2

O uso do booleano and ou or.

Respondente 4

Excluir textos que contenham determinada palavra.

Respondente 5

Não tenho sugestões.

Respostas

Respondente 6

Certamente, por todos os itens mencionados na pergunta anterior. Também tentaria fazer um serviço de tradução de textos legais relevantes sobre temas em discussão no Brasil e em outros países.

Respondente 7

Não posso opinar porque uso muito pouco a pesquisa por assunto no Portal da Câmara.

Respondente 8

Não sei.

Respondente 9

A possibilidade de utilizar filtros diversos. Por exemplo, escolher uma ou mais expressão que a norma deve conter e excluir termos ou expressões que não devem constar da norma. Selecionar normas que contenham termo ou expressão somente na ementa. Selecionar normas promulgadas em determinado intervalo temporal.

Respondente 10

Sugeriria que, nos casos dos pactos internacionais celebrados pelo Brasil, quer bilaterais, multilaterais etc., independentemente de seu formato (tratado, convenção, convênio, acordo, protocolo etc., sejam indexados como "ato internacional", que é o nome genérico no qual cabem todos os tipos e feitios de avenças internacionais.

Respondente 12

Link para alterações e legislação citada, informação sobre PL que originou Lei, Lei que decreto está regulamentado e outras correlações.

Respondente 13

Os resultados da busca têm apresentado uma grande revocação. Quando fazemos uma busca por dois termos ou mais, sem " ", parece que o sistema utiliza como default o conectivo "ou". Em minha opinião, quando alguém faz a busca conjugada intuitivamente ele espera como resposta a presença de todos os termos utilizados na busca.

Respondente 14

Creio que as formas atuais de filtro e refino atendem a quase totalidade das necessidades.

Continuação	<p>Respondente 15 Colocaria alguns critérios de pesquisa avançada. Por exemplo: filtrar legislação entre normas em vigor e normas revogadas. Colocaria a opção de retorno para o início da pesquisa.</p> <p>Respondente 16 Não uso busca por assunto.</p> <p>Respondente 17 Gostaria de otimizar os produtos: apresentação de listagens, navegação pelos anos e tipos de legislação (como no site do planalto).</p> <p>Respondente 18 Indexar cada parte do texto legal, mas sobretudo a matéria ali contida, não necessariamente as palavras.</p> <p>Respondente 19 Não tenho contribuições.</p> <p>Respondente 20 Uso de conectores para excluir aspectos em uma busca. Isto e não aquilo, por exemplo. A partir de uma busca inicial vc refinar por aspectos. Mulher -1, 1 e violência doméstica, 1 e profissão, 1 e mercado de trabalho, e assim por diante.</p> <p>Observação: Não responderam a questão 6) os respondentes 3 e 11.</p>
--------------------	---

7) Quando realiza uma pesquisa por assunto, o qual supostamente é um termo composto (ex.: reforma tributária), você usa palavras isoladas na sua expressão de busca (ex.: reforma, tributo, imposto), ou o próprio termo composto (reforma tributária), ou tenta ambas as formas?

	<p>Respondente 1 Depende da necessidade do momento. Se eu souber exatamente a expressão que consta da lei, procuro diretamente pelo termo composto. Se a pesquisa for mais ampla, procurando normas que tratem da matéria, uso palavras isoladas.</p> <p>Respondente 2 Ambas as formas.</p> <p>Respondente 3 Depende, se o objetivo for atingido com o termo composto não procuro por palavras. Normalmente tenho conseguido apenas com termo composto.</p> <p>Respondente 4 Depende da situação, no caso acima colocaria Reforma Tributária porque é um termo consagrado. Em outras situações poderia colocar, por exemplo: imposto de renda, dedução, juros sobre capital próprio.</p>
Respostas	<p>Respondente 5 Uso o próprio termo composto.</p> <p>Respondente 6 Tento ambas as formas.</p> <p>Respondente 7 Não posso opinar porque uso muito pouco a pesquisa por assunto no Portal da Câmara. Mas quanto necessário uso ambas as formas.</p> <p>Respondente 8 Tento ambas as formas.</p> <p>Respondente 9 Tento por expressão ou por termos pertinentes, conforme o assunto. Há termos cuja pesquisa resulta num volume enorme de legislação, sendo necessário ser mais específico quanto ao que se quer, por isso seria necessário melhorar as possibilidades de filtrar o resultado da pesquisa.</p>

Continuação	<p>Respondente 10 Tento ambas as formas: ato internacional, por exemplo. Mas já descobri que, para exaurir todas as mensagens pertinentes recebidas pelo Congresso, tenho de verificar uma a uma de todas as mensagens encaminhadas pela Presidência da República, o que acaba incluindo assuntos que nada têm a ver com o tema da pesquisa, tais como afastamentos da Presidente; vetos; promulgações, concessões de rádio e televisão etc.</p> <p>Respondente 12 Uso todas as formas possíveis.</p> <p>Respondente 13 Reforma tributária, entretanto a aspas só é possível quando se sabe exatamente a expressão de busca. Quando os termos buscados não estão ligados diretamente, existem numa relação de proximidade, não utilizo aspas.</p> <p>Respondente 14 Inicialmente eu utilizo o termo composto. Caso não encontre a norma desejada dessa forma, passo a pesquisar por palavras isoladas.</p> <p>Respondente 15 As duas formas.</p> <p>Respondente 17 Tento ambas as formas.</p> <p>Respondente 18 Ambas as formas.</p> <p>Respondente 19 Tento ambas as formas.</p> <p>Respondente 20 Utilizo o termo composto. Gostaria de usar conectores na pesquisa.</p> <p>Observação: Não responderam a questão 7) os respondentes 11 e 16.</p>
--------------------	---
