

# Organização do Conhecimento e Diversidade Cultural

Organizadores: José Augusto Chaves Guimarães e Vera Dodebei



ISKO-BRASIL



Organizadores

José Augusto Chaves Guimarães  
Vera Dodebei

## Organização do Conhecimento e Diversidade Cultural

Marília, São Paulo, Brasil

Sociedade Brasileira de Organização do Conhecimento  
(ISKO-Brasil)

Fundação para o Desenvolvimento do Ensino, Pesquisa e Extensão  
(FUNDEPE)

2015

O68o      Organização do conhecimento e diversidade cultural [recurso eletrônico] / José Augusto Chaves Guimarães, Vera Dodebei, organizadores. -- Marília: ISKO-Brasil ; FUNDEPE, 2015.  
835 f. ; 30 cm.

**ISBN: 978-85-98176-70-3**

Livro digital

1. Organização do conhecimento. I. Título.

CDD 025.4

**Comissão Científica**

Vera Lúcia D. Louzada de Mattos Dodebei (UNIRIO)

Gercina Ângela Borém de Oliveira Lima (UFMG)

Marisa Bräscher Basílio Medeiros (UFSC)

Maria Aparecida Moura (UFMG)

Fabio Assis Pinho (UFPE)

**Editoração**

Isadora Victorino Evangelista

Gilberto Gomes Cândido

Rafael Aparecido Moron Semidão

Rafael Cacciolari Dalessandro

Suellen Oliveira Milani

**Capa e arte gráfica**

Maíra Fernandes Alencar

© A reprodução deste livro na íntegra ou em parte é permitida, desde que citados os créditos.  
Proibida a venda.

## SUMÁRIO

<b>APRESENTAÇÃO.....</b>	<b>13 - 19</b>
 <b>A DIMENSÃO EPISTEMOLÓGICA DA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO</b>	
<b>Epistemic, ontic, axiologic, and praxic constructs in knowledge organization research.....</b>	<b>21 - 29</b>
<i>Joseph T. Tennis</i>	
<b>Abordagens e paradigmas na organização do conhecimento.....</b>	<b>30 - 36</b>
<i>Leila Cristina Weiss e Marisa Bräscher</i>	
<b>A dimensão histórico-epistemológica da Organização do Conhecimento: contributos da taxonomia Gesneriana, Séc. XVI.....</b>	<b>37 - 46</b>
<i>Andre Vieira de Freitas Araujo e Giulia Crippa</i>	
<b>Configurações terminológicas e interdisciplinares na pesquisa em organização do conhecimento (2010-2014).....</b>	<b>47 - 54</b>
<i>Leilah Santiago Bufrem, Aline Elis Arboit e Juliana Lazzarotto Freitas</i>	
<b>A dimensão conceitual da organização do conhecimento nos congressos da NASKO: análise de conteúdo Bardiana.....</b>	<b>55 - 72</b>
<i>José Augusto Chaves Guimarães, Rodrigo de Sales, André Ynada dos Santos e Daniela Fernanda de Oliveira Matos</i>	
<b>A relação entre organização do conhecimento e ciência da informação na comunidade científica brasileira: uma investigação no âmbito da ISKO-Brasil.....</b>	<b>73 - 84</b>
<i>Rodrigo de Sales</i>	
<b>Poole, o índice e as fraturas: Indexação e publicações seriadas nos Estados Unidos do Oitocentos.....</b>	<b>85 - 93</b>
<i>Gustavo Silva Saldanha e Naira Christofolletti Silveira</i>	
<b>Os processos de representação do conhecimento arquivístico: elementos históricos e conceituais da classificação e descrição.....</b>	<b>94 - 99</b>
<i>Natália Bolfarini Tognoli e Thiago Henrique Bragato Barros</i>	
<b>Ontologias e Knowledge Organization System (KOS): aproximações e diferenças.....</b>	<b>100 - 107</b>
<i>Rogério Aparecido Sá Ramalho</i>	
<b>Estudos de genealogia acadêmica como abordagem para análise de domínio.....</b>	<b>108 - 116</b>
<i>Renata Cristina Gutierrez Castanha e Maria Cláudia Cabrini Grácio</i>	

<b>A dimensão epistemológica da análise documental de conteúdo de obras de ficção na organização do conhecimento.....</b>	<b>117 - 124</b>
<i>Roberta Caroline Vesú Alves e João Batista Ernesto de Moraes</i>	
<b>As linguagens de indexação e a análise de domínio.....</b>	<b>125 - 134</b>
<i>Vera Lucia Ribeiro Guim e Mariângela Spotti Lopes Fujita</i>	
<b>Reader-Interest classification as a method of stock control: the McClellan Legacy.....</b>	<b>135 - 142</b>
<i>Daniel Martínez-Ávila e M. P. Satija</i>	
<b>Terminologia, exaustividade e especificidade: uma relação conceitual.....</b>	<b>143 - 153</b>
<i>Isadora Victorino Evangelista, Walter Moreira e João Batista Ernesto de Moraes</i>	
<b>Aspectos dos metadados de assunto na catalogação de fotografias.....</b>	<b>154 - 162</b>
<i>Ana Carolina Simionato</i>	
<b>Linguagens documentárias e tratamento de recursos audiovisuais: interfaces teóricas e metodológicas.....</b>	<b>163 - 172</b>
<i>Francisnaira Cristina Ravazzi e Walter Moreira</i>	
<b>Organização do conhecimento e design da informação: um estudo de convergência.....</b>	<b>173 - 183</b>
<i>Natalia Nakano, Mariana Cantisani Padua e Maria José Vicentini Jorente)</i>	
<b>Análisis de dominio de las dos primeras ediciones del capítulo ISKO-Brasil.....</b>	<b>184 - 199</b>
<i>Exequiel Fontans e Mario Barité</i>	
<b>Ensino em organização e representação do conhecimento: linguagens e normas.....</b>	<b>200 - 209</b>
<i>Cibele Araújo Marques dos Santos, Eduardo de Abreu de Jesus e João Ricardo de Luca</i>	
<b>A genealogia do conceito de Ciência da Informação no Brasil: uma análise discursiva a partir de periódicos fundacionais na área .....</b>	<b>210 - 218</b>
<i>Larissa de Melo Lima, João Batista Ernesto de Moraes e Daniel Martínez-Ávila</i>	

## A DIMENSÃO APLICADA DA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

<b>A Discourse Analysis of the Semantic Web.....</b>	<b>220 - 223</b>
<i>D. Grant Campbell</i>	
<b>Tecnologias da web semântica aplicadas a organização do conhecimento: padrão SKOS para construção e uso de vocabulários controlados descentralizados.....</b>	<b>224 - 233</b>
<i>José Eduardo Santarem Segundo e Caio Saraiva Coneglian</i>	
<b>Análise preliminar sobre a conversão de plano de classificação em vocabulário controlado.....</b>	<b>234 - 241</b>
<i>Luciana Davanzo e Walter Moreira</i>	
<b>Reengenharia de tesauro: caso do Thesagro.....</b>	<b>242 - 250</b>
<i>Benildes Maculan, Gercina Lima, Ivo Pierozzi Jr. e Leandro Oliveira</i>	
<b>Modelo conceitual de inteligência competitiva organizacional.....</b>	<b>251 - 261</b>
<i>Thiciane Mary Carvalho Teixeira e Marta Lígia Pomim Valentim</i>	
<b>Controle de vocabulário em periódicos científicos eletrônicos.....</b>	<b>262 - 271</b>
<i>José Carlos Francisco dos Santos e Brígida Maria Nogueira Cervantes</i>	
<b>As linguagens documentárias na descrição arquivística.....</b>	<b>272 - 282</b>
<i>Maria de Fátima Santos de Lima e Francisco Aragão Pedroza da Cunha</i>	
<b>As pesquisas em sistemas de organização do conhecimento nas conferências DCMI.....</b>	<b>283 - 294</b>
<i>Felipe Augusto Arakaki, Plácida Leopoldina Ventura Amorim da Costa Santos e Rachel Cristina Vesú Alves</i>	
<b>Modelos conceituais e representação descritiva da informação.....</b>	<b>295 - 302</b>
<i>Elisabete Gonçalves de Souza e Wellington Freire Cunha Costa</i>	
<b>Nanopublicações e indexação: experimento metodológico.....</b>	<b>303 - 317</b>
<i>Lorena Tavares de Paula e Maria Aparecida Moura</i>	
<b>Aplicação de ontologias no processo de recuperação da informação aplicado em ambientes acadêmicos.....</b>	<b>318 - 328</b>
<i>Caio Saraiva Coneglian, Elvis Fusco e José Eduardo Santarém Segundo</i>	
<b>Atuação bibliotecária na análise de assunto de teses nas dimensões teóricas da catalogação de assunto e indexação.....</b>	<b>329 - 337</b>
<i>Roberta Cristina Dal'Evedove Tartarotti, Paula Regina Dal'Evedove e Mariângela Spotti Lopes Fujita</i>	
<b>A contribuição da identificação documental para a organização do conhecimento em arquivos pessoais.....</b>	<b>338 - 343</b>
<i>Gabrieli Aparecida da Fonseca e Sonia Maria Troitiño Rodriguez</i>	

<b>Análise documental de conteúdo e o percurso gerativo de sentido: na representação do documento de arquivo.....</b>	<b>344 - 354</b>
<i>Gilberto Gomes Cândido, João Batista Ernesto de Moraes e Deise Sabbag</i>	
<b>Contribuições do percurso gerativo de sentido para a leitura documentária de textos narrativos de ficção.....</b>	<b>355 - 364</b>
<i>Deise Sabbag e João Batista Ernesto de Moraes</i>	
<b>O modelo teórico do percurso gerativo de sentido para a elaboração de resumos de textos científicos: uma releitura do produto informacional resumo.....</b>	<b>365 - 372</b>
<i>Roberta Caroline Vesú Alves e João Batista Ernesto de Moraes</i>	
<b>El control de vocabulario y su aporte a la investigación: el caso de las ciencias agrarias de Uruguay.....</b>	<b>373 - 380</b>
<i>Lucía Simón e Mario Barité</i>	
<b>Bens culturais rurais: relato de experiência da construção de taxonomia no contexto das fazendas históricas paulistas.....</b>	<b>381 - 391</b>
<i>Mayara Cristina Bernardino, Luciana de Souza Gracioso, Maria da Graça Melo Simões e Luzia Sigoli Fernandes Costa</i>	
<b>Harmonization of CIDOC CRM ontology in the context of archives, libraries and museums.....</b>	<b>392 - 398</b>
<i>Laís Barbudo Carrasco, Silvana Aparecida Borsetti Gregório Vidotti e Phil Manfred Thaller</i>	
<b>Catalogación de recursos interactivos: análisis de recomendaciones y prácticas en catálogos internacionales.....</b>	<b>399 - 412</b>
<i>Víctor Amor, Daniel Martínez-Ávila e Rosa San Segundo</i>	
<b>Estrutura de metadados para banco de imagens em patologia.....</b>	<b>413 - 424</b>
<i>Elisabete Gonçalves Souza, Joice Cleide Cardoso Ennes de Souza e Elan Cardozo Paes de Almeida</i>	
<b>Indexação por imagens: acessibilidade via OPACs imagéticos.....</b>	<b>425 - 448</b>
<i>Rita de Cássia do Vale Caribé e Marcílio de Brito</i>	
<b>Mapas conceituais interativos como instrumento didático no campo da organização da informação e do conhecimento.....</b>	<b>449 - 460</b>
<i>Marilda Lopes Ginez de Lara, Gabriela Previdello e Nair Yumiko Kobashi</i>	
<b>Análise de domínio na organização do conhecimento: explorando as relações temáticas e de citação.....</b>	<b>461 - 468</b>
<i>Bruno Henrique Alves, Ely Francina Tannuri de Oliveira e Maria Cláudia Cabrini Grácio</i>	
<b>Um estudo do domínio em organização do conhecimento por meio da análise de citação e cocitação de autores.....</b>	<b>469 - 476</b>



*Lidyane Silva Lima, Pollyana Ágata Gomes da Rocha Custódia, Ely Francina Tannuri de Oliveira e Leilah Santiago Bufrém*

**Abordagens de classificação nos trabalhos apresentados ao GT2 do ENANCIB: classificação profissional e classificação não-profissional.....477 - 486**  
*Walter Moreira e Isabela Santana de Moraes*

**O tema comportamento informacional no âmbito da ISKO - International Society For Knowledge Organization: estudo referente ao período de 2004 – 2014.....487 - 497**  
*Marli Vitor da Silva e Helen de Castro Silva Casarin*

**A representação de assunto no escopo da arquivologia: uma análise de artigos científicos nacionais e internacionais.....498 - 506**  
*Graziela Martins de Medeiros, Leolibia Luana Linden, Luciane Paula Vital e Marisa Bräscher*

**Análise da metodologia em estudos bibliométricos: uma proposição de indicadores de contexto.....507 - 517**  
*Maria Guiomar da Cunha Frota e Ana Cláudia Ribeiro*

**Organização e representação do conhecimento: contribuições aos estudos métricos.....518 - 525**  
*Cibele Araújo Marques dos Santos*

**Relativizando os índices h e g: estudo aplicado ao domínio dos estudos métricos.....526 - 533**  
*Deise Deolindo Silva e Maria Cláudia Cabrini Grácio*

**A abordagem da análise de domínio na organização e representação do conhecimento em arquivística.....534 - 544**  
*Cynthia Maria Kiyonaga Suenaga e Brígida Maria Nogueira Cervantes.*

**O Familysearch Indexing como iniciativa de crowdsourcing no contexto da organização do conhecimento.....545 - 552**  
*Paula Carina de Araújo e José Augusto Chaves Guimarães*

**Estudo de ações para disponibilização de datasets governamentais em linked open data.....553 - 564**  
*Fernando de Assis Rodrigues e Ricardo César Gonçalves Santana*

**Organização do conhecimento e o patrimônio industrial em São Paulo: o projeto Eletromemória.....565 - 573**  
*Vânia Mara Alves Lima, Marcia Cristina e Carvalho Pazin Vitoriano e Cristina Hilsdorf Barbanti*

## A DIMENSÃO SOCIAL, CULTURAL E POLÍTICA DA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

**Cultural Interoperability and Knowledge Organization Systems.....575 - 606**  
*Widad Mustafa El Hadi*

**Reflexões acerca do desenvolvimento de metodologia para análise de assunto.....607 – 614**  
*Paula Regina Dal'Evedove, Roberta Cristina Dal'Evedove Tartarotti e Mariângela Spotti Lopes Fujita*

**Organização, representação, recuperação e acesso à informação: (re)configuração do Formato MARC21 e do BIBFRAME pelos propósitos da diversidade cultural nos ambientes informacionais digitais?.....615 - 620**  
*Zaira Regina Zafalon e Marcela Cristina Nespoli*

**Música, literatura e audiovisual: as contribuições da organização do conhecimento (OC) nas relações de interseccionalidade das obras de Dorival Caymmi e Jorge Amado.....621 - 629**  
*Fabio Assis Pinho, Francisco Arrais Nascimento, e Andréa Carla Melo Marinho*

**A mediação no domínio da organização do conhecimento.....630 - 642**  
*Mona Cleide Quirino da Silva Farias, Carlos Cândido de Almeida e Daniel Martínez-Ávilla*

**O valor da informação e da linguagem na sociedade de consumo.....643 - 649**  
*Luciana de Souza Gracioso*

**Formação acadêmica do docente em descrição arquivística.....650 - 657**  
*Laura Maria do Rego, José Augusto Chaves Guimarães e Natália Bolfarini Tognoli*

### PERSPECTIVAS DA PESQUISA EM ORGANIZAÇÃO DO CONHECIMENTO NO BRASIL

**Knowledge Organization and the core of Information Science: tales of big data, computing clouds and social networks.....659 - 669**  
*Renato Rocha Souza, Mauricio Barcellos Almeida e Renata Abrantes Baracho Porto*

**Organização do conhecimento: pesquisa e desenvolvimento.....670 - 687**  
*Gercina Ângela Borém de Oliveira Lima*

**A ISKO Brasil e os grupos de pesquisa em organização do conhecimento.....688 - 695**  
*Evelyn Goyannes Dill Orrico*

**Perspectivas da pesquisa em organização do conhecimento: reflexões a partir de trabalhos acadêmicos do PPGCI IBICT-UFRJ.....696 - 708**  
*Rosali Fernandez de Souza*

<b>Perspectivas da pesquisa sobre organização do conhecimento no Brasil.....</b>	<b>709 - 724</b>
<i>Leilah Santiago Bufrem</i>	

## **PESQUISA EM ANÁLISE DOCUMENTÁRIA NO BRASIL: A INFLUÊNCIA DE JEAN-CLAUDE GARDIN**

<b>A busca pela eficiência na representação da informação e do conhecimento – desdobramentos posteriores no pensamento de Gardin.....</b>	<b>726 - 735</b>
<i>Johanna W. Smit</i>	

<b>Jean-Claude Gardin e a análise documentária: trajetória da constituição de uma semiologia da representação.....</b>	<b>736 - 747</b>
<i>Maria de Fátima Gonçalves Moreira Tálamo e Giovana Deliberali Maimone</i>	

<b>Linguagem documentária na ótica de J.-C. Gardin.....</b>	<b>748 - 757</b>
<i>Marilda L. G de Lara</i>	

<b>Da análise documentária à terminologia: percurso teórico e metodológico.....</b>	<b>758 - 764</b>
<i>Vania M. A. Lima</i>	

<b>Ordenação de documentos: fundamentos e relações com a classificação bibliográfica.....</b>	<b>765 - 797</b>
<i>Cristina D. Ortega</i>	

<b>A ISKO-Brasil e o fortalecimento de intercâmbios e visibilidade científica em organização do conhecimento.....</b>	<b>798 – 802</b>
<i>Vera Dodebei</i>	

**Resumos dos trabalhos ganhadores do prêmio ISKO Brasil 2015, a serem  
publicados, na íntegra, na revista Knowledge Organization**

<b>A organização do conhecimento no contexto da pós-modernidade sob a perspectiva da teoria da classificação.....</b>	<b>803</b>
<i>Daniel Abraão Pando e Carlos Cândido de Almeida</i>	

<b>O princípio da garantia semântica e os estudos da Linguagem: reflexões conceituais.....</b>	<b>804</b>
<i>Roger de Miranda Guedes e Maria Aparecida Moura</i>	

<b>Interoperabilidade e Mapeamentos entre Sistemas de Organização do Conhecimento: Metatesouro-Unified Medical Language System- National Library of Medicine.....</b>	<b>805</b>
<i>Julietti de Andrade e Marilda Lopes Ginez de Lara</i>	

<b>Aspectos metodológicos nas teorias críticas em Organização do Conhecimento.....</b>	<b>806</b>
<i>Daniel Martinez-Ávila, Rafael Aparecido Moron Semidão e Marcio Ferreira da Silva</i>	

**Resumos dos trabalhos ganhadores do prêmio ISKO Brasil 2015, a serem publicados, na íntegra, na revista Brazilian Journal of Information Science.**

**Propuesta de representación del Tesauro de la UNESCO para la web semántica mediante ISO-25964.....807**

*Juan-Antonio Pastor-Sánchez*

**Modelagem Metaconceitual: aplicação à Infraestrutura de Qualidade com base na ISO 25964.....808**

*Catarina Felix dos S. Soares e Vera Dodebei*

**Tres enfoques de *bias* en organización del conocimiento: *bias* neutro, *bias* negativo y *bias* positivo.....809**

*Stephanie Colombo e Mario Barité*

## Estudo de ações para disponibilização de datasets governamentais em linked open data



**Fernando de Assis Rodrigues**  
Universidade Estadual Paulista  
fernando@elleth.org



**Ricardo César Gonçalves Sant'Ana**  
Universidade Estadual Paulista  
ricardosantana@marilia.unesp.br

### 1 Introdução

Os princípios de *Linked Open Data* (LOD) estabelecem um novo modo de compartilhamento de *datasets* (conjuntos de dados) abertos pela internet, com objetivo de promover a ampla distribuição de dados estruturados em linguagens como a *eXtensible Markup Language* (XML) e em concomitância com as recomendações do padrão *Resource Description Framework* (RDF) (BERNERS-LEE, 2009; BIZER; HEATH; BERNERS-LEE, 2009; HEATH, 2015; W3C, 2014, 2015).

Neste cenário, *datasets* governamentais têm papel de destaque: representam 18,58% da quantidade total de *datasets* de LOD existentes e 41,54% destes *datasets* governamentais possuem ao menos um vínculo com ontologias ou vocabulários controlados, segundo resultados do mapeamento elaborado pelo projeto *Linking Open Data cloud diagram* (SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Entretanto, segundo Schmachtenberg, Bizer e Paulheim (2014a, p. 12–16) ainda existem características nas estruturas dos *datasets* de LOD no momento da recuperação de dados que não são consideradas ideais ou boa prática adotada, tais como a ausência de metadados e de licenças de uso.

Ações para tornar acessíveis dados governamentais públicos são parte integrante de discussões sobre tendências de modernização dos modelos de administração pública, que procuram redistribuir competências e recursos dentre diferentes organizações

intra e extragovernamentais, permitindo um maior pluralismo institucional nas funções públicas (MALIN, 2006; SANT'ANA; RODRIGUES, 2013).

O fortalecimento de ações de transparência pode ser ampliado com a construção de ambientes de compartilhamento de informação que, dentre outras características, proporcionem a ampliação de fluxos informacionais entre a administração pública e a sociedade, garantindo maior visibilidade das atividades do Estado (BOHMAN, 2000; MARCONDES; JARDIM, 2003). Estes ambientes se tornam componentes de uma maior participação cidadã, estendendo possibilidades de participação além do voto; e o Estado pode melhorar a efetividade e o acompanhamento das atividades e resultados de suas ações, além de atender obrigatoriedade de publicação de dados governamentais (BRASIL, 2011; SANT'ANA; RODRIGUES, 2013).

O acesso aos *datasets* governamentais sobre resultados de votações legislativas é importante no acompanhamento das atividades dos representantes, servindo de apoio a construção de análises, tais como “[...] a identificação de *clusters* partidários” e na “[...] coerência de cada um de nossos representantes nas votações durante seus mandatos” (SANT'ANA; RODRIGUES, 2013, p. 58).

### **Objetivo e delimitação do tema**

O objetivo deste trabalho é o estudo de ações necessárias para a disponibilização de *datasets* governamentais em *Linked Open Data*, a partir da aplicação de um modelo de recomendações para publicação de dados “*Linked Data Best Practices in Different Topical Domains*”, proposto por Schmachtenberg, Bizer e Paulheim (2014a), em bases de dados disponíveis sobre votações legislativas do Senado brasileiro.

O objeto de pesquisa foi delimitado a dados disponíveis nas Tecnologias de Comunicação e Informação (TIC) oficiais do Senado brasileiro, mais precisamente sobre votações existentes no ‘Portal e-Cidadania – Dados Abertos’, analisados entre janeiro e março de 2015.

### **Procedimentos metodológicos**

A metodologia adotada foi a análise exploratória do objeto de pesquisa, de caráter qualitativa, através da explicitação das características: do *dataset* existente (a localização do recurso no web site, informações sobre página descritiva e arquivos disponíveis), e; da estrutura dos dados encontrada no momento da recuperação. Estas características formaram um conjunto de informações que serviram como subsídio para a proposta de uma estratégia de ações necessárias para a reestruturação destes dados, já existentes, em concomitância com as recomendações estabelecidas de boas práticas de disponibilização de *datasets* em LOD, proposta por Schmachtenberg, Bizer e Paulheim (2014a).

## 2 Referencial teórico

Schmachtenberg, Bizer e Paulheim (2014a, 2014b) propõem um modelo com recomendações para publicação de dados, com o objetivo de identificar a concomitância aos conceitos de LOD e boas práticas para compartilhamento de dados por *datasets* públicos armazenados em diversos domínios da web. Estas recomendações foram elaboradas a partir das práticas de comunidades e dos resultados apresentados pelo mapeamento de *datasets* LOD elaborado pelo projeto *Linking Open Data cloud diagram* (JENTZSCH; CYGANIAK; BIZER, 2011). O modelo está dividido em nove recomendações:

### *Fornecimento de identificadores únicos de localização*

No processo de recuperação de dados é necessário que *datasets* possuam identificadores únicos para auxiliar o processo de recuperação de dados por agentes externos, em concomitância com o primeiro princípio de LOD (BIZER; HEATH; BERNERS-LEE, 2009). Estes identificadores únicos devem estar em conformidade com regras estabelecidas pelo *Uniform Resource Identifier (URI)* e pelo padrão RDF (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

### *Definição de ligações com outros datasets*

É necessário que o *dataset* possua ligações com outros *datasets* através das regras de relacionamentos estabelecidas pelo padrão RDF. Esse procedimento facilita a coleta automatizada por agentes externos de dados disponíveis do *dataset* e dos *datasets* a qual foram vinculados (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

#### *Uso de vocabulários controlados e ontologias existentes*

Como o dado é um elemento básico “[...] formado por signo ou conjunto finito de signos que não contém, intrinsecamente, um componente semântico, mas somente elementos sintáticos” (SANTOS; SANT’ANA, 2002, s. p.), é necessário o uso de vocabulários controlados e ontologias para ampliar a carga semântica no momento de coleta de dados por agentes externos, tais como o *Dublic Core (DC)*, o *Friend of a Friend (FOAF)*, o *Simple Knowledge Organization System (SKOS)*, entre outros (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

#### *Definições de termos, elementos e atributos referenciáveis em vocabulários e ontologias*

Os documentos complementares vinculados ao *dataset*, contendo informações sobre ontologias e vocabulários controlados, devem possuir URI únicas para termos, elementos e atributos (SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b). Por exemplo, no FOAF as definições dos termos ‘name’ e ‘birthday’ devem ser acessíveis por URI únicas, seja através do uso de símbolos de divisão (‘/’), ou uso de *hashtag* (#) para diferenciar o acesso aos termos (BRICKLEY; MILLER, 2014).

#### *Vinculação de termos entre vocabulários*

Caso ocorra a necessidade de elaboração de vocabulários próprios, é importante que termos deste novo vocabulário sejam vinculados a termos de vocabulários já existentes e com maior abrangência, como: o DC, o FOAF, o SKOS, entre outros. A vinculação de novos vocabulários com vocabulários abrangentes



proporcionam um maior repertório de informações sobre termos desenvolvidos para o vocabulário próprio aos agentes externos (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

#### *Fornecimento de metadados*

No momento da recuperação, *datasets* devem ter elementos de metadados para garantir qualidade no processo de recuperação de dados, identificar a origem dos dados e avaliar a qualidade (JENTZSCH; CYGANIAK; BIZER, 2011). Os metadados “[...] são um fator chave para minimizar os problemas de busca e recuperação dos recursos nos diversos ambientes informacionais [...]” (SANTOS; ALVES, 2009) e é recomendado que: seus elementos estejam disponíveis no elemento raiz, e; a utilização de elementos do DC (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

#### *Uso de licença nos metadados*

Os metadados do *dataset* precisam conter licenças de autoria em seus elementos e atributos, tais como: o *Creative Commons*, o *Open Data Commons Attribution License*, o *Open Database License (ODbL)*, entre outros (JENTZSCH; CYGANIAK; BIZER, 2011; W3C, 2011).

#### *Disponibilização de metadados sobre a estrutura do dataset*

Na recuperação do *dataset* devem existir metadados contendo informações sobre a sua estrutura – disponibilizadas junto aos dados ou em documentos suplementares – delimitando elementos, iterações, termos e atributos utilizados (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

#### *Uso de métodos alternativos para recuperação de dados*

A forma mais comum de disponibilização de *datasets* estruturados no padrão RDF é através de um *SPARQL Endpoint* (JENTZSCH; CYGANIAK; BIZER, 2011), que habilita agentes externos a realizar buscas estruturadas na linguagem de consulta *SPARQL*. Todavia, é recomendado que também estejam disponíveis arquivos de despejo (*RDF Dump*), explicitados no padrão RDF/XML ou equivalente (SEMANTICWEB.ORG, 2011).

### **Características do *dataset***

O Portal e-Cidadania tem o objetivo de promover transparência de ações e atividades do Senado Brasileiro, através do acesso a dados (BRASIL, 2015a). Em janeiro de 2015, o *web site* possuía quarenta e cinco *datasets*, divididos em oito grupos: ‘Projetos e Matérias’, ‘Plenário’, ‘Parlamentares’, ‘Composição’, ‘Comissões’, ‘LexML’, ‘Legislação’, e ‘Processo Legislativo’.

O grupo ‘Plenário’ contém oito subdivisões: ‘Diários do Senado’ e do Congresso’; ‘Legislaturas e ‘Sessões Legislativas’; ‘Matérias com prazos’; ‘Pronunciamentos de senador’; ‘Questões de Ordem’; ‘Sessões do Plenário’; ‘Tabelas de tipos relacionados a plenário’, e; ‘Votações nominais’ – este último contendo dados sobre “[...] votos registrados nas sessões de votação em plenário [...]” e “[...] informações sobre as sessões, as matérias votadas, totalização de votos e o voto de cada parlamentar” (BRASIL, 2015b).

Votações nominais é composto por 11 itens: 9 arquivos de despejo no formato XML, contendo dados sobre votações agrupadas anualmente; 1 hiperlink para um *web service* de consulta automatizada, e; 1 hiperlink para uma página de consulta de votações – este último, por se tratar de consultas em formato *HyperText Markup Language* (HTML) não é objeto deste estudo.

### **Características da estrutura dos dados na recuperação**

Cada arquivo de despejo no formato XML possui uma URL para identificação, formado pela composição: domínio/hierarquia ‘*http://legis.senado.leg.br/dadosabertos/dados/*’, prefixo ‘*ListaVotacoes*’, ano referente aos dados, e; extensão de arquivo ‘*.xml*’.

Na recuperação de dados via *web service*, os *datasets* apresentam apenas resultados agrupados por consultas diárias. Por exemplo, para realizar uma coleta de dados sobre votações ocorridas em um determinado mês, é necessário realizar ‘x’ consultas, onde ‘x’ representa o total de dias do mês. Essa recuperação de *datasets* possui uma URL para identificação de resultado diário, cujo valor da URL é formado pela composição: domínio/hierarquia ‘<http://legis.senado.leg.br/dadosabertos/plenario/lista/votacao/>’; ano; mês, e; dia.

Em ambos os casos – na coleta dos arquivos de despejo e na recuperação de dados via *web service* – os *datasets* estão explicitados na linguagem XML e as coleções de elementos, atributos e termos disponíveis são idênticas (Quadro 1).

Elemento	Associado ao elemento	Tipo de Dado	Atributos
ListaVotacoes	Nenhum (Elemento Raiz)	Elem. Agrupador	'xmlns:xsi' e 'xsi:noNamespaceSchemaLocation'
Metadados	ListaVotacoes	Elem. Agrupador	Nenhum
Votacoes	ListaVotacoes	Elem. Agrupador	Nenhum
Versao	Metadados	Texto	Nenhum
VersaoServico	Metadados	Número inteiro	Nenhum
DescricaoDataSet	Metadados	Texto	Nenhum
Votacao	Votacoes	Elem. Agrupador	Nenhum
CodigoSessao	Votacao	Número inteiro	Nenhum
SiglaCasa	Votacao	Texto	Nenhum
CodigoSessaoLegislativa	Votacao	Número inteiro	Nenhum
TipoSessao	Votacao	Texto	Nenhum
NumeroSessao	Votacao	Número inteiro	Nenhum
DataSessao	Votacao	Texto	Nenhum
HoraInicio	Votacao	Texto	Nenhum
CodigoTramitacao	Votacao	Número inteiro	Nenhum
CodigoSessaoVotacao	Votacao	Número inteiro	Nenhum
SequencialSessao	Votacao	Número inteiro	Nenhum
Secreta	Votacao	Texto	Nenhum
DescricaoVotacao	Votacao	Texto	Nenhum
Resultado	Votacao	Texto	Nenhum
TotalVotosSim	Votacao	Número inteiro	Nenhum
TotalVotosNao	Votacao	Número inteiro	Nenhum
TotalVotosAbstencao	Votacao	Número inteiro	Nenhum
CodigoMateria	Votacao	Número inteiro	Nenhum
SiglaMateria	Votacao	Texto	Nenhum
NumeroMateria	Votacao	Número inteiro	Nenhum
AnoMateria	Votacao	Número inteiro	Nenhum
Votos	Votacao	Elem. Agrupador	Nenhum

VotoParlamentar	Votos	Elem. Agrupador	Nenhum
CodigoParlamentar	VotoParlamentar	Número inteiro	Nenhum
NomeParlamentar	VotoParlamentar	Texto	Nenhum
SexoParlamentar	VotoParlamentar	Texto	Nenhum
Url	VotoParlamentar	Texto	Nenhum
Foto	VotoParlamentar	Texto	Nenhum
Tratamento	VotoParlamentar	Texto	Nenhum
Voto	VotoParlamentar	Texto	Nenhum

**Quadro 1** – Estrutura de elementos e atributos do *dataset*

**Fonte:** Elaborado pelos autores

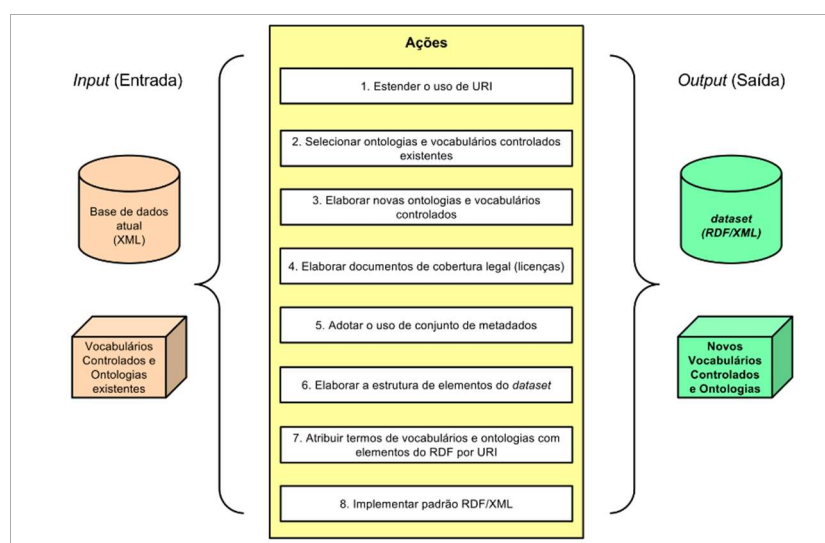
O elemento raiz é denominado ‘ListaVotacoes’ e possui dois atributos e dois elementos. Seus dois atributos são responsáveis por vincular o *dataset* com um documento suplementar (*XML Schema*), contendo a delimitação de elementos, tipos de conteúdos e atributos disponíveis.

Seus elementos a) ‘Metadados’ e b) ‘Votacoes’ são elementos agrupadores com seu valor sendo um conjunto de um ou mais elementos; ambos sem atributos. O elemento agrupador ‘Metadados’ possui três elementos; ‘Votacoes’ contém um ou mais elementos ‘Votacao’; e o elemento ‘Votacao’ possui dezenove elementos. Nenhum dos elementos possui atributos.

O elemento agrupador ‘Votos’ (filiado ao elemento ‘Votacao’) contém um ou mais elementos ‘VotoParlamentar’, sem atributos. O elemento ‘VotoParlamentar’ possui sete elementos, também sem atributos.

### 3 Resultados

A partir da análise, foram propostas oito ações para a disponibilização de *datasets* governamentais em LOD a partir da recuperação de dados já existente (Figura 1).



**Figura 1** – Síntese de ações necessárias ao desenvolvimento de *datasets* governamentais

**Fonte:** Elaborado pelos autores.

Anteriormente a execução destas ações é importante que estejam disponíveis informações (*input*) sobre as características das bases de dados já existentes e um conhecimento prévio dos envolvidos sobre ontologias e vocabulários existentes que possam ser parte da explicitação de relacionamentos e de características de elementos do novo *dataset* LOD.

As ações identificadas neste trabalho podem ser resumidas em:

- Ação 1: estender o uso de identificadores URI para a identificação de elementos, atributos e termos que compõem o RDF, explicitados em documentos suplementares ao *dataset* para a compreensão das regras deste sistema;
- Ação 2: selecionar ontologias e vocabulários controlados de uso por muitas comunidades que possam ser parte da explicitação de relacionamentos e elementos do novo *dataset* LOD;
- Ação 3: elaborar ontologias e vocabulários específicos para a explicitação de relacionamentos e elementos não existentes em ontologias e vocabulários da segunda ação;
- Ação 4: elaborar documentos complementares contendo a cobertura legal como licenças de uso e direitos autorais, e vincular estes documentos com o *dataset*. É importante que estas licenças sejam explicitadas nos metadados (quinta ação);

- Ação 5: adotar o uso de conjuntos de elementos de metadados de iniciativas populares para explicitar o conteúdo do *dataset* na recuperação de dados por agentes externos;
- Ação 6: elaborar a estrutura lógica dos elementos do *dataset*, atributos, títulos de elementos, valores, agrupadores e regras de validação;
- Ação 7: atribuir termos das ontologias e vocabulários selecionados aos elementos do *dataset*, vinculando-os através de URI, para ampliar a carga semântica destes dados no momento de coleta;
- Ação 8: implementar as regras do padrão RDF sobre a estrutura da linguagem de marcação XML, respeitando as formas estabelecidas para documentos RDF/XML.

Estas ações devem desenvolver como resultado (*output*) o *dataset* LOD (explicitado em formato RDF/XML ou equivalente); e vocabulários e ontologias elaboradas para o atendimento das necessidades do contexto dos dados (votações nominais).

#### 4 Conclusões

A forma de recuperação atual dos conjuntos de dados não contempla características importantes na contextualização dos dados, como o uso de vocabulários controlados e de ontologias, interferindo diretamente na independência do agente externo na coleta dos dados.

Apesar da existência de três elementos de metadados não existem informações como: autor, licença, fonte, data de criação e data de publicação.

A aplicação do modelo de recomendação serviu como norteador para a formação de ações propostas neste estudo, principalmente ao fornecer subsídios para elaboração do conjunto de ações necessárias para disponibilização de *datasets* governamentais em LOD, fornecendo aos gestores públicos pontos de intervenção necessários para que dados já disponíveis seus *websites* possam ser reestruturados em *datasets* LOD e, conseqüentemente, passíveis de coleta por agentes externos.

Espera-se que a aplicação destas ações em dados sobre votações sirva de estímulo para a aplicação destas ações em outras bases de dados, em outras esferas

e em outros *websites*; como também estimular o surgimento de novos *datasets* governamentais brasileiros nesta temática.

## Referências

BERNERS-LEE, T. **The next Web**. In: TED2009. Estados Unidos da América: fev. 2009.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data – The Story So Far: **International Journal on Semantic Web and Information Systems**, v. 5, n. 3, p. 1–22, 33 2009.

BOHMAN, J. **Public deliberation: Pluralism, complexity, and democracy**. [s.l.] MIT press, 2000.

BRASIL. 12.527. Lei Nº 12.527, de 18 de Novembro de 2011 (Lei de Acesso à Informação). . 18 nov. 2011.

BRASIL. **Portal e-Cidadania**, 2015a. Disponível em: <<http://dadosabertos.senado.gov.br>>. Acesso em: 3 abr. 2015

BRASIL. **Votações Nominais**, 2015b. Disponível em: <<http://dadosabertos.senado.gov.br/dataset/votaces-nominais>>. Acesso em: 3 abr. 2015  
BRICKLEY, D.; MILLER, L. **FOAF Vocabulary Specification 0.99**, 14 jan. 2014. Disponível em: <<http://xmlns.com/foaf/spec/20140114.html>>. Acesso em: 27 mar. 2015.

HEATH, T. **Frequently Asked Questions** Linked Data - Connect Distributed Data across the Web, , 2015. Disponível em: <<http://linkeddata.org/faq>>. Acesso em: 22 mar. 2015

JENTZSCH, A.; CYGANIAK, R.; BIZER, C. **State of the LOD Cloud**, set. 2011. Disponível em: <<http://lod-cloud.net/state/>>. Acesso em: 24 mar. 2015

MALIN, A. M. B. Gestão da Informação Governamental: em direção a uma metodologia de avaliação. **DataGramaZero - Revista de Ciência da Informação**, v. 7, n. 5, out. 2006.

MARCONDES, C. H.; JARDIM, J. M. Políticas de Informação Governamental: a construção de Governo Eletrônico na Administração Federal do Brasil. **DataGramaZero - Revista de Ciência da Informação**, v. 4, n. 2, abr. 2003.

SANT'ANA, R. C. G.; RODRIGUES, F. DE A. Visualização de afinidades entre parlamentares mediante dados de votações no Senado Brasileiro. **Informação & Sociedade: estudos**, v. 23, n. 1, p. 49–59, jan. 2013.

SANTOS, P. L. V. A. DA C.; ALVES, R. C. V. Metadados e Web Semântica para

estruturação da Web 2.0 e Web 3.0. **DataGramaZero – Revista de Ciência da Informação**, v. 10, n. 6, dez. 2009.

SANTOS, P. L. V. A. DA C.; SANT'ANA, R. C. G. Transferência da Informação: análise para valoração de unidades de conhecimento. **DataGramaZero - Revista de Ciência da Informação**, v. 3, n. 2, abr. 2002.

SCHMACHTENBERG, M.; BIZER, C.; PAULHEIM, H. Adoption of the Linked Data Best Practices in Different Topical Domains. In: MIKA, P. et al. (Eds.). . **The Semantic Web – ISWC 2014**. Cham: Springer International Publishing, 2014a. v. 8796p. 245–260.

SCHMACHTENBERG, M.; BIZER, C.; PAULHEIM, H. **State of the LOD Cloud 2014** University of Mannheim, 30 ago. 2014b. Disponível em: <<http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/>>. Acesso em: 25 mar. 2015

SEMANTICWEB.ORG. **SPARQL endpoint** Semantic Web.org, , 26 nov. 2011. Disponível em: <[http://semanticweb.org/index.php?title=SPARQL\\_endpoint&oldid=52541](http://semanticweb.org/index.php?title=SPARQL_endpoint&oldid=52541)>. Acesso em: 4 jan. 2015

W3C. **Data Licensing** W3C, 11 abr. 2011. Disponível em: <<http://www.w3.org/wiki/index.php?title=TaskForces/CommunityProjects/LinkingOpenData/DataLicensing&oldid=49411>>. Acesso em: 20 mar. 2015

W3C. **RDF 1.1 XML Syntax** W3C, 25 fev. 2014. Disponível em: <<http://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225/>>. Acesso em: 22 mar. 2015

W3C. **Linked Data Platform 1.0** W3C, 26 fev. 2015. Disponível em: <<http://www.w3.org/TR/2015/REC-ldp-20150226/>>. Acesso em: 24 mar. 2015