UNIVERSIDADE DE SÃO PAULO
Escola de Artes, Ciências e Humanidades – EACH
Programa de Pós-graduação em Sistemas de Informação – PpgSI

MARCELO TAVARES DE SANTANA

**Uma proposta de publicação de dados do
orçamento público na Web**

São Paulo

2014

Marcelo Tavares de Santana

Uma proposta de publicação de dados do orçamento público na Web

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências. Programa: Sistemas de Informação.

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 04 de dezembro de 2013. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientadora: Prof.^a Dr.^a Gisele da Silva Craveiro

São Paulo
Fevereiro de 2014

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

CATALOGAÇÃO-NA-PUBLICAÇÃO

Biblioteca

Escola de Artes, Ciências e Humanidades da Universidade de São Paulo

Santana, Marcelo Tavares de

Uma proposta de publicação de dados do orçamento público na Web
/ Marcelo Tavares de Santana ; orientador, Gisele da Silva Craveiro. –
São Paulo, 2014.

101 f. : il.

Dissertação (Mestrado em Ciências) - Programa de Pós-
Graduação em Sistemas de Informação, Escola de Artes, Ciências
e Humanidades, Universidade de São Paulo.

Versão corrigida.

1. Administração pública – Processamento de dados. 2.
Informação pública. 3. Tecnologia da informação. 4. Orçamento
público. 5. Processamento eletrônico de dados. 6. Web
semântica. I. Craveiro, Gisele da Silva, orient. II. Título.

CDD 22.ed. – 350.0285

Folha de Aprovação

Dissertação sob o título “*Uma proposta de publicação de dados do orçamento público na Web*”, defendida por Marcelo Tavares de Santana e aprovada em 04 de dezembro de 2013, em São Paulo, Estado de São Paulo, pela banca examinadora constituída pelos doutores:

Prof.^a Dr.^a Gisele da Silva Craveiro
Orientadora

Prof. Dr. Edson Luiz Riccio
Universidade de São Paulo

Prof. Dr. José de Jesús Pérez Alcázar
Universidade de São Paulo

Dedico esse trabalho a toda minha família, parentes próximos e distantes, que num ambiente de solidariedade e humildade me ajudaram na minha construção.

Agradeço o apoio de todos os colegas do meio acadêmico e ativista, que em nossas conversas me ajudaram a pensar num tema desafiador e necessário à sociedade. Agradeço aos professores da pós-graduação pelos exemplos de excelência e exigência científica. Agradecimentos especiais à Prof^a Dr^a Gisele da Silva Craveiro por me acompanhar e auxiliar num caminho de busca por excelência científica que continuará sendo difícil e prazeroso.

Todo o poder emana do povo, que o exerce por meio de representantes eleitos ou diretamente, nos termos desta Constituição.

(Brasil, 1988)

Resumo

SANTANA, Marcelo Tavares de. **Uma proposta de publicação de dados do orçamento público na Web**. 2014. 101 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2013.

Este trabalho apresenta uma proposta de publicação de dados de execução do orçamento público em acordo com a legislação brasileira. A partir das leis e da revisão da literatura foram levantados requisitos que devem ser considerados numa publicação de dados que seja processável por máquina. São apresentadas dimensões de dados como qualidade de dados, metadados e taxonomia, além das abordagens Dados Governamentais Abertos e Linked Data. A proposta de publicação resultante além de alinhada com os requisitos estudados prevê processamento por máquina em três níveis: tratamento de dados, validação de dados e colecionamento de conjuntos de dados. Com as tecnologias selecionadas, como XBRL e XSLT, alcançou-se um modelo que atendeu cartorze requisitos levantados, abrindo caminho para estudos como a de uma rede de conjuntos de dados baseada nesta proposta.

Palavras-chave: orçamento público, Dados Governamentais Abertos, dados abertos, metadados, Linked Data, XBRL

Abstract

SANTANA, Marcelo Tavares de. **A proposal for public budget data publication on the web**. 2014. 101 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2013.

This paper presents a proposal for publishing public budget data in accordance with Brazilian law. From the laws have raised requirements that must be considered in the publication data format that is machine-processable. From the literature review were also raised requirements that are part of the proposed publication of this work, such as controlled vocabularies, metadata and taxonomy. Two approaches of open data were also considered in the preparation of the publication format: Open Government Data and Linked Data. The proposed publication resulting from this work, that use technologies like XBRL and XSLT, is aligned with the raised requirements provides three levels of machine-processing: data processing, data validation and collecting datasets.

Keywords: public budget, Open Government Data, Linked Dada, metadata, machine-processable

Lista de Figuras

1	Publicação de orçamento no Brasil como DGA	16
2	Evolução da Web	17
3	Elaboração de Projeto de LOA na Administração Pública Federal .	22
4	Exemplo de classificação de receita	26
5	Exemplo de classificação de despesa	27
6	Interpretação de página <i>web</i> por programas navegadores e humanos	41
7	Modelo de 5 estrelas de Linked Open Data	42
8	XBRL publicado como Linked Data	44
9	Estrutura de validação e publicação de documentos XBRL	45
10	Requisitos para publicação de dados na literatura	46
11	Camadas de um sistema de informação	49
12	Exemplos de dados serializados	53
13	Modelo de proposta de publicação de dados orçamentário	56
14	Exemplo de serialização YAML	58
15	Transformação de <i>links</i> com XSLT	61
16	Visualização da implementação em XBRL e JSON	64
17	Modelo da proposta XBRL	73
18	Conversões de formatos com XSLT	77

Lista de Tabelas

1	Requisitos identificados na legislação	29
2	Requisitos levantados neste trabalho	50
3	Características de formatos para serialização de dados	55

Sumário

1	Introdução	12
1.1	Objetivos	18
1.2	Justificativa	18
1.3	Organização do texto	19
2	Arcabouço legal brasileiro	20
2.1	Classificação orçamentária	26
2.2	Requisitos de publicação	28
3	Dimensões conceituais sobre dados	30
3.1	Dados Governamentais Abertos	30
3.2	Qualidade de dados	33
3.3	Metadados	34
3.4	Taxonomia	35
3.5	Vocabulários controlados	36
3.6	Proveniência de dados	37
3.7	Versionamento de dados	38
3.8	Validação de dados	39

3.9	Linked Data	40
3.10	Trabalhos relacionados	43
3.11	Requisitos de publicação	44
4	Proposta de publicação	47
4.1	Metodologia	47
4.2	Requisitos e sistemas de informação	48
4.3	Projeto da proposta	54
4.4	Modelo da proposta	56
4.5	Implementação da proposta	57
4.6	Importação de dados	62
5	Discussão da proposta	65
5.1	Comparação com trabalhos relacionados	74
5.2	Contribuições da proposta	75
5.3	Trabalhos futuros	75
6	Considerações finais	79
Anexo A Instância em XBRL		81
Anexo B Taxonomia em XBRL		87
Anexo C Glossário		96
Referências		98

Capítulo 1

Introdução

Em setembro de 2011, o Governo Brasileiro assinou, junto a África do Sul, Estados Unidos, Indonésia, México, Noruega, Filipinas e Reino Unido, a "Declaração de Governo Aberto" (OGP, 2011)[1], assumindo alguns compromissos como o apoio à participação cívica, onde os governos devem "tornar a formulação de políticas e a tomada de decisões mais transparentes, criando e usando canais para requisitar a opinião pública, aprofundando a participação pública na elaboração, monitoramento e avaliação das atividades governamentais" (OGP, 2011)[1].

Para isso, o governo deve investir em tecnologia, criando espaços *online* em sistemas acessíveis e seguros, atraindo a sociedade civil e a comunidade empresarial na identificação, desenvolvimento e implantação de novas tecnologias que promovam a transparência nos governos (OGP, 2011)[1]. A Controladoria Geral da União (CGU) também destaca esse importante papel da participação cidadã sobre a gestão pública, através do controle social que "pode ser entendido como a participação do cidadão na gestão pública, na fiscalização, no monitoramento e no controle das ações da Administração Pública" (CGU, 2010)[2].

A expressão governo eletrônico significa bem mais do que governo informatizado: trata-se do uso da tecnologia da informação e comunicação para se construir um governo aberto e ágil, ampliar a cidadania, aumentar a transparência da gestão e a participação dos cidadãos na fiscalização do poder público, democratizar o acesso aos serviços e informações na Internet e aumentar a eficiência dos serviços públicos. (CGI, 2010)[3]

A legislação brasileira aponta a Internet como meio de divulgação do orçamento, como apresentado na Seção Arcabouço legal brasileiro. Segundo Craveiro *et al.*, 2013[4], a Organização para a Cooperação Econômica e Desenvolvimento, o Fundo Monetário Internacional (FMI) e a Federação Internacional de Contabilistas, além de sugerir ações que se traduzem em boas práticas, apontaram também a Internet como um meio de comunicação com o potencial de fornecer informações de forma oportuna e transparente para os diversos atores sociais. Um exemplo de boa prática é a recomendação brasileira e-MAG (Brasil & Ministério do Planejamento, Orçamento e Gestão, 2011)[5], onde "para se criar um ambiente online efetivamente acessível é necessário, primeiramente, que o código esteja dentro dos padrões Web internacionais definidos pelo W3C", um consórcio internacional que define os padrões abertos utilizados em sites na Internet. O ambiente formado por locais que utilizam os padrões W3C é chamado de Web, onde se deve encontrar os dados de execução do orçamento público.

A partir da publicação de dados governamentais na Web, vemos o desenvolvimento de algumas iniciativas pela sociedade, que busca se apropriar da informação obtida e repassar ou tornar mais acessível esse conhecimento derivado a uma parcela mais ampla da população, como nos exemplos a seguir.

Além disso, o uso da Internet pode potencializar o desenvolvimento da democracia. Um exemplo de participação cidadã é o *site* "Onde Acontece?"¹ que, desenvolvido por cidadãos, constrói visualizações gráficas sobre o mapa do Estado do Rio Grande do Sul baseadas na quantidade de incidências por tipo de delito. Bertot *et al.*(2010:265)[6] relatam, entre alguns exemplos, um sistema de compras no Chile que tem sido usado para permitir que funcionários do governo e cidadãos comparem os custos de licitações e de serviços contratados pelo governo.

Das informações governamentais, aquelas referentes ao orçamento público são as que tratam de como o dinheiro será gasto na saúde, educação, transportes,

¹O sistema foi criado usando dados da segurança pública do Estado do Rio Grande do Sul e pode ser acessado em <http://ondeacontece.com.br/>

segurança, e em outras áreas, sendo muito importante para acompanhar e fiscalizar o uso dos recursos públicos. A legislação brasileira sobre orçamento público tem evoluído significativamente como forma de propiciar maior controle social, inclusive os órgãos públicos devem publicar em tempo real informações detalhadas do orçamento em meios eletrônicos de acesso público(Brasil, 2009)[7]. Porém, mesmo com estruturas e vocabulários definidos em lei, as informações encontradas em páginas de transparência de entes da Federação distintos trazem informações diferentes mesmo com uma legislação orçamentária comum para todos. Isso sugere que um cidadão precisa entender separadamente o que cada um desses entes fizeram ao publicarem informações estruturalmente diferentes, dificultando uma análise comparativa entre essas informações.

Para se ter uma visão dessa evolução, baseado em Corrêa (2010)[8], a evolução da transparência nas contas públicas segue abaixo:

- 2000 - Lei Complementar 101(Brasil, 2000)[9], também chamada Lei de Responsabilidade Fiscal (LRF), pregão eletrônico;
- 2004 - Portal da transparência do Governo Federal;
- 2005 - Páginas de transparência e pregão eletrônico ampliado;
- 2006 - Portais de transparência nos Estados e Municípios;
- 2009 - Lei Complementar 131(Brasil, 2009)[7] amplia responsabilidades na LRF;
- 2011 - Aprovação da Lei Geral de Acesso à Informação;
- 2013 - Publicação de orçamento na Internet obrigatória para todos entes da federação.

Como em qualquer meio profissional ou acadêmico, são necessários dados para tomadas de decisões, caso contrário a participação fica comprometida; "o uso da

Web para o envolvimento é significativo, por abrir novas maneiras de falar com o governo, mas essas conversas são complexas pois os limites entre participação e envolvimento muitas vezes são obscuros” (W3C, 2009)[10].

Mesmo um cidadão especialista em contas públicas pode não ser atendido pelos portais de transparência, por exemplo, pelo tempo que se leva para achar a informação, o que abre oportunidades de aprimoramento desses portais, oferecendo melhor localização, mais agilidade e facilidade na consulta do orçamento na Internet. Com esses problemas, todo o contexto de participação, governo eletrônico, transparência e forma de publicação levou à discussão sobre a disponibilidade de dados e como eles são postos para a sociedade.

Muitas informações do setor público estão sendo publicadas em formatos proprietários, dificultando o acesso a todas as partes interessadas, causando falta de informação para computadores sem um determinado software proprietário e ou barreiras de acessibilidade para portadores de deficiências (W3C, 2009)[10]. Nesse ponto a padronização se torna necessária na promoção da interoperabilidade de sistemas, inclusive na comunicação entre estado e cidadãos.

Um estudo de 2013[4] trás um exame da qualidade dos dados de execução do orçamento de 88 portais de transparência no Brasil em relação aos oito princípios de Dados Governamentais Abertos (DGA) e, embora em alguns aspectos uma boa avaliação tenha sido feita, uma das principais conclusões leva à falta de um nível primário adequado de dados, sendo esses apenas 17,16% de todos os conjuntos de dados analisados. Além disso, outro resultado mostra que 30% dos conjuntos de dados eram processáveis por máquina. No entanto, na maioria dos casos os dados foram publicados de forma semi-estruturada, sem metadados nem vocabulário comum, podendo inferir em limitações no processamento de dados e exigir maior intervenção humana.

A Figura 1 resume a conformidade da amostra nesse estudo com os oito primeiros princípios. Os autores observaram que apenas os princípios ”completo”, ”acessível” e ”não-discriminatório” estavam em conformidade superior a 50%. Por

outro lado, apenas 4,4% dos *sites* analisados forneciam informações atuais e nenhum apresentou licença livre para uso dos dados. Podemos considerar que em relação as questões analisadas no estudo, no cenário brasileiro, falta qualidade de publicação de dados de execução do orçamento público.

fonte:Craveiro et al., 2013[4]

Figura 1 – Publicação de orçamento no Brasil como DGA

Nota-se que a publicação da execução do orçamento no Brasil precisa avançar para atender aos princípios e leis de DGA. Alguns desses princípios, como "completo", "primário", "atual" e "não-proprietário", têm correspondência na legislação brasileira, outros não. Portanto, é preciso entender quais as diferenças entre legislação e princípios de DGA para aprimorar as próprias leis de publicação de dados governamentais.

Para se conseguir uniformidade na publicação de dados, padrões são de grande importância para que exista interoperabilidade entre sistemas, tanto que o Governo Federal, através do Comitê Executivo de Governo Eletrônico, publica anualmente a e-PING, um documento que "define um conjunto mínimo de premissas, políticas e especificações técnicas que regulamentam a utilização da Tecnologia de Informação e Comunicação (TIC) na interoperabilidade de serviços de Governo Eletrônico" (MPOG, 2012)[11] e é obrigatório para todos os órgãos do executivo federal.

A interoperabilidade é fator importante na formação de uma nova proposta de Web. Na Figura 2 é mostrada a evolução da Web em quatro fases. Na primeira fase os documentos eram colocados na Web até que, na segunda fase, ficou caracterizada uma Web de Documentos. O movimento por publicação de dados é visto como a fase de dados na Web, para na quarta fase termos uma Web de dados. No entanto, colocar dados na Web por si só não garante que exista um ambiente de interconexão desses dados para que seja possível navegar na Web de dados como na atual Web de documentos, ou seja, visitando outros conteúdos clicando em *hyperlinks* nas páginas Web. São necessários estudos para que o incremento de semanticalidade sobre os dados permitam o processamento por máquina.

fonte: Florian e Martin, 2012[12]

Figura 2 – Evolução da Web

Nessa visão de evolução da Web, publicar documentos significa publicar dados na primeira fase da Figura 2, ou seja, publicar documentos na Web como na década de 90. Dessa forma, publicar informações no formato de documentos PDF significa olhar no sentido contrário a futura Web de Dados. Neste trabalho

são abordados aspectos de padronização de publicação de dados da execução do orçamento público que permita a formação de uma Web de dados com informações da execução de orçamento público, que inclui requisitos legais e outros encontrados na literatura e tecnologias que atendam a esses requisitos.

1.1 Objetivos

Objetivo geral:

- desenvolver um padrão de publicação de dados governamentais abertos da execução do orçamento público.

Objetivos específicos:

- analisar a legislação orçamentária vigente para extrair a estrutura de dados orçamentários;
- selecionar tecnologias Web para publicação de DGA;
- desenvolver formatos de divulgação do orçamento público;

1.2 Justificativa

Publicar dados do orçamento num meio eletrônico de amplo acesso e de forma padronizada pode auxiliar no combate à corrupção, participação democrática, controle social e abrir portas para cidadania, além de aproximar sociedade e governo. Para preservar a autonomia dos entes da federação é necessário um ambiente independente de plataforma, linguagem e fornecedor, e que também permita a colaboração no desenvolvimento entre os entes federativos, diminuindo custos e melhorando a interoperabilidade de sistemas. Investir no processamento por máquina promove melhorias na gestão de um grande volume de informação digital.

1.3 Organização do texto

No Capítulo 2, são apresentados os aspectos legais que devem ser considerados na publicação de dados do orçamento pelos entes da federação. Na literatura são encontrados outros aspectos que devem ser aplicados à publicação de dados e que são apresentados no Capítulo 3. A proposta deste trabalho, projeto, modelo e implementação estão no Capítulo 4, onde ela também é associada a requisitos de publicação extraídos dos Capítulos 2 e 3. No Capítulo 5, é feita a discussão da proposta relacionando-a com as abordagens de publicação encontradas na literatura, além da própria legislação. O Capítulo 6 contém as considerações finais deste trabalho.

Capítulo 2

Arcabouço legal brasileiro

Parte das informações do orçamento público que deve ser publicada à sociedade é definida durante o planejamento do orçamento, no ano anterior a sua execução. Neste capítulo, são abordados aspectos do planejamento onde são definidas as informações que devem ser publicadas durante a execução do orçamento e requisitos de publicação estabelecidos em lei.

Para o planejamento de orçamento público cada órgão de governo é dividido em Unidades Orçamentárias (UO), cada uma com o papel de elaborar proposta orçamentária com suas unidades administrativas dentro do propósito do órgão(MPOG,2012)[13]. Conforme a LRF, os entes da federação (União, Estados, Distrito Federal e Municípios) para cada ano devem registrar o resultado do planejamento do orçamento público na forma de lei, chamada de Lei Orçamentária Anual(LOA), que compreenderá receitas e despesas de todas UOs conforme classificação orçamentária prevista em lei; tal classificação é melhor abordada no Capítulo 2 entre outros aspectos legais. Com a Lei Capiberibe todos os detalhes da execução do orçamento devem ser publicados por todos os entes da federação em meios eletrônicos de acesso público, no caso a Web.

A LOA é o documento final de um processo de discussão que, no caso da Administração Pública Federal(BRASIL, 2013)[13], está exemplificado na Figura 3 com a linguagem de processo de negócios BPMN¹. Observa-se que na Administração Pública Federal, cabe à Secretaria do Orçamento Federal(SOF), estudar e comparar limites, definir estratégias, ajustar propostas e consolidar o projeto de lei orçamentária. Os órgãos setoriais, como os ministérios, estabelecem diretri-

¹<http://www.bpmn.org/>

zes para que suas Unidades Orçamentárias elaborem os programas e as ações do orçamento, também chamada de Estrutura Programática do projeto de lei. Feita essa estrutura a UO formaliza a proposta que será enviada ao Órgão Setorial para consolidação e validação, posteriormente à SOF para ajustes na proposta, antes da decisão da Presidência da República(PR) e do Ministério Público(MP) enviarem o projeto de lei ao Congresso. Decidido pelo envio, o projeto entra na SOF para ajustes finais e é enviado ao Congresso pela PR/MP.

Processos similares acontecem nos outros entes federativos, sempre resultando numa LOA, e é durante esse planejamento que são estabelecidas algumas das informações que deverão ser publicadas para sociedade em páginas de transparência como, por exemplo, a classificação dos gastos públicos por função: saúde, educação, previdência, etc. Por isso, faz-se importante conhecer o orçamento público.

Os aspectos legais relacionados à publicação de informações da execução do orçamento público à sociedade, constantes na Constituição de 1988 até o Decreto 7.724 de 2012, que regulamenta a Lei nº 12.527, conhecida Lei de Acesso à Informação, são apresentados em ordem cronológica. Desse período são extraídos elementos que devem ser apresentados para a sociedade de forma automatizada na rede mundial de computadores, a Internet.

No Art. 5º da Constituição de 1988[14] é estabelecido que todos têm direito de receber informações dos órgãos públicos, sejam de interesse particular ou coletivo, nos prazos previstos em lei, ressalvadas as informações relativas à segurança da sociedade e do Estado. Se não houver essa prestação de informação o órgão poderá ser penalizado. No Art. 37, o acesso dos usuários aos registros administrativos e de gestão devem ser estabelecidos em lei. Conforme Art. 216, a documentação governamental parte do patrimônio cultural brasileiro, que inclui bens materiais e não-materiais.

Em 2000, é publicada a Lei Complementar 101(Brasil, 2000)[9], chamada Lei de Responsabilidade Fiscal (LRF), que estabelece regras para a gestão da execução do orçamento público no Brasil e a responsabilização da administração pública so-

Figura 3 – *Elaboração de Projeto de LOA na Administração Pública Federal*

bre os atos dessa execução. É em 2009 que a essa lei é acrescida a responsabilidade de publicação eletrônica, pela Lei Complementar 131 (Brasil, 2009)[7], conhecida como Lei Capiberibe, onde se estabelece, no Art. 48, que a transparência sobre os planos orçamentários deve ser assegurada mediante "liberação ao pleno conhecimento e acompanhamento da sociedade, em tempo real, de informações pormenorizadas sobre a execução orçamentária, em meios eletrônicos de acesso público", por todos os entes da Federação em sistema integrado de administração financeira

que atenda a um padrão mínimo de qualidade estabelecido na mesma lei.

Na LRF as informações mínimas a serem publicadas de todos os atos praticados por unidades gestoras do entes da Federação quanto à despesa são: número do processo, bem ou serviço prestado, pessoa física ou jurídica beneficiária e procedimento licitatório quando for o caso, e quanto à receita, lançamento e recebimento, inclusive de recursos extraordinários; adiante, mais explicações sobre informações de despesa e receita. Na LRF também está estabelecido o prazo máximo para todos se adaptarem à publicação de dados, que se encerrou em 27 de maio de 2013.

No Art. 2º do Decreto nº 7.185 de 2010 (Brasil, 2010)[15] foi estabelecido um padrão mínimo de qualidade de sistema integrado e ratificada a "liberação em tempo real das informações pormenorizadas sobre a execução orçamentária e financeira das unidades gestoras, referentes à receita e à despesa" de todas entidades da administração direta, as autarquias, as fundações, os fundos e as empresas estatais dependentes. Também é estabelecido como tempo real a publicação até um dia útil subsequente ao registro contábil e a Internet como o meio eletrônico de amplo acesso. Até esse decreto a Internet era subentendida como um meio de publicação. Conforme Art. 4º, o sistema integrado deve permitir importação e exportação de dados e ter mecanismos que permitam verificar a integridade, confiabilidade e disponibilidade da informação registrada e exportada.

Alguns aspectos mais técnicos são abordados nesse decreto, pois se deve atender preferencialmente a e-PING(MPOG, 2012)[11], um conjunto de padrões de interoperabilidade do Governo Eletrônico, e a e-MAG(MPOG, 2011)[5], um conjunto de recomendações para acessibilidade dos *sites* e portais do governo brasileiro e, no Art. 6º, "aplicar soluções tecnológicas que visem simplificar processos e procedimentos de atendimento ao cidadão e propiciar melhores condições para o compartilhamento das informações".

No Decreto 7.185, Art. 7º, também são mais discriminadas as informações da execução do orçamento que devem ser publicados:

1. quanto à despesa:

- (a) o valor do empenho, liquidação e pagamento;
- (b) o número do correspondente processo da execução, quando for o caso;
- (c) a classificação orçamentária, especificando a unidade orçamentária, função, subfunção, natureza da despesa e fonte dos recursos que financiaram o gasto;
- (d) a pessoa física ou jurídica beneficiária do pagamento, inclusive nos desembolsos de operações independentes da execução orçamentária, exceto no caso de folha de pagamento de pessoal e de benefícios previdenciários;
- (e) o procedimento licitatório realizado, assim como sua dispensa ou inexigibilidade, quando for o caso, com o número do correspondente processo;
- (f) o bem fornecido ou serviço prestado, quando for o caso;

2. quanto à receita, os valores de todas as unidades gestoras, compreendendo no mínimo sua natureza, relativas a:

- (a) previsão;
- (b) lançamento, quando for o caso; e
- (c) arrecadação, inclusive referente a recursos extraordinários.

Na Lei de Acesso à Informação (LAI), como é conhecida a Lei nº 12.527 de 2011(Brasil, 2011)[16], e no Decreto 7.724 de 2012 (Brasil, 2012)[17], que regulamenta a LAI, são estabelecidos que os sítios governamentais na Internet devem conter ferramenta de pesquisa de conteúdo, possibilitar a gravação de relatórios em diversos formatos eletrônicos não-proprietários, possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina, divulgar em detalhes os formatos utilizados para estruturação da informação, e utilizar linguagem de fácil compreensão no fornecimento de informação primária, íntegra, autêntica, atualizada, de forma gratuita e independente de requerimentos

no caso de repasses, transferências ou despesas. Outros conceitos sobre informação também são encontrados na LAI, Art. 4º, onde destacamos:

- informação: dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato;
- disponibilidade: qualidade da informação que pode ser conhecida e utilizada por indivíduos, equipamentos ou sistemas autorizados;
- autenticidade: qualidade da informação que tenha sido produzida, expedida, recebida ou modificada por determinado indivíduo, equipamento ou sistema;
- integridade: qualidade da informação não modificada, inclusive quanto à origem, trânsito e destino;
- primariedade: qualidade da informação coletada na fonte, com o máximo de detalhamento possível, sem modificações.

O orçamento é instrumento de planejamento de qualquer entidade pública ou privada e representa o fluxo de receitas e despesas de recursos em determinado período (MPOG, 2012)[13]. Durante a execução do orçamento público a receita pode estar em três fases: lançamento, arrecadação e recolhimento; respectivamente, a inscrição de débito a pessoa física ou jurídica devedora, o pagamento desse débito em agentes arrecadadores ou instituições financeiras autorizadas, e a transferência do recurso para a entidade pública. No Decreto 7.185 também é colocada a previsão de receita, uma fase que antecede o lançamento por estar no planejamento orçamentário, mas não o recolhimento onde se espera ser igual à arrecadação por ser uma formalidade financeira da transferência do recurso do agente arrecadador para o ente federativo.

2.1 Classificação orçamentária

Para identificar detalhadamente as receitas é dada uma classificação da natureza da receita, com seis números, exemplificada na Figura 4, ou seja, todo imposto de renda sobre pessoa física é lançado, arrecadado e recolhido sob o código 1.1.1.2.04.10 que correspondem, respectivamente, à estrutura hierárquica categoria econômica, origem, espécie, rubrica, alínea e subalínea².

fonte: MPOG, 2012[13]

Figura 4 – Exemplo de classificação de receita

Além da classificação por natureza da receita, existem também as classificações por indicador de resultado primário e por fonte/destinação de recurso (MPOG, 2012)[13]. Quando uma receita é incluída na "apuração do resultado primário (diferença entre as receitas e as despesas primárias)", é classificada com a letra "P" como indicador de resultado primário, caso contrário com "F", se não primária ou financeira. A classificação por fonte/destinação de recurso foi criada para "assegurar que receitas vinculadas por lei a uma finalidade específica sejam exclusivamente aplicadas em programas e ações que visem a consecução de despesas ou políticas

²além do Manual do Orçamento Técnico[13], detalhes sobre a legislação orçamentária são encontrados no site do Tesouro Nacional: http://www3.tesouro.gov.br/contabilidade_governamental

públicas associadas a esse objetivo legal” (MPOG, 2012)[13] e também utilizada na despesa.

CÓDIGO COMPLETO*		10.	39.	252.	26.	782.	2075.	7M64.	0043.	9999.	0.	100.	4490.	2
Q U A L I T A T I V A	Esfera: Orçamento Fiscal	10												
	CLASSIFICAÇÃO INSTITUCIONAL		39											
	Órgão: Ministério dos Transportes													
	Unidade Orçamentária: Departamento Nacional de Infraestrutura de Transportes - DNIT			252										
	CLASSIFICAÇÃO FUNCIONAL				26									
	Função: Transporte					782								
	Subfunção: Transporte Rodoviário													
Q U A L I T A T I V A	PROGRAMA: Transporte Rodoviário						2075							
	AÇÃO: Construção de Trecho Rodoviário							7M64						
	SUBTÍTULO: Rio Grande do Sul								0043					
Q U A N T I T A T I V A	IDOC: Outros recursos									9999				
	IDUSO: Recursos não destinados à contrapartida										0			
	Fonte de Recursos: Recursos do Tesouro - Exercício Corrente (1) Recursos Ordinários (00)											100		
	Natureza da Despesa: Categoria Econômica: Despesas de Capital (4); Grupo de Natureza: Investimentos (4); Modalidade de Aplicação: Aplicação Direta (90)												4490	
	Identificador de Resultado Primário: Primária Discrecional													2

*Código visualizado no SIAFI.

fonte: MPOG, 2012[13]

Figura 5 – Exemplo de classificação de despesa

A classificação da despesa é dividida em programação qualitativa e quantitativa, compondo uma sequência de treze códigos alfanuméricos. Na Figura 5 está exemplificada uma classificação de despesa relativa à construção de um trecho rodoviário, onde a fonte de recurso é "Recursos do Tesouro - Exercício Corrente(1), Recursos Ordinários (00)", código "100", que é o mesmo utilizado nas receitas do Tesouro Nacional. Uma despesa é dividida em uma fase de planejamento e três de execução, sequencialmente, ela é orçada, empenhada, liquidada e paga. Despesa empenhada é o valor que o Estado reservou para efetuar um investimento planejado no orçamento do ano de sua execução. O empenho ocorre após a assinatura de um contrato para aquisição de bens ou prestação de serviço. Nesse caso, quando o bem for entregue ou o serviço prestado, o valor é liquidado e, quando o beneficiário

de fato receber o valor, ele é considerado valor pago.

2.2 Requisitos de publicação

Com base na legislação apresentada na Tabela 1, reunimos os requisitos legais que devem ser aplicados à publicação de dados da execução do orçamento.

Tabela 1 – Requisitos identificados na legislação

Legislação	Requisitos
Lei de Responsabilidade Fiscal(2000) e Lei Capiberibe(2009) Decreto 7.185(2010)	informações pormenorizadas publicadas em tempo real em meios eletrônicos <ul style="list-style-type: none"> • tempo real equivale a um dia útil; • publicação na Internet; • verificação de integridade, confiabilidade e disponibilidade da informação; • atender a e-PING e a e-MAG; • da despesas publicar empenho, liquidação, pagamento, número de processo, classificação orçamentária, beneficiário, procedimento licitatório e bem ou serviço fornecido; • da receita publicar previsão, lançamento e arrecadação.
Lei de Acesso à Informação(2011) e Decreto 7.724(2012)	<ul style="list-style-type: none"> • ferramentas de pesquisa; • gravar relatórios em diversos formatos não proprietários; • possibilitar acesso automatizado; • utilizar formatos abertos legíveis por máquina; • fornecer informação primária, íntegra, autêntica, atualizada e gratuita.

Capítulo 3

Dimensões conceituais sobre dados

Segundo Houaiss(2009)[18], no âmbito da informática *dado* é informação que pode ser processada por computador. Este significado encontrado entre outros catorze apresentados pelo dicionário leva diretamente a dois desdobramentos, que dado é informação e processável por computador. No Capítulo 2 foram apresentadas as leis relacionadas a publicação de dados da execução de orçamento público na Internet, ou seja, a dimensão legal apresentada na forma de requisitos de publicação.

Além da legislação brasileira, outras abordagens de publicação de dados apontam requisitos que devem ser considerados para o caso brasileiro. Alguns desses requisitos convergem com a lei, outros a complementam. Neste capítulo, são elencados conceitos que agregam requisitos de publicação de dados encontrados na literatura, bem como pontos de convergência com a lei quando aplicável. Tais conceitos configuram outras dimensões sobre dados por agregarem mais de um item que pode ser verificado na publicação de dados.

3.1 Dados Governamentais Abertos

As informações podem ser difíceis de achar, estar em formatos complicados e inacessíveis a mecanismos de buscas eletrônicas na Internet. Neste contexto de disponibilidade de dados do governo, o termo de Dados Governamentais Abertos foi cunhado, e se tornou popular no início de 2008, após a publicação de um conjunto

de princípios por ativistas nos EUA em dezembro de 2007 (OGWG, 2007)[19], que devem ser aplicados em todos os dados produzidos pelo setor público, pois esses princípios visam assegurar o direito à informação. São eles:

1. Completo: Todos os dados públicos são disponibilizados. Dados públicos são dados que não estão sujeito a limitações válidas de segurança, privacidade ou privilégio.
2. Primário: dados são coletados como na fonte, com o maior nível possível de granularidade, não em formas agregadas ou modificadas.
3. Atual: dados são disponibilizados tão rapidamente quanto necessário para preservar os seus valores.
4. Acessível: dados estão disponíveis para a mais ampla gama de usuários, para as mais diversas finalidades.
5. Processável por máquina: dados são razoavelmente estruturados para permitir processamento automatizado.
6. Não-discriminatório: dados estão disponíveis a qualquer pessoa, sem necessidade de cadastramento.
7. Não-proprietário: dados são disponíveis em formato sobre o qual nenhuma entidade tem controle exclusivo.
8. Licença-livre: dados não estão sujeitos a qualquer direito autoral, patente, marca ou a regulação do comércio segredo.

No contexto da legislação brasileira (Brasil, 2009)[7], dados completos de orçamento público significa: publicar informações de execução de receitas e despesas sem incluir a etapa anterior de planejamento do orçamento de forma detalhada, que não significa exatamente informação primária. E dado atual significa ser publicado até um dia útil após o registro contábil (Brasil, 2010)[15]. Em 2010 outros princípios foram adicionados, totalizando dez princípios de DGA (SF, 2010)[20]:

9. Permanente: estar disponível ao longo do tempo se refere aos restos pois, muitas vezes, as informações são atualizadas, alteradas ou removidas sem qualquer indicação, nem a devida indicação de versão.
10. Custo de uso: uma grande barreira de acesso às informações é o custo imposto ao público para acesso à informação, esse deve ser o mínimo possível.

Além dos princípios também surgiram três leis de DGA, em 2009: 1) se o dado não pode ser encontrado na web e indexado, ele não existe; 2) se não estiver aberto e disponível em formato processável por máquina, ele não pode ser utilizado; 3) se qualquer dispositivo legal não permitir que ele seja reutilizado, não é útil. A primeira lei implicitamente trata da disponibilidade e descoberta dos dados. As informações devem estar na rede mundial de computadores, portanto, acessíveis de qualquer parte do planeta, e precisam estar num formato que permita indexação, para facilitar sua leitura pelos mecanismos de busca. Se o cidadão não encontrar os dados de maneira simples é o mesmo que eles não estarem disponíveis. A segunda lei, corrobora com a primeira mas focando a tecnologia. Para o dado ser encontrado e indexado ele deve estar num formato adequado ao uso nas tecnologias da Web. Se os computadores não puderem tratar os dados, eles não aparecerão nos mecanismos de busca nem serão mostrados nas máquinas aos cidadãos, portanto, ficariam inacessíveis e seria o mesmo que não existirem. A terceira lei trata do âmbito jurídico. Se informações sobre gastos públicos não podem legalmente ser utilizadas pelos cidadãos para questionar a atuação do governo, se o cidadão for obrigado a ter uma postura passiva em relação às informações governamentais, não existe participação e as informações se tornam inúteis.

Os princípios *Completo*, *Primário*, *Processável por máquina* e *Não-proprietário* podem ser atendidos serializando os dados no formato XML ou XBRL. Os princípios *Atual* e *Licença-livre* são atendidos anexando dados de proveniência nos metadados, o princípio *Não-discriminatório* não está relacionado diretamente com a publicação dos dados, mas a uma decisão política de não exigência de cadastro para acesso a eles.

Sheridan e Tennison (2010)[21] relatam sobre a experiência de publicação de dados abertos do Governo do Reino Unido, que novos formatos como o RDF são impenetráveis sem programas especiais, tornando difícil o uso de dados por não-especialistas. Por isso, esse governo provê visões nos formatos JSON e XML dos dados[21]. Nesse contexto, a linguagem eXtensible Stylesheet Language for Transformation (XSLT)¹ é uma tecnologia que permite converter padrões baseados na linguagem XML, o que inclui o XBRL, para qualquer outro formato baseado em texto.

3.2 Qualidade de dados

Segundo Madnick *et al.*(2009)[22], existe uma tendência em se usar *qualidade de dados* para questões técnicas e *qualidade de informação* para se referir a questões não-técnicas sobre os dados. Os autores dividiram sua revisão sobre pesquisas em qualidade de dados em quatro categorias principais e subcategorias, de onde destacamos a subcategoria *Sistemas de Informação* dentro de *Contexto de Ciências de Computação e Tecnologia de Informação*. Nas pesquisas em sistemas de informação, identificam-se questões de qualidade de dados nas organizações, práticas que aumentam ou deterioram a qualidade dos dados, técnicas e soluções para a gestão da qualidade de dados em ambientes organizacionais.

Nessa perspectiva, os princípios de DGA são questões que devem ser consideradas como requisitos numa publicação que vise assegurar o direito à informação ao sociedade, assim, não cumprir os princípios deterioram esse direito. Analogamente, cumprir tecnicamente os requisitos legais sumarizados na Tabela 1 na publicação na Internet aumenta a qualidade de dados num sistema de informação nas questões legais.

Nessa linha de pensamento, na proposta deste trabalho, qualidade de dados é atender tecnicamente na publicação de dados a questões não-técnicas, como

¹A especificação do padrão XSLT é encontrada em <http://www.w3.org/TR/xslt>

legislação, princípios de DGA e outros requisitos encontrados na literatura.

3.3 Metadados

Metadados são informações que facilitam a busca, classificação, utilização dos dados, ou seja, são de vital importância para o reuso dos dados no futuro (NISO, 2004:1)[23]. Também são tratados como dados que explicam outros dados. Zuiderwijk *et al.* (2012) [24] afirmam que a ligar dados entre si não é suficiente, metadados são obrigatórios para se fazer uso efetivo dos dados, pois também melhoram a capacidade de encontrar padrões para criar novas hipóteses.

Metadados também podem ser divididos em níveis como coleção (usados para agrupar dados em grupos formando coleções), variável (Powell, 2009)[25], descoberta (flat), contextual e detalhado (Zuiderwijk *et al.*, 2012)[24], além de conter outras informações de apoio, tais como licença, atualização e proveniência. Algumas organizações padronizam conjuntos de elementos de metadados como vocabulários controlados, por exemplo, Dublin Core, TEI, METS e VRA (NISO, 2004:3-5)[23]. O padrão de arquivo CSV suporta apenas a descrição de nomes de campos, não sendo adequado para serialização de metadados, salvo se os metadados estiverem num arquivo diferente.

Segundo Powell(2009:89)[25], o padrão de metadados Dublin Core tem uma ótima amplitude de elementos de metadados e documentação relativamente simples, permitindo sua aplicação imediata, além de ter como grande vantagem a experiência e o conhecimento de centenas de profissionais que o criaram. No Brasil, o VCGE é o vocabulário a ser usado no Padrão de Metadados do Governo Eletrônico (e-PMG) (MPOG, 2010)[26], esse também baseado no padrão Dublin Core, no campo "*assunto.categoria*". Percebe-se que há dois níveis de vocabulários controlados no Governo Eletrônico: um que descreve os metadados e outro que delimita os valores para um campo dos metadados.

3.4 Taxonomia

Taxonomia é o nome dado a qualquer sistema de classificação criado para agrupar elementos com características semelhantes (Campos e Gomes, 2010)[27]. Uma das taxonomias mais conhecidas é a classificação das espécies. A classificação do orçamento público pode ser representadas numa taxonomia, que pode ser obtida mediante a análise de documentos para a elaboração de uma estrutura classificatória que se pauta na categorização, formação e ordenação de classe de conceitos (Campos e Gomes, 2010)[27].

Segundo (Campos e Gomes, 2010)[27], para a elaboração de uma taxonomia algumas ações devem ser tomadas:

1. Captura do conhecimento – que pode se dá mediante entrevistas com especialistas da área, averiguação de documentos existentes na instituição, de outros instrumentos classificatórios ou terminológicos;
2. Análise dos documentos/informações que serão agregados à taxonomia – adequação das informações existentes nos acervos à terminologia apresentada na taxonomia;
3. Elaboração da estrutura classificatória da taxonomia – que deve se pautar nos princípios teóricos de categorização, de formação e de ordenação das classes de conceitos;
4. Validação – apoia-se como base para a certificação da proposta classificatória, visando atender às necessidades da comunidade para a qual se destina. Esta ação permitirá que possamos atingir critérios de comunicabilidade, estimulação e compatibilidade.

Para publicação de dados financeiros, a eXtensible Business Reporting Language (XBRL) é um padrão que pode ser usado para divulgação de informações financeiras pela Internet (Riccio et al, 2006)[28]. Silva (2004)[29] relata que na

experiência do Banco Central do Brasil, que XBRL tornou o intercâmbio de informações mais eficiente, facilitando análises e tomadas de decisão. XBRL também foi adotado pela Secretaria do Tesouro Nacional no projeto SICONFI (Brasil, 2013), que é um sistema de informações contábeis e fiscais.

XBRL foi criado a partir dos padrões W3C XML, XML Schema, XML Namespace, Xlink (XBRL International, 2012)[30]. Também foram considerados para o XBRL padrões que favorecem a internacionalização dos relatórios, como é o caso do ISO 41272 para nomenclatura de moedas correntes no mundo. Uma vantagem de utilizar XBRL é que ele permite a inserção de semântica de um domínio financeiro que pode ser processável por máquina, facilitando no processamento de dados, além da capacidade de representar taxonomias serializadas no formato XML Schema.

3.5 Vocabulários controlados

Vocabulário controlado é uma linguagem artificial elaborada para padronizar e facilitar entradas e saídas de dados de sistemas de informações (Kobashi, 2008:1)[31]. Sem seu uso o grande volume e complexidade da estrutura governamental pode tornar a busca por informações uma tarefa difícil, até mesmo impossível para os cidadãos(MPOG, 2011:4)[32]. Além disso, vocabulários controlados são uma forma de ajudar os cidadãos na busca de informações, pois "alterações na semântica de termos em vocabulários dados ligados podem ter uma influência dramática sobre a interpretação de conjuntos de dados" (Käfer *et al.*, 2012) [33].

Em geral os vocabulários controlados são apresentados com um esquema temático global (macroestrutura) abrangendo as categorias mais amplas dos campos do conhecimento, e com unidades dos vocabulários – os descritores(Kobashi, 2008:2)[31]. No e-PING[11] do Governo Eletrônico do Brasil há um vocabulário controlado para uso em campos de assunto chamado Vocabulário Controlado de Governo

Eletrônico(VCGE)², como exemplo, nesse vocabulário existem os descritores *Educação indígena* e *Política trabalhista*.

3.6 Proveniência de dados

Hartig (2009)[34] desenvolveu um modelo de proveniência de dados Web, que inclui duas dimensões - criação de dados e acesso a dados - a fim de "apoiar a avaliação das qualidades de dados, tais como precisão, confiabilidade e pontualidade". Este modelo distingue-se modelo de proveniência aplicado a banco de dados para incluir o acesso a dados de informações, como o tipo de serviço de acesso, por exemplo, um servidor ou Serviço Web.

É importante ressaltar que um dado é a forma serializada da informação previamente obtida. Essa informação, no caso de uma pesquisa de campo, pode ser coletada em um questionário seguindo métodos estabelecidos antes da coleta. Esse questionário é um dos elementos da proveniência de informação em sua gênese. O modelo de Hartig trata da proveniência dos dados publicados na Web e não retorna à história da informação.

Dados de proveniência podem ser representados nos elementos de metadados, sendo a URL um dos elementos da proveniência de acesso, por ser o endereço eletrônico de acesso aos dados.

A Open Knowledge Foundation (2011)[35] citou a importância de se tornar os dados descobríveis: "postar na web e, talvez, organizar um catálogo central para listar seus dados abertos". Este catálogo terá a função de uma pesquisa de referência para os dados, facilitando a descoberta por parte dos cidadãos. Assim, como o sistema de Código de Endereçamento Postal no Brasil, é necessário que os dados estejam inseridos numa sistemática de identificador único para que seja possível construir um catálogo desses dados.

²o VCGE é encontrado no endereço <http://vocab.e.gov.br/2011/03/vcge>

Na Web a forma de endereçar locais é o Uniform Resource Locator(URL), um tipo de Uniform Resource Identifier(URI)³, ou seja, é o endereço eletrônico utilizado pelos usuários para acessar as páginas eletrônicas.

Mas uma URL em si não é suficiente. É preciso garantir que os endereços não mudem para que os dados sejam sempre encontrados e o catálogo não se torne inexpressivo. Esse conceito é chamado de Persistent URL(PURL). Um exemplo de domínio que implementa PURL é o <http://purl.pt>, gerenciado pela Biblioteca Nacional de Portugal, ou seja, um órgão teoricamente permanente é mantenedor da URLs permanentes.

O governo do Reino Unido desenvolveu um conjunto de URLs para identificar objetos do mundo real como escolas e rodovias, documentos, vocabulários e dados (Sheridan e Tennison, 2010)[21], dada a importância de se ter endereços persistentes.

3.7 Versionamento de dados

”Alterações na semântica de termos nos vocabulários de dados ligados podem ter uma influência dramática na interpretação sobre conjuntos de dados que os usam”(Käfer *et al.*, 2012) [33]. Por isso, é importante apontar qual versão do vocabulário está sendo usada nos dados publicados.

Indicar a versão de vocabulários é uma necessidade transversal aos padrões de vocabulários controlados, metadados e taxonomias. As organizações que criam esses padrões, conforme a necessidade, alteram-nos no tempo quando necessário, sendo possível criar, mudar e até mesmo excluir elementos dos padrões. Uma prática comum na Web é utilizar um URL permanente com a indicação da versão no próprio endereço, como a versão 1.1 do conjunto de metadados Dublin Core no endereço <http://purl.org/dc/elements/1.1/>, a versão 2003 de instâncias XBRL

³a especificação de URL e URI é encontrada no endereço <http://tools.ietf.org/html/rfc3305>

no endereço <http://www.xbrl.org/2003/xbrl-instance-2003-12-31.xsd>, que inclusive contém a data de publicação no próprio endereço.

Trabalhar com padrões diferentes, cada um com sua versão, exige que o formato de serialização suporte trabalhar com termos de domínios distintos. O formato XML suporta a combinação de múltiplos vocabulários pelo mecanismo chamado XML Namespace, conseqüentemente o XBRL também pode ser utilizado para esse fim.

3.8 Validação de dados

Validar dados significa compará-los a alguma declaração sobre o que esses dados devem ser. Portanto, é necessário um conhecimento prévio sobre eles, o que são e como devem estar organizados, ou seja, quais os tipos dos dados e sua estrutura. Para uma validação automática de dados é necessário que a declaração sobre eles também seja processável por máquina.

O sucesso da Web é derivado do desenvolvimento do HTML como uma estrutura para apresentação visual, mas na cultura de bancos de dados é utilizado um vocabulário de esquemas e diagramas entidade e relacionamentos para descrever estruturas de dados (Abiteboul, 2000:3)[36]. XML representa um primeiro passo na convergência dessas duas visões de estrutura de informação, que mesmo sendo sintaticamente relacionado ao HTML tem o propósito de transmitir uma estrutura de dados (Abiteboul, 2000:4)[36].

Mesmo existindo uma estrutura implícita à própria serialização de dados em XML, informações como tipos de dados, obrigatoriedade de elementos de dados e hierarquia, não são declaradas nessa serialização. Para tais declarações sobre a estrutura de dados existe o XML Schema⁴, uma linguagem para definição de regras de validação de dados no formato XML, onde se pode declarar nome de

⁴detalhes sobre a especificação XML Schema são encontrados em <http://www.w3.org/XML/Schema>

elemento, tipo de dado, número mínimo e máximo de ocorrência dos elementos, além da própria estrutura. XML Schema é opcional nos documentos XML, mas é obrigatório na especificação XBRL, sendo chamado de Taxonomia XBRL.

3.9 Linked Data

A publicação de dados na Web, por mais que cresça em volume, não forma uma infraestrutura de informação correlacionada entre si como um grande ecossistema de dados. Para formar a chama Web de Dados, as informações devem estar relacionadas umas às outras como o conjunto de *sites* visitados pelos usuários e interligados entre si. Este capítulo trata dos aspectos relacionados a essa Web de Dados, como eles devem ser publicados para possibilitar o processamento por máquina e facilitar a extração de informações num grande volume de dados.

Na Figura 6 é mostrado como uma página *web* tradicional é interpretada por programas navegadores (à esquerda) e por humanos (à direita). Os programas interpretam a página como um conjunto de elementos gráficos sem identificar a natureza da informação, se é um título, nome próprio ou palavra-chave. Um humano, como sugerido na figura, pode diferenciar entre os *links* o que é palavra-chave e o que é licença de Copyright.

Observou-se na Figura 2 que ao caminharmos para uma Web de Dados há um aumento na "semanticalidade" da Web. Para a evolução da Web, uma alternativa é representar o conteúdo numa forma mais fácil de processamento por máquina e usar técnicas inteligentes para aproveitar as vantagens desse tipo de representação, constituindo assim uma Web Semântica, envolvendo gradualmente a Web existente nessa nova abordagem (Antoniou e Harmelen, 2008)[37]. "As tecnologias da Web Semântica podem promover um enorme desenvolvimento no modo como se pensa e se usa a Internet" (W3C, 2009)[10].

Uma abordagem dessas tecnologias, o Linked Open Data (LOD), é um movimento em crescimento que vai habilitar os usuários em criar e combinar conjuntos

fonte: W3C Brasil, 2009[10]

Figura 6 – Interpretação de página web por programas navegadores e humanos

de dados para construir suas próprias interpretações de dados disponibilizados em formatos compreensíveis por máquina (Florina e Martin, 2012)[12]. Os princípios de dados abertos suportam esse movimento ao incluir que os dados devam ser interpretáveis por máquina, ou seja, esses princípios preparam a atual fase para futura ligação de dados.

Bernes-Lee em 2009[38], desenvolveu um modelo de LOD em cinco níveis (chamados de estrelas) para classificar progressivamente a facilidade de uso por consumidores de dados. Esse modelo é uma classificação de referência de dados abertos na abordagem LOD e complementa os oito princípios de DGA.

1. disponível na Web (qualquer formato), mas com uma licença aberta;
2. como o item 1 e disponível como dados estruturados legíveis por máquina;
3. como 2, e em formato não-proprietário;

4. como 3, com uso de padrões abertos da W3C para identificar as coisas, para poderem apontar para seus dados;
5. todos anteriores, mais vincular os seus dados com os de outras pessoas para fornecer contexto.

Os três primeiros níveis, ou estrelas, referem-se a *Licença-livre*, *Processáveis por máquina* e *Não-proprietário* dos princípios de DGA, mais três questões são adicionados por este modelo: a definição que a via de acesso é a Web, a criação de identificador único para conjuntos de dados, e vínculo como outros dados mediante seus identificadores. Na Figura 7, estão exemplificadas algumas tecnologias que podem estar associadas a cada uma dessas estrelas, portanto nesse modelo, publicar dados no formato PDF é classificado com uma estrela, e o formato não-proprietário CSV como três estrelas.

fonte: <http://5stardata.info/>

Figura 7 – Modelo de 5 estrelas de Linked Open Data

3.10 Trabalhos relacionados

Nos trabalhos apresentados nesta seção encontram-se relações entre informações contábeis, dados abertos e Linked Data e o padrão XBRL. Tais relações abrem caminho para uma proposta de serialização de dados que inclua esse padrão para atender aos requisitos levantados no Capítulo 2 e neste capítulo.

O’Riain *et al.* (2012)[39] investigaram a possibilidade de utilizar XBRL num ambiente mais amplo de dados abertos e como XBRL pode ser integrado na abordagem Linked Data. Para eles tanto XBRL quanto Linked Data podem ser tratados como *arquiteturas customizáveis* para ajudar na interpretação e integração de informações. Taxonomias XBRL podem ser usadas como recursos lexicais de domínio para ajudar a interpretar conteúdo não-estruturados como os casos de uso na indústria e na academia. Como parte de um ecossistema de dados mais amplo em XBRL, também deve ser considerado o uso URIs (identificadores únicos) para identificação de recursos financeiros na Web.

Na Figura 8 é mostrada uma proposta de arquitetura de publicação de dados por Garcia e Gil (2009)[40]. Eles afirmam que é possível mapear dados em XML para documentos em XBRL e conseqüentemente gerar dados semânticos no formato RDF, preservando toda informação e sua estrutura. Eles utilizaram ferramentas informatizadas para importar os documentos XBRL para um ambiente na abordagem Linked Data, a plataforma Rhizomer, de onde se pode gerar visualizações em XHTML, ou seja, visível em navegadores Web.

Fernandes e Campos (2008)[41] propuseram uma central de balanços no Brasil, que agregaria demonstrações contábeis das empresas nacionais e permitiria estudos setoriais, análise de risco e geração de estatísticas. Na figura 9 são representadas as possibilidades técnicas com o uso de XBRL: validar um documento a partir de um esquema (taxonomia), publicar nos formatos HTML, PDF e DOC, mostrar as relações entre conceitos e recursos chamados *linkbases*. Como explicam, cada *linkbase* no XBRL é utilizado para acrescentar informação à taxonomia (Fernan-

fonte: Garcia e Gil, 2009

Figura 8 – XBRL publicado como *Linked Data*

des e Campos, 2008:8)[41]. Como exemplo, no *linkbase calculation* são definidas relações de cálculos entre os conceitos e no *linkbase reference* são feitas relações entre conceitos e elementos como atos normativos, legais e até páginas na Web.

3.11 Requisitos de publicação

Na literatura consultada foram levantados conceitos que se configuram em requisitos que devem ser considerados na publicação de dados abertos. A Figura 10 contém os requisitos levantados na literatura consultada onde alguns são suportados por mais de um autor, como o princípio *não-proprietário* que é apontado por OGWG[19] e Berners-lee[38]. E existem correlações entres requisitos que não estão demonstradas na Figura 10, por exemplo, o requisito *Dados devem ser acessíveis*[19] está associado aos requisitos *Dados devem estar disponíveis na Web*[38] e *Deve-se prover alternativas de acesso*[21], e atender a esses dois últimos infere em atender ao primeiro. Essas correlações são discutidas no Capítulo 5.

fonte: Fernandes e Campos, 2008

Figura 9 – Estrutura de validação e publicação de documentos XBRL

Os requisitos da Figura 10 junto aos requisitos da Tabela 1 formam o conjunto de requisitos que são considerados na proposta de padrão de publicação processável por máquina deste trabalho.

fonte: elaborado pelo autor

Figura 10 – Requisitos para publicação de dados na literatura

Capítulo 4

Proposta de publicação

Nos Capítulos 2 e 3 foram abordados questões legais e conceituais que devem ser consideradas na proposta de publicação de dados de execução orçamentária. Publicar dados pressupõe a existência de um padrão para serialização de dados que atenda a requisitos previamente estabelecidos. Os requisitos legais de publicação em alguns pontos convergem com os conceitos apresentados no Capítulo 3, em outros os complementam e vice-versa. A proposta deste trabalho é desenvolver um padrão de publicação que atenda à legislação, às abordagens de DGA e LD, e outros aspectos que podem ser considerados requisitos para publicação de dados.

Neste capítulo são apresentados requisitos levantados nos Capítulos 2 e 3, metodologia de desenvolvimento, decisões de projeto, modelo da proposta, seleção de dados e como as tecnologias podem atender aos requisitos levantados.

4.1 Metodologia

Este estudo é uma pesquisa exploratória das exigências legais no Brasil em relação à publicação dos dados de execução de orçamento público na Web. Paralelamente, outra pesquisa foi realizada a partir de documentos sobre os princípios de Dados Governamentais Abertos e Linked Data e suas referências, na busca de requisitos aplicáveis à publicação de dados abertos. Foi notado que alguns artigos que traziam conceitos relevantes eram publicados no Linked Data on the Web Workshop. Assim optamos por verificar todos os artigos a partir da primeira edição em 2008, até a edição de 2012. Esse *workshop* acontece na WWW Conference, uma

conferência bem conhecida que teve sua primeira edição em 1994¹.

Os requisitos encontrados foram organizados em ordem cronológica, retiradas as redundâncias das referências encontradas, e relacionados com um modelo de sistema de informação, usado para auxiliar na identificação dos requisitos de publicação de dados. A partir dos requisitos de publicação foi criada uma lista de características necessárias para serialização de dados e comparada aos formatos encontrados na literatura. Passada essa fase, as tecnologias selecionadas foram organizadas num modelo utilizado na implementação desta proposta.

4.2 Requisitos e sistemas de informação

Esta seção apresenta os elementos obtidos a partir do levantamento preliminar realizado neste trabalho e representam os desafios e exigências a serem implementadas em uma plataforma para a publicação de dados. É importante notar que alguns dos requisitos encontrados transcendem o modelo de sistema de informação, consequentemente, o escopo dessa proposta. Não se pretende neste trabalho fazer uma revisão de modelos de sistema de informação, mas identificar quais requisitos devem ser considerados para a publicação dos dados ao usuário, ou seja, aquilo que é acessado pelo usuário. Portanto, um modelo simples de sistema de informação atende ao propósito de auxílio de tomada de decisão de projeto.

A Figura 11 mostra um modelo do sistema de informação em quatro camadas. Por ser um modelo *top-down* o foco de desenvolvimento é colocado no usuário final, no caso o cliente que utiliza o canal de acesso. A camada de apresentação é acessada pelo cliente, onde são definidos formatos e protocolos para dados que o cliente acessa através dos canais de sua plataforma. A lógica da aplicação deve entregar o conteúdo no formato especificado na camada de apresentação a partir da camada de gerenciamento de recursos, que define as fontes de dados e a organização

¹mais detalhes sobre a história da WWW Conference são encontrados em <http://www.iw3c2.org/conferences>

necessária para a lógica da aplicação.

fonte: Alonso, 2004[42]

Figura 11 – Camadas de um sistema de informação

A Tabela 2 lista os requisitos que foram identificados na literatura, e na legislação, e onde requisitos devem ser implementados num sistema de informação baseado no modelo da Figura 11. Nesta tabela é considerado o mínimo que precisaria ser implementado num sistema, portanto, é somente uma possibilidade de implementação. Por exemplo, o requisito *dados devem ser atuais* pode ser implementado na camada de *gerenciamento de recurso*, mas isso não é necessário para inclusão da informação de atualidade dos dados na camada de *apresentação*, para isso basta a informação ser incluída pelo *lógica de aplicação*.

A completude e primariedade de dados devem ser apresentadas ao usuário, mas para que isso seja possível, é necessário que os dados sejam completos e primários no recurso, porque a lógica de aplicação não cria dados nem deve agregá-los. Portanto, os dados devem ser completos e primários no gerenciamento do recursos e na camada de apresentação. Da mesma forma, a estrutura de dados que existe no gerenciamento de recurso deve estar representado na camada de apresentação para permitir que o cliente facilmente reorganize dados para (re)uso nos seus meios de análise. Esta reutilização pode ser melhorada pelo uso de metadados, onde alguns só podem ser apresentados ao cliente se existirem no gerenciamento de recursos.

Tabela 2 – Requisitos levantados neste trabalho

Requisito	Camadas de SI
Dados devem ser completos	gerenciamento de recurso, apresentação
Dados devem ser primários	gerenciamento de recurso, apresentação
Dados devem ser atuais	lógica de aplicação, apresentação
Dados devem ser acessíveis	lógica de aplicação, apresentação
Dados devem ser processáveis por máquina	gerenciamento de recurso, apresentação
Dados devem ser não-discriminatórios	lógica de aplicação
Dados devem estar em formatos não proprietários	apresentação
Dados devem ter licença livre	apresentação
Disponíveis na Web	apresentação
Identificador único para dados	apresentação
Ligar dados a outros dados	lógica de aplicação, apresentação
Ter informações de proveniência	gerenciamento de recurso, apresentação
Dados devem ser permanentes	gerenciamento de recurso, apresentação
Baixo custo de uso de dados	*
Prover alternativas de acesso	lógica de aplicação, apresentação
Criar lista de dados	*
Usar vocabulários controlados	gerenciamento de recurso, apresentação
Usar metadados	gerenciamento de recurso, lógica de aplicação, apresentação
Dados devem ser validáveis	lógica de aplicação, apresentação

** este requisito está além do modelo de SI*

Além da completude e primariedade, questões como a criação de dados, acesso, versionamento de dados, vocabulários e suas versões, atualização, informações de proveniência e permanência, devem ser implementados no gerenciamento de recurso e ou lógica de aplicação para, devidamente, ser apresentado ao cliente como metadados.

A frequência com que os dados são atualizados e a variedade de formatos que o mesmo recurso é apresentado depende da lógica da aplicação, considerando que o recurso existe na camada de gerenciamento de recursos. É importante notar que

algumas das informações produzidas na lógica da aplicação devem ser representadas nos dados, pois esses podem ser copiados e passados de mão em mão, de um sistema a outro, por exemplo, se a data de atualização não estiver representada nos dados, essa informação é perdida na redistribuição de informação. Portanto, alguns metadados gerados na lógica de aplicação devem estar junto aos dados.

Na camada de lógica da aplicação é definido como o conteúdo será entregue ao cliente na camada de apresentação, por isso, nessa camada deve ser garantido acesso não discriminatório aos dados. O gerenciamento de recurso e lógica da aplicação podem ser implementados com tecnologias proprietárias, mas os dados devem estar em formatos não-proprietários, ter licença livre, estar na Web e ter identificador único na camada de apresentação.

O custo do acesso aos dados deve ser o mínimo possível. Esse requisito depende da plataforma e dos canais de acesso do cliente, mas fornece orientação sobre como as informações devem ser apresentadas e onde (Web) no sistema de informação. O requisito de se criar uma lista de dados para facilitar a descoberta desses, não está relacionado direto com um padrão de publicação de dados, pois um catálogo é um mecanismo externo aos dados publicados que deve registrar uma lista de identificadores de dados, não os dados em si, portanto não faz parte do escopo desta proposta.

A maioria dos requisitos legais são contemplados pelos requisitos da literatura. Informações pormenorizadas e tempo real na LRF estão relacionadas aos princípios de DGA *completo*, *primário* e *atual*. Questões como gravar em diversos formatos de relatório e possibilitar acesso automatizado estão relacionadas aos princípios de DGA *acessível* e *processável por máquina*. A exceção é a disponibilização de ferramenta de pesquisa, que está além de uma proposta de padrão de publicação, portanto, não faz parte desta proposta.

Apresentados os requisitos relacionados às camadas de um sistema de informação, pode-se observar que dezesseis dos dezenove requisitos estão relacionados à apresentação desses dados.

Serializar dados significa representá-los de uma forma que possam ser transmitidos e armazenados em sistemas de informação. Para isso são criados padrões cujas características são desenvolvidas para fins de como armazenar texto, transmitir áudio e publicar na Web. Antes da decisão de qual padrão utilizar, é preciso conhecer a necessidade de representação dos dados e das possibilidades tecnológicas. Nesta seção são abordados aspectos da serialização de dados importantes para a compreensão das tecnologias que serão citadas neste capítulo.

A Figura 12 contém exemplos de dados serializados em quatro padrões de formato de arquivo. As três primeiras linhas contém um pequeno trecho de uma imagem no formato Portable Network Graphics (PNG) que continha texto escrito. Da linha 5 a 8 estão dados no formato Comma-Separated Values (CSV), onde os valores são separados por vírgulas, sendo a linha 5 uma sequência dos nomes das variáveis nas linhas seguintes, assim na linha 8, *March* é um nome de *Modelo*. Da linha 10 até 15 há uma lista de nomes de frutas no formato HyperText Markup Language (HTML), padrão W3C para páginas eletrônicas na Web. A partir da linha 17 ao fim, um exemplo de dados no formato Extensible Markup Language (XML)².

Extrair texto de uma imagem exige processamento de reconhecimento de caracteres, pois os dados representam pontos coloridos na imagem, não o texto diretamente, resultando numa sequência de caracteres como no exemplo da imagem em PNG. No exemplo em CSV, "*Grand Vitara*" é escrito entre aspas por conter espaço em branco; se nos valores de *Preço* existissem centavos também seriam necessárias aspas, no caso do separador decimal ser vírgula. Os exemplos em HTML e XML contém códigos delimitados pelos sinais *j* e *j* que são chamados de marcas ou, em inglês, *tags*, onde as marcas com */* representa o fim de um dado ou grupo de dados. As *tags* HTML descrevem uma apresentação visual a ser construída, no caso *h3* é um subtítulo, *ul* e *li* combinados compõem uma lista com marcado-

²Especificações dos formatos citados são encontrados respectivamente nos endereços <http://www.w3.org/TR/PNG/>, <http://tools.ietf.org/html/rfc4180>, <http://www.w3.org/TR/html401/> e <http://www.w3.org/TR/REC-xml>

```
1 de59 b524 4c30 ac64 afc7 34f7 c068 19b2
2 72bb 3fe3 1c98 0e72 dff7 7e6f 7644 8077
3 90e1 a13f 9047 90df 6d93 af10 2e65 f961
4
5 Ano, Fabricante, Modelo, Preço
6 2003, Ford, Ka, 23000
7 2005, Susuki, "Grand_Vitara", 40000
8 2013, Nissan, March, 36000
9
10 <h3>Frutas</h3>
11 <ul>
12 <li>maçã</li>
13 <li>limão</li>
14 <li>abacaxi</li>
15 </ul>
16
17 <despesa>
18 <item>
19 <tipo>gasolina</tipo>
20 <valor moeda="Real">320,00</valor>
21 </item>
22 <item>
23 <tipo>cartão de crédito</tipo>
24 <valor moeda="Dólar">200,00</valor>
25 </item>
26 </despesa>
```

fonte: elaborado pelo autor

Figura 12 – Exemplos de dados serializados

res. Uma das diferenças do XML em relação ao HTML é a descrição do nome do dado nas *tags*. Diferente do CSV, onde se representa os dados na mesma linha, em HTML e XML é possível estabelecer uma hierarquia entre as *tags*, no exemplo XML, *tipo* e *valor* contêm dados, *item* é um grupo de dados que contém as duas informações, e *despesa* contém dois grupos de informações do tipo *item*. Outra característica do XML, e do HTML, é a possibilidade de extensão da descrição do dado usando atributos dentro das *tags*. No exemplo, XML *moeda* é um atributo de *valor* que descreve uma unidade monetária.

Diferentes tipos de formatos possuem características distintas. Numa imagem não se descreve as variáveis dos dados como em CSV ou XML, sendo que CSV

não representa hierarquia nem possui a possibilidade de uso de atributos, como em HTML e XML. Nas seções seguintes serão abordados outros padrões de formatos que possuem características similares às apresentadas.

4.3 Projeto da proposta

Algumas decisões de projeto são necessárias para a decisão de quais formatos podem ser utilizados para atender aos requisitos levantados. Algumas dessas decisões vêm direto dos requisitos, por exemplo, o formato deve ser não proprietário e processável por máquina. Neste trabalho, será considerado processável por máquina o formato que contiver alguma informação de estrutura sobre os dados. No entanto, só será considerado um formato validável se essa estrutura incluir uma declaração da mesma. Outra importante decisão é utilizar metadados para contextualizar a publicação da execução dos dados de orçamento público, não a história desde quando a informação foi coletada.

Na Tabela 3, estão relacionadas as características necessárias para publicação de dados da execução do orçamento e formatos de serialização considerados para implementação. Os formatos de arquivo na tabela foram selecionados a partir da literatura consultada. Quando determinada característica pode ser atendida pelo formato, é colocado um tique(✓) como marcação positiva. Se não, é colocado o nome de uma especificação que pode ser combinada ao formato para atender a característica desejada. Caso não exista especificação auxiliar um xis(✗) é usado como marcação negativa.

Sabe-se que no XBRL o *linkbase presentation* é utilizado para criar outras visualizações dos dados. Na Figura 9 há exemplos de visualização, formatos HTML, DOC e PDF, mas decidimos utilizar o padrão XSLT nesta proposta por ser reconhecido pelos programas navegadores Web. Outras tecnologias podem ser utilizadas para se verificar integridade e autenticidade dos dados, como o GPG que pode ser utilizado em qualquer tipo de arquivo para qualquer propósito, mas na Tabela 3

Tabela 3 – Características de formatos para serialização de dados

Característica	CSV	JSON	JSON-LD	XML	XBRL	RDF
Processável por máquina	✓	✓	✓	✓	✓	✓
Inclusão de metadados	✗	✓	✓	✓	✓	✓
Representa taxonomia	✗	✓	✓	✓	✓	✓
Declaração de estrutura	✗	JSON Schema	✓	XML Schema	✓	RDF Schema
Combina vocabulários	✗	✗	✓	XML Namespace	✓	✓
Ligação entre dados e ou conceitos	✗	✗	✗	XLink	✓	✓
Conversão de tipos	✗	✗	✗	XSLT	XSLT	XSLT
Integridade e autenticidade	✗	✗	✗	XML Signature	XML Signature	XML Signature

só foram consideradas tecnologias que têm relação direta com os formatos listados. Nesse sentido, existe uma proposta de assinatura do formato JSON³ mas ainda está em desenvolvimento. Por isso, não foi incluída na tabela. Na especificação XBRL deliberadamente não há mecanismo de integridade (XBRL International, 2012)[30], mas por ser baseado em XML foi considerada a especificação XML Signature⁴. O mesmo ocorre com RDF.

Como elementos de metadados, Powell (2009:89)[25] afirma o conjunto Dublin Core atende às necessidades de metadados de coleção, é bem documentado e é mantido por uma comunidade experiente. Hartig (2009) aponta Dublin Core como um dos vocabulários para metadados de proveniência. Assim, o formato para seria-

³essa proposta de assinatura JSON encontra-se em <http://tools.ietf.org/html/draft-ietf-jose-json-web-signature-14>

⁴a recomendação XML Signature encontra-se em <http://www.w3.org/TR/xmlsig-core/>

lização de dados desta proposta é XBRL, com transformações XSLT, metadados com elementos Dublin Core e verificação de integridade e autenticidade XML Signature. Para identificadores únicos dos dados publicados, é necessário uma política de nomeação de endereços permanentes que será proposta com a implementação.

4.4 Modelo da proposta

A figura 13 contém o modelo da proposta deste trabalho, relacionando de baixo para cima as dependências de conceitos ou padrões utilizados para cumprir os requisitos na camada de apresentação ao usuário de dados.

A proposta depende diretamente de XBRL, XSLT, XML Signature e Persistent URL. A instância XBRL e a taxonomia XBRL baseiam-se, além da própria especificação, na taxonomia que é elaborada a partir da legislação como indicado por Campos e Gomes (2010), nos elementos Dublin Core selecionados para informações de coleção e proveniência, e de padrões incorporados da Família XML (XML Namespace, XML Schema, XLink e XPointer). XSLT e XML Signature também são padrões da Família XML, mas estão em destaque porque não fazem parte do XBRL e são necessários para cumprir requisitos de acesso, integridade e autenticidade. Persistent URL não é um padrão em si, mas um conceito que significa utilizar o padrão de endereços eletrônicos URL que são permanentes.

Publicação da Execução do Orçamento Público (proposta)			
XBRL Instância e Taxonomia (validação e internacionalização)		XSLT (conversão)	XML Signature (integridade e autenticidade)
Legislação (taxonomia)	Dublin Core (metadados)	Família XML (meta linguagem XML e outros padrões W3C relacionados)	
			Persistent URL (identificador único)

fonte: elaborado pelo autor

Figura 13 – Modelo de proposta de publicação de dados orçamentário

No modelo da Figura 13, com XBRL podem ser atendidos os requisitos *dados devem ser completos*, *dados devem ser primários*, *dados devem ser processáveis por máquina*, *dados devem estar em formatos não proprietários*, *ligar dados a outros dados* (utilizando XLink), *usar vocabulários controlados* (utilizando XML Namespace) e *dados devem ser validáveis*. Com os elementos Dublin Core pode-se incorporar metadados que atendam aos requisitos *dados devem ter licença livre*, *usar metadados*, *dados devem ser atuais* e *ter informações de proveniência*. Com XSLT é possível converter dados publicados em formatos baseados em XML, como XBRL, para formatos como JSON e CSV, incrementando as formas de obtenção dos dados e, assim, atendendo aos requisitos *dados devem ser acessíveis* e *prover alternativas de acesso*. Os requisitos *disponíveis na Web*, *identificador único para dados* e *dados devem ser permanentes* dependem de URLs permanentes e da manutenção do sistema de informação que contém os recursos.

O uso dos padrões da proposta não garante o cumprimento dos requisitos. Como exemplo dessa problemática, o princípio de primariedade dos dados depende de como a taxonomia XBRL é elaborada, se não forem incluídos elementos sobre dados primários na taxonomia não será possível validar se o dado primário existe.

4.5 Implementação da proposta

Neste capítulo, é relatado o processo de implementação da proposta, como os padrões selecionados são utilizados, a construção de um protótipo, a seleção de dados, dificuldades encontradas e as validações realizadas. O primeiro passo foi serializar metadados e dados em XBRL, instância e taxonomia, pois isso era necessário para criar as transformações de formato com XSLT. O resultado da serialização em XBRL estão no Anexo A (instância) e no Anexo B (taxonomia).

Uma instância em XBRL foram utilizados três grupos de elementos da especificação: *context* para dados de contexto e onde foram incluídos os metadados, *unit* para unidade de moeda, e *fact* para inclusão dos fatos, no caso os dados da

execução do orçamento. Esse último é o local onde foi representada a taxonomia da execução do orçamento com receitas e despesas.

Algumas ferramentas textuais e gráficas foram testadas para elaboração da instância e da taxonomia XBRL e, no final, optamos por ferramentas em modo texto como descrito a seguir. A Figura 14 contém um pequeno trecho de código serializado na notação YAML, onde a hierarquia dos elementos é dada pela organização de espaços em branco antes de cada elemento. Desse modo, *xbrl* é o elemento principal e, hierarquicamente, *context* e *unit* estão diretamente ligados a *xbrl*. O elemento *orcamento* será o local para se inserir os fatos no XBRL.

```
1 xbrl:
2   context:
3     entity:
4       scheme: "http://orgao.gov.br/"
5       identifier: "Órgão_do_Governo"
6     period:
7       startDate: "2012-01-01"
8       endDate: "2012-12-31"
9   unit:
10    measure:
11      iso4217: "BRL"
12   orcamento:
13     receita:
14       categoria:
15         nome: "Receita_Corrente"
16       origem:
17         nome: "Receita_Tributária"
```

fonte: elaborado pelo autor

Figura 14 – Exemplo de serialização YAML

Antes de criar a taxonomia em XBRL foi preciso escolher quais elementos Dublin Core seriam utilizados para as informações de coleção e proveniência. Foram selecionados os elementos do trabalho de Hartig(2009)[34] mais os elementos *type*, *title* e *license* com os seguintes propósitos:

- *type*: tipo de informação, no caso *Dataset* em vocabulário do Dublin Core

- *title*: nome do conjunto de dados
- *creator*: nome de quem cria os dados, o executor do orçamento
- *publisher*: nome de quem publica o orçamento, pode ser igual ao *creator*
- *created*: data de criação dos dados
- *modified*: data da última modificação dos dados
- *source*: endereço Web (URL) dos dados
- *license*: licença de uso dos dados ou citação da legislação (LRF, LAI)
- *provenance*: detalhes de criação de dados, opções de pesquisa ou simples *download*

A partir da análise da legislação foi construída uma taxonomia no formato YAML, com receitas, despesas e todos os campos obrigatórios nesses dois grupos do orçamento, numa organização baseada em XBRL. Também foram acrescentados os elementos de metadados Dublin Core selecionados. Terminada a taxonomia, com a ferramenta *online* (UN)Serialize.me⁵, esta foi convertida para um documento XML que foi a base para o documento XBRL, faltando identificar os vocabulários com XML Namespace e os atributos dos elementos.

XML Namespace usa rótulos para indicar de qual vocabulário vem o elemento. No Anexo A, entre as linhas 9 e 14, estão as declarações dos vocabulários utilizados na proposta, identificados por *xmlns[:rótulo]* seguido do endereço Web onde esse vocabulário é encontrado. Assim, *xmlns:dc* refere-se ao vocabulário Dublin Core e *dc:license* é o elemento *license* desse vocabulário. Pode-se definir um vocabulário padrão omitindo-se o rótulo, recurso utilizado nos elementos dos fatos da execução do orçamento no Anexo A.

⁵as ferramentas utilizadas no processo de criação da instância e taxonomia, (UN)Serialize.me, Trang e XMLStarlet podem ser acessadas, respectivamente, nos endereços <http://unserialize.me/>, <http://www.thaiopensource.com/relaxng/trang.html> e <http://xmlstar.sourceforge.net/>

Para completar a instância XBRL foi necessário adequar alguns atributos, acrescentar um caminho para uma taxonomia XBRL, linha 16 do Anexo A, e adicionar rótulos a elementos que não eram do vocabulário padrão. Também foram incluídos referências às transformações XSLT, da linha 3 a linha 5, deixando a transformação para XHTML como padrão, CSV e JSON como alternativas.

Na elaboração da instância, notamos que itens da classificação como *função* e *subfunção* poderiam ter o mesmo código numérico. O atributo *id* usado para colocar os códigos, por definição, não pode ter valores repetidos em elementos diferentes. Portanto, foi adotada a notação húngara⁶ para diferenciar esses atributos, colocando-se duas letras extraídas do nome do elemento antes do número. Por exemplo, *og39* refere-se ao órgão com código 39.

Terminada a instância XBRL, a ferramenta *trang* foi utilizada para iniciar a elaboração da taxonomia XBRL pois essa ferramenta cria um XML Schema a partir de qualquer tipo de XML. Se os vocabulários forem corretamente referenciados, a *trang* busca esquemas de validação desses vocabulários automaticamente. No XML Schema produzido pela ferramenta ainda foi necessário adequar os tipos de dados, incluir outros elementos quando necessários, indicar o número de ocorrências de elementos com *minOccurs* e *maxOccurs* quando necessário, acrescentar atributos e as declarações de XML Namespace. O resultado desse processo se encontra no Anexo B. A validação da instância XBRL com a taxonomia XBRL foi realizada com a ferramenta *xmllstarlet*.

Com a instância XBRL em mãos, foram criadas as transformações XSLT para XHTML, JSON e CSV. As transformações para XHTML e JSON foram validadas nos navegadores Web Firefox, Internet Explorer, Google Chrome, Opera e Safari para Windows. A transformação para CSV apresentou o problema de conversão do caracter " (aspas duplas) nos dados para a especificação CSV, onde deveria ser convertido para " " (duas aspas duplas). Descobriu-se que era necessário utilizar uma função da especificação XSLT 2.0 chamada *replace* e que os navegadores têm

⁶ver detalhes em http://en.wikipedia.org/wiki/Hungarian_notation

implementado suporte a XSLT 1.0, dessa forma, a conversão para CSV só foi validada com a ferramenta *xmlstarlet*.

Estruturalmente, na serialização de dados, os formatos XHTML e JSON são próximos do XBRL, hierárquico com marcas de início e fim, mas CSV tem uma estrutura bidimensional similar a uma tabela. Mesmo assim, foi possível transformar uma estrutura hierárquica em uma estrutura em tabela com a especificação XSLT 1.0, excluindo o problema da conversão de " (aspas duplas).

Também foi realizado um teste de conversão de *links* XLink para XHTML Link, ou seja, ligações para outros dados em XBRL podem ser convertidas para *links* de páginas Web com XSLT. Na Figura 15, é mostrado como foi feita essa validação. Instâncias XBRL diferentes, que têm uma ligação real entre si, foram transformadas no navegador Web em visualizações XHTML no momento do acesso.

O XLink é convertido em XHTML Link e aponta para outra visualização XHTML, que não existe no mesmo instante da primeira visualização. Essa transformação foi validada nos cinco navegadores Web anteriormente citados.

No entanto, foi descoberta uma limitação ao se utilizar XPointer, que associado ao XLink pode apontar direto para elementos no meio dos dados. Essa limitação é utilizar somente a forma abreviada de XPointer para a transformação ser compatível com XHTML Link.

fonte: elaborado pelo autor

Figura 15 – Transformação de links com XSLT

XML Signature e Permanent URL não foram implementados porque o ambiente de validação estava limitado a uma máquina com sistemas instalados localmente. Seria necessário um ambiente de rede com mais máquinas, registro de domínio e certificado digitais para as assinaturas. No entanto, XML Signature é uma recomendação desde 2008 e a definição de endereços permanentes é política que só depende de configurações de domínio no governo.

Visto que no governo os órgãos são divididos em Unidades Orçamentárias, a proposta para Permanent URLs é fazer a publicação de execução de dados do orçamento por UO, no domínio do órgão seguindo o modelo `http://orcamento.[orgao].[uf].gov.br/[ano]/uo[nmero_da_UO]`, por exemplo, `http://orcamento.saude.sp.gov.br/2013/uo252`. Nesses endereços seriam encontradas as instâncias XBRL como a do Anexo A.

Uma assinatura XML Signature pode ser feita de três formas: num arquivo externo a instância XBRL (*desanexada*), inserindo a assinatura como uma parte da instância (*envelopada*) ou inserindo a instância dentro da assinatura (*envelopando*). Para ter uma assinatura *envelopada* é preciso alterar a taxonomia XBRL e uma parte dos dados não seria assinada. A assinatura do tipo *envelopando* vai exigir alteração na taxonomia XBRL, mas permite assinar todos os dados. No entanto, não pode ser aplicada a taxonomia em si, sem que seja feito um XML Schema da taxonomia XBRL com a previsão da assinatura. A assinatura *desanexada* pode ser aplicada à instância e taxonomia XBRL e implantada a qualquer momento. Portanto, essa é a forma proposta neste trabalho.

4.6 Importação de dados

Para validação da proposta, dados do orçamento do Estado de São Paulo foram baixados para serem inseridos num banco de dados relacional; foram pesquisados dados do Governo Federal, mas somente os gastos diretos tinham todas as informações; e dados da Prefeitura do Município de São Paulo, mas também esta-

vam incompletos. Esses dados estavam no formato CSV com 1,7 GBytes de dados de despesas entre os anos de 2008 e 2012. As primeiras tentativas de importação falharam porque o arquivo continha erros que foram descobertos manualmente, totalizando 46(quarenta e seis) linhas erradas no arquivo CSV.

Superados esses problemas, foram importados 2.417.979 (dois milhões, quatrocentos e dezessete mil, novecentos e setenta e nove) registros de despesas entre 2008 a 2012, incluindo a classificação orçamentária das mesmas, em 27 minutos. Os dados ocuparam 213 MBytes em dados no banco de dados relacional, 493,7 MBytes com os índices do banco de dados, e 381,5 MBytes em arquivos XBRL de instâncias das Unidades Orçamentárias.

A Figura 16 mostra a visualização da publicação de dados pela implementação desta proposta no navegador Web Firefox, da Mozilla Foudation. Acima como o navegador mostra o XBRL sem transformações, abaixo a transformação para o formato JSON realizado pelo próprio navegador com a tecnologia XSLT.

The image shows a web browser window with two panes displaying XML data. The top pane shows the raw XBRL markup, and the bottom pane shows the corresponding JSON representation.

Top Pane (XBRL Markup):

```

- <xbrli:xbrl targetNamespace="http://www.gpopai.usp.br/orcamento/">
  <link:schemaRef xlink:type="simple" xlink:href="http://www.gpopai.usp.br/orcamento/
  /esquema.xsd"/>
  <!-- contexto dos fatos -->
  - <xbrli:context id="cx1">
    - <xbrli:entity scheme="http://www.gpopai.usp.br/orcamento/">
      <xbrli:identifier scheme="http://www.gpopai.usp.br/orcamento/">
      </xbrli:entity>
    - <xbrli:period>
      <xbrli:startDate>2012-01-01</xbrli:startDate>
      <xbrli:endDate>2012-12-31</xbrli:endDate>
    - <xbrli:scenario>
      - <dc:dcterms>
        <dc:type>Dataset</dc:type>
        - <dc:title>
          Execução orçamentária de UNIVERSIDADE DE SAO PAULO - USP
        </dc:title>
        <dc:creator>Secretaria do Planejamento do Estado de São Paulo</dc:creator>
        <dc:publisher>Marcelo Tavares de Santana</dc:publisher>
        <dc:date>2013-12-04</dc:date>
        <dc:source>http://www.gpopai.usp.br/orcamento/uo.php?id=43058</dc:source>
        <dc:rights>Constituição de 1988, LRF, LAI, Decreto 7185</dc:rights>
        <dc:provenance>download de arquivo</dc:provenance>
      </dc:dcterms>
    - </xbrli:scenario>
  </xbrli:context>
  <!-- unidade monetária -->
  - <xbrli:unit id="brl">

```

Bottom Pane (JSON Representation):

```

/* conversões pré-definidas, por padrão os navegadores vão escolher XHTML
<?xml-stylesheet type="text/xsl" href="xhtml.xsl" title="XHTML" alternate="yes" ?>
/** relatório XBRL */ /** xmlns (XML Namespace) descreve o caminho dos padrões/vocabulários para o documento
schemaRef:null, /* contexto dos fatos */
/*@id="cx1"*/ /** xbrli:context */: {
  /*@scheme="http://www.gpopai.usp.br/orcamento/"*/ /** xbrli:entity */:
  identifier:null,
  "xbrli:period": {
    "xbrli:startDate": new Date(2012, 01 - 1, 01),
    "xbrli:endDate": new Date(2012, 12 - 1, 31)},
  "xbrli:scenario": {
    "dc:dcterms": {
      "dc:type": "Dataset",
      "dc:title": "Execução orçamentária de UNIVERSIDADE DE SAO PAULO - USP",
      "dc:creator": "Secretaria do Planejamento do Estado de São Paulo",
      "dc:publisher": "Marcelo Tavares de Santana",
      "dc:date": new Date(2013, 12 - 1, 04),
      "dc:source": "http://www.gpopai.usp.br/orcamento/uo.php?id=43058",
      "dc:rights": "Constituição de 1988, LRF, LAI, Decreto 7185",
      "dc:provenance": "download de arquivo"}}, /* unidade monetária */
  /*@id="brl"*/ /** xbrli:unit */: {
    "xbrli:measure": "iso4217:BRL"}, /* fatos da execução do orçamento */
  orcamento: {
    despesa: {
      /*@id="og43000"*/ /** orgao */: {
        nome: "SECRETARIA DE ENSINO SUPERIOR",
        /*@id="uo43058"*/ /** uo */: {

```

Figura 16 – Visualização da implementação em XBRL e JSON

Capítulo 5

Discussão da proposta

Na legislação brasileira está regulamentada a publicação de dados da execução orçamentária dos entes da federação. O acesso a esses dados é um direito de todos os cidadãos. Na mesma linha, os princípios de DGA foram pensados para preservar o direito à informação à sociedade. A forma como os dados do orçamento são publicados podem influenciar na forma como são (re)usados pelos cidadãos. Neste capítulo, são resgatados conceitos, requisitos de publicação e tecnologias que compõe a proposta deste trabalho, e discutido como cada requisito pode ser atendido pela proposta.

Na LRF as informações do orçamento devem ser publicadas de forma pormenorizada e na LAI essas informações devem ser primárias, considerando que no Decreto 7.185 é estabelecido quais informações de receitas e despesas, os dois grupos que compõem a execução do orçamento. A legislação brasileira converge com os princípios de DGA que estabelecem que os dados devem ser completos e primários.

Na proposta, a Taxonomia XBRL é a forma como a estrutura dos dados podem ser descritas e é utilizada para fazer a validação da Instância XBRL de modo automatizado. Porém, a qualidade do quanto os dados são completos e primários depende de como essa taxonomia é elaborada, pois depende de como é capturada da documentação legal. Se a lei estabelecer publicação de informação agregada e a taxonomia for feita seguindo essa regra, a Taxonomia XBRL somente poderá validar informação agregada.

Por outro lado, uma Taxonomia XBRL facilita a análise do que é validado na publicação, inclusive porque ela própria é processável por máquina. Independente

de como é feita a elaboração da taxonomia, na proposta deste trabalho é possível representar a estrutura de dados, número de ocorrências e tipos de dados, ou seja, informações sobre o que são os dados, características herdadas da Taxonomia XBRL e XML Schema.

No Decreto 7.185 é estabelecido que tempo real é um dia útil após a execução de receitas ou despesas para a publicação dos dados. Na proposta do trabalho há cinco elementos diferentes de data para auxiliar na compreensão do que está sendo publicado: *xbrli:startDate* e *xbrli:endDate* são utilizados para indicar o intervalo de tempo relativo aos dados de execuções de receitas e despesas, *dc:created* para indicar a data da primeira publicação dos dados, *dc:modified* para a data da última modificação, e *data* para indicar em cada receita ou despesa sua data de execução.

Através da observação dessas datas é possível avaliar se a publicação está atualizada dentro do que é estabelecido na lei, mais que um dia útil de diferença entre *xbrli:endDate* e *dc:modified* pode indicar um problema na publicação dos dados. Portanto, além de ser possível identificar a execução a cada receita e despesa, pode-se recuperar as informações de criação da publicação da execução do orçamento como um todo de forma automatizada, pois os metadados também são processáveis por máquina.

Na LRF é determinado a publicação da execução do orçamento em meios eletrônicos. Posteriormente fica definido no Decreto 7.185 que esse meio é a Internet. Considerando que o requisito na literatura é a disponibilização dos dados na Web, há uma pequena diferença entre lei e literatura. Internet é o nome dado à rede mundial de computadores, uma estrutura física por onde trafega a informação. A Web é uma rede de informações baseadas em recomendações do W3C cujos dados são trafegados na Internet. Portanto, a Internet é o meio físico onde a Web está inserida. Existem outros tipos de redes de dados que utilizam a Internet como meio físico. Não foi encontrada, na legislação, exigência de publicação de dados na Web. No Decreto 7.185 há uma sugestão de ser esse o meio preferencial. Na proposta de publicação deste trabalho, todos os padrões de formatos selecionados

são recomendações W3C ou baseadas nela, de forma que a Web é o meio onde essa proposta pode ser implantada.

O princípio de DGA *acessível* indica que os dados devem ser acessíveis para o maior número de usuários e mais amplas finalidades. Isso se relaciona com o Decreto 7.185 no atendimento preferencial ao padrão de acessibilidade do governo e-MAG, e com a LAI ao definir que se deve possibilitar gravar relatórios em diversos formatos não-proprietários. Na proposta deste trabalho, as transformações por XSLT foram validadas para os formatos de arquivo CSV, JSON e XHTML; este último de acordo com a e-MAG. Além disso, documentos XBRL podem ser convertidos em outros formatos com o uso do *linkbase presentation*(Fernandes *et al.*, 2008)[41]. Sabe-se também que dados em XBRL podem ser publicados como Linked Data com ferramentas automatizadas(Garcia e Gil, 2009)[40]. Portanto, na proposta de publicação deste trabalho há três meios para conversão dos dados publicados: XSLT, *linkbase presentation* e publicação como Linked Data. Os formatos para de conversão conhecida, validados ou na literatura consultada, são CSV, JSON, XHTML, PDF, DOC e RDF, mas com XSLT se pode transformar dados para outros formatos baseados em XML.

Na LAI e o Decreto 7.724 consta que a publicação dos dados deve possibilitar o acesso automatizado e utilizar formatos abertos legíveis por máquina. Esses requisitos legais convergem com o princípio de DGA *processável por máquina*, equivalente a duas estrelas na abordagem Linked Data. As conversões de formatos discutidas anteriormente são feitas por máquina, assim essa proposta atende aos requisitos legais e aos princípios de DGA citados. Porém, há outros dois aspectos que não estão no escopo desses requisitos, pois os metadados e a taxonomia também são processáveis por máquina. É essa possibilidade de processamento por máquina de outros aspectos relativos aos dados que permite que XBRL seja publicado como Linked Data e a verificação sobre a atualidade dos dados possa ser automatizada.

Como esta proposta de publicação é baseada em recomendações abertas do W3C; dados, metadados e taxonomia são representados em formatos abertos. Na

LAI, no Decreto 7.724, no princípio de DGA *não-proprietário* e na classificação três estrelas de Linked Data, consta o uso de formatos não-proprietários, assim a proposta deste trabalho atende todas essas abordagens. O uso de recomendações abertas também contribui com o baixo custo de uso de dados, um dos princípios de DGA, e com o fornecimento de informação gratuita indicado na LAI e no Decreto 7.724. O padrão XBRL, particularmente, é uma recomendação no e-PING (MPOG, 2012)[11] para troca de informações financeiras.

Um dos requisitos da literatura é criar lista de dados, e na legislação, LAI e Decreto 7.724, outro requisito é criar ferramentas de pesquisa. Esta proposta de publicação trata somente da apresentação dos dados, mas pelas possibilidades de processamento por máquina de dados, metadados e taxonomia, potencializa a indexação por sistemas automatizados.

Na legislação, Decreto 7.185, LAI e Decreto 7.724, a validação de dados é um requisito, pois os governos devem fornecer informação íntegra e autêntica. Na literatura consultada não continha essa questão, inclusive, na recomendação XBRL consta que deliberadamente optaram por não propor uma tecnologia de integridade e autenticidade de dados(XBRL, 2012)[30]. Na proposta deste trabalho, a Taxonomia XBRL valida os dados quanto à estrutura, tipo e ocorrência, mas existe a possibilidade de um valor de receita ou despesa ser trocado por erro na transmissão de dados. Para garantir integridade e autenticidade dos dados foi necessário propor XML Signature, que é adotado pelo Governo Federal, conforme descrito na e-PING(MPOG, 2012)[11], portanto, já é uma prática de governo.

Outra questão que consta na legislação, LAI e Decreto 7.724, é a publicação detalhada dos formatos utilizados na estruturação da informação. Não há indicação de como disponibilizar esse detalhamento, nem onde. Uma Taxonomia XBRL é uma descrição detalhada dos dados que pode ser processada por máquina. Com o uso dos *linkbases* do XBRL pode-se descrever também relações entre cálculos e conceitos, associações com normas, relações entre conceitos. Essas são mais opções de descrição dos dados que podem ser utilizadas na proposta de publicação deste

trabalho, cabendo a decisão de uso de *linkbases* a quem publica os dados.

Sobre o princípio de DGA *licença livre*, Craveiro *et al.* (2013)[4] relataram que dos dados observados nenhum tinha licença livre. No entanto, no Decreto 7.724 é regulamentado que todas as informações de governo são, *a priori*, de domínio público e mesmo considerando isso uma licença livre, não é uma garantia de que eles não possam ser publicados com uma licença restritiva. Na proposta deste trabalho o elemento Dublin Core *license* é utilizado para especificar sob qual condição legal os dados foram publicados, sendo seu uso obrigatório. Caso não exista uma licença, recomenda-se indicar a legislação vigente, como no exemplo do Anexo A, linha 35, onde é citado o Art. 5º da Constituição, a Lei 12.527 (LAI) e o Decreto 7.724. Essas informações são importantes para, num futuro acesso, ser possível recuperar a parte legal do histórico dos dados.

Até este ponto foram apresentadas discussões sobre requisitos que tinham correspondência com a legislação brasileira sobre publicação de dados da execução do orçamento público. Outros requisitos na literatura não tem tal correspondência mas dão suporte aos requisitos anteriores, como discutido será a seguir.

Os princípios de DGA *não-discriminatório* tem relação direta com o acesso que é oferecido aos dados pelo governo. Não há, na legislação, mecanismo que impeça a solicitação de cadastro para se ter acesso aos dados. No Decreto 7.724 é definido o conceito de Transparência Ativa, a publicação de dados independente de requerimentos, mas não independente de cadastro. Nesse decreto também são regulamentados os Serviços de Informação ao Cidadão, onde o cidadão pode solicitar qualquer informação dos órgãos públicos sem precisar explicar o motivo, mas para isso deverá se identificar fornecendo seu nome, identificação pessoal válida e endereço pessoal ou eletrônico, contrariando assim o princípio *não-discriminatório*, ou seja, esse princípio não é coberto por essa proposta, pois exigir um cadastro para acesso aos dados é uma decisão política que independe das tecnologias adotadas.

Um problema bem conhecido na Web é a quebra de *link*. Endereços Web podem deixar de existir por desinteresse de continuidade de um domínio por seu

proprietário, troca ou atualização de sistema, reorganização do site, entre outros motivos. O mesmo pode acontecer com a ligação entre dados, enfraquecendo a possibilidade de integração entre esses dados. Para prevenir tal problema deve-se adotar uma política de identificador único permanente, que consiste na garantia de existência de um endereço eletrônico ao longo do tempo para um mesmo conjunto de dados. Nesse contexto, inserem-se dois requisitos: usar "*identificador único para dados*" e "*ligar dados a outros dados*".

Uma lista de dados tem uma forte dependência de identificadores únicos permanentes, pois se eles deixam de existir, o catálogo perde valor. Na proposta deste trabalho, é sugerida uma forma de endereçamento Web compatível com a própria organização dos órgãos na execução do orçamento, que é pela Unidade Orçamentária, uma entidade física que deverá existir indefinidamente no orçamento de um determinado ano, pois o orçamento público é regulamentado por lei, ou seja, por um documento histórico.

Dados sobre a execução do orçamento público representam o resultado de um conjunto de ações que foi realizado num determinado período, portanto é um documento histórico da administração pública. Preservar esse documento permite estudos para o aprimoramento da própria gestão pública. Preservar séries históricas pode ajudar na identificação de padrões que levem a boas práticas no uso do dinheiro público.

Para a preservação de séries históricas é necessário o uso de metadados de nível de coleção para permitir que sistemas de informação classifiquem, organizem, agrupem dados de milhares de entes federativos no Brasil. É como considerar os dados como obras de arte que devem ser preservadas para conhecimento das gerações futuras. Dados devem ser colecionáveis pelos metadados anexados aos próprios dados. Informações de proveniência dos dados também são importantes para se recuperar a história dos dados, sua origem, criação, acesso. Para atender os requisitos "*dados devem ter informações de proveniência*" e "*dados devem ter metadados*". Nesta proposta, foi utilizado o conjunto de metadados Dublin Core,

conforme estudos de Powell (2009)[25] e Hartig(2009)[34], que pode ser utilizado para o colecionamento de dados e recuperação de informação de proveniência dos dados.

”*Dados devem ser permanentes*” é um requisito que depende de recursos materiais para suportar a permanência da dados na Web. No entanto, a ideia de permanência leva à discussão sobre resiliência de princípios. Dados baixados sem um conjunto de metadados ao serem transferidos dentro da sociedade civil podem perder informações importantes, por exemplo, o primeiro cidadão que coletar dados saberá sob qual licença esses estavam publicados, mas quem copiar dele pode não ter contato com essa informação se ela não existir junto aos dados. Essa ideia é similar à venda de unidades de CDs onde cada unidade tem informações de direitos autorais marcado na embalagem. Portanto, cada porção de dados fornecida deve conter metadados para preservar os princípios ao longo do tempo e das cópias. Assim, metadados também devem ser permanentes, para que o dados não percam valor.

”*Vocabulários controlados*” são combinados na proposta deste trabalho e suas versões estão definidas no próprio endereço do vocabulário. O ”*versionamento*” de metadados, vocabulários e taxonomia permitem a decisão de como esses vocabulários devem ser processados. No exemplo do Anexo A, a versão de metadados Dublin Core utilizada é a 1.1 e qualquer sistema automatizado que processe essa versão pode colecionar os dados publicados na Web, mesmo que esse sistema ainda não interprete a taxonomia do orçamento.

Neste trabalho, foram percebidos três níveis do processamento por máquina: tratamento de dados, validação de dados e colecionamento. O tratamento de dados e validação são suportados pelo Taxonomia XBRL por conter as informações sobre os elementos de dados, o colecionamento de dados é suportado pelos elementos de metadados. Tais possibilidades de processamento por máquina demonstram que se deve investir em formatos que suportem metadados e taxonomia que também sejam processáveis por máquina, o que não ocorre com o formato CSV encontrado

na publicação de dados de execução orçamentária no Brasil.

Este trabalho foi focado em informações de execução de orçamento público, no entanto, os requisitos aqui apresentados podem ser aplicados a outros tipos de dados, pois os padrões utilizados na proposta independem da estrutura de dados e do tipo de informação. A taxonomia que descreve os fatos pode ser elaborada de diversas formas, mesmo que o XBRL costuma ser associado a informações contábeis, não há restrição que impeça a representação de outro tipo de informação, como relatórios educacionais.

A Figura 17 contém um modelo da proposta XBRL onde é mostrado como os dados são divididos.

XBRL é um padrão robusto que tem os recursos necessários para problemas simples e complexos. Pela experiência da implementação, pode-se fazer um uso simplificado do XBRL para representar dados abertos como no caso desta proposta de publicação. Um manual direcionado para esse uso, com um subconjunto da documentação do XBRL para publicação de dados abertos, pode auxiliar na adoção dessa especificação pelos órgãos públicos para outros propósitos.

Outro aspecto levantado durante a implementação é a possibilidade de ligação entre dados e sua possível correspondência na visualização. Investigações sobre como essas ligações podem integrar dados na Web ainda precisam ser realizadas para se conhecer como tais ligações podem agregar valor aos dados de forma similar a abordagem Linked Data.

Neste trabalho optou-se por focar os dados publicados. Sabe-se que existe uma história interna do governo de como esses dados foram coletadas. Portanto, os metadados são utilizados para recuperar a história de publicação dos dados. Não fez parte deste trabalho, analisar toda qualidade da informação desde sua gênese, mas permitir que mediante os dados sejam descobertos os responsáveis pela criação e publicação de dados. Assim, quem desejar mais informações sobre os dados poderá descobrir os responsáveis pela publicação e questionar sobre qualquer

fonte: elaborado pelo autor

Figura 17 – Modelo da proposta XBRL

outra informação desejada sobre os dados.

Dos dezenove requisitos levantados na legislação e literatura, onze têm correspondência na lei, sete não têm e para um requisito na lei não foi encontrado correspondente na literatura. Destes requisitos, dezesseis são referentes a apresentação de dados ao usuário, sendo que catorze são atendidos pela proposta de publicação deste trabalho, pois a permanência depende de recurso para manter um serviço de publicação disponível à sociedade, e a possibilidade de ligação entre dados precisa ser mais estudada para se considerar que a proposta deste trabalho atenda a esse requisito.

5.1 Comparação com trabalhos relacionados

A proposta deste trabalho foi desenvolvida a partir do caso específica da publicação de dados da execução do orçamento público em XBRL. Garcia e Gil (2009)[40] investigaram a publicação de dados como Linked Data a partir da instância e taxonomia ambas em XBRL. Consequentemente, a proposta de publicação deste trabalho pode ser republicada como Linked Data para uso na Web Semântica. Pensando na evolução apresentada na Figura 2 significa estar preparado para as tecnologias que serão utilizadas em 2020. Assim, esta proposta e o trabalho de Garcia e Gil (2009)[40] são complementares, compondo uma solução maior. Essa visão corrobora com O’Riain *et al.* (2012)[39] na formação de um ecossistema de dados, governamentais, devido a publicação utilizando o padrão XBRL.

Fernandes e Campos (2008)[41] propuseram uma central de balanços no Brasil com o uso de XBRL. Como nesta proposta os dados tratados são governamentais, os dois trabalhos se complementam numa solução de central de balanços para o governo, que também permitiriam estudos setoriais, análise de risco e geração de estatísticas da administração pública. De fato, com a proposta deste visou uma publicação com mais qualidade de dados, orientada para Web Semântica

num contexto sem um padrão de publicação, os trabalhos relacionados somente poderiam ser passíveis de propostas de integração.

5.2 Contribuições da proposta

Neste trabalho foi desenvolvido uma proposta de publicação que incluiu dados, metadados e uma taxonomia para publicação de dados do orçamento público. Considerando que a prática nos portais de transparência é publicar dados em CSV, segundo estudo de Craveiro *et al.* (2013)[4], a taxonomia XBRL e os metadados selecionados são contribuições imediatas sobre a prática brasileira.

Com a taxonomia XBRL passa a ser possível validação automática dos dados, evitando alguns tipos de erros de publicação, como a falta de dados e erros textuais encontrados nos dados do orçamento do Estado de São Paulo. Os metadados selecionados e integrados na taxonomia XBRL permitem colecionamento de dados automático, ou seja, agregação de séries históricas, identificação de proviniência de dados e agregação por qualquer informação dos metadados, levando a uma gestão dos dados por esses elementos.

Outra contribuição é um padrão de publicação que pode ser usado por qualquer ente da federação, passível de ser estendido para qualquer particularidade graças as possibilidades de extensibilidade do padrão XML Schema, utilizado na taxonomia XBRL. Dessa forma, todas as publicações de orçamento podem ter uma base comum, permitindo a correlação entre orçamentos públicos e abrindo caminho para outros estudos.

5.3 Trabalhos futuros

Neste trabalho foram utilizados dados do orçamento do Estado de São Paulo para validar a publicação. Nos trabalhos relacionados e contribuições da proposta verificam-se outros estudos a realizar para validar esta proposta na perspectiva de

Linked Data e de uma central de balanço da administração pública.

Os metadados dessa proposta também podem ser utilizados na pesquisa por novas hipóteses[24] num grande volume dados colecionados por máquina. Dados de um ente isoladamente pode ser visto com um pequeno volume de dados, mas integrado num ecossistema de dados com XBRL ou republicado como Linked Data pode levar a estudos com enormes volumes de dados, também chamado de Bid Data.

A partir do padrão implementado nesta proposta outras implementações são necessárias para suporte a pesquisa, análise de dados e publicação como Linked Data. A seguir um lista de alguns possíveis estudos a partir desta proposta:

- extensibilidade do padrão da proposta para necessidade de outros entes;
- desenvolvimento de colecionadores de dados automáticos usando metadados;
- republicação de dados de orçamento como Linked Data;
- ligação de dados de orçamento a outros dados governamentais abertos;
- desenvolvimento de ecossistema de dados governamentais abertos baseados em XBRL;
- geração de relatórios automatizados a partir da publicação em XBRL;
- mecanismo de tratamento e análise de dados usando metadados;
- auditoria automatizada de execução de orçamento público;
- reconhecimento de padrões de execução de orçamento público;
- impacto de reuso de dados pela sociedade.

A integração desta proposta com os trabalhos relacionados também proporcionam outros estudos. Além disso a apropriação de dados da execução do orçamento

com metadados, taxonomia e outros elementos inexistentes na prática brasileira pode levar a um aprimoramento nas aplicações de desenvolvedores na sociedade, consequentemente, novas soluções e estudos.

A Figura 18 exemplifica dois cenários possíveis de publicação de dados. No cenário (a) o governo publica dados no formato CSV, para os consumidores de dados visualizarem como tabela. No cenário (b) os dados são fornecidos com a estrutura que existe em seus sistemas internos no padrão XBRL. Dados publicados em XML, o que inclui XBRL, podem ser convertidos em OWL com a linguagem XSLT (Antoniou e Harmelen, 2008:241); OWL é um formato, também baseado em XML, recomendado pelo W3C para representar ontologias. Da mesma maneira, os próprios consumidores de dados podem converter dados em XBRL para outros formatos como HTML, para visualizações pelo público em geral, JSON sistemas que entendem esse formato, RDF para ser processado por máquinas semânticas, e CSV para os que preferirem ver uma tabela, ou seja, diversos públicos podem consumir dados na forma que desejar, ampliando o valor de (re)uso dos dados governamentais. A proposta de publicação deste trabalho é alinhada com o cenário (b).

fonte: elaborado pelo autor

Figura 18 – Conversões de formatos com XSLT

Madnick *et al.*(2009)[22] ressaltam a importância de profissionais de outras áreas entrarem na discussão sobre qualidade de dados. Esta proposta estabelece um ponto de partida para discussão da qualidade de DGA, incluindo a qualidade para algumas formas de processamento por máquina. Para se construir um repositórios de dados com captura automática são necessários localizadores permanentes (PURL) e metadados de coleção e proveniência. Para validação automática é preciso uma taxonomia publicada na Web, para integridade uma assinatura. Dessa forma o modelo desta proposta é o princípio de discussão de um modelo de qualidade de dados governamentais.

Capítulo 6

Considerações finais

Neste trabalho foi discutido a falta de um padrão de publicação de dados de orçamento público pelos entes da federação no Brasil e a necessidade de se pensar uma publicação para a Web que proporcionasse mais informação para a sociedade exercer a participação no governo e o controle social. Para tanto foram levantados requisitos da legislação brasileira, e a partir da literatura de Dados Governamentais Abertos e Linked Data. Cada requisito com relacionado a um modelo de sistema de informação para identificar quais desses requisitos devem ser atendidos na perspectiva do consumidor de dados, ou seja, dos dados na Web. Posteriormente, foram investigadas as possibilidades técnicas de implementação dos requisitos para a implementação do protótipo.

Foram apresentadas questões sobre dados como qualidade de dados, metadados e taxonomia. Uma questão emergente foi a publicação de dados como obras históricas que devem ser colecionadas para gerações futuras. Assim, verificou-se que além da necessidade de dados processáveis por máquinas, metadados e taxonomia também precisavam ser processáveis por máquina.

Dos dezesseis requisitos referentes a apresentação de dados ao usuário, catorze são atendidos pela implementação da proposta de publicação deste trabalho, pois a permanência dos dados na Web depende exclusivamente da administração pública e a validação da ligação entre dados precisa de uma infraestrutura maior que a utilizada no desenvolvimento deste trabalho, com mais conjuntos de dados publicados. A proposta de publicação resultante além de alinhada com os requisitos estudados prevê processamento por máquina em três níveis: tratamento de dados, validação de dados e colecionamento de conjuntos de dados.

A maioria dos requisitos têm correspondência na legislação brasileira, mais a prática de publicação está aquém desses requisitos. A publicação de dados, metadados e taxonomia processáveis por máquina permitiria fiscalização automatizada da qualidade dos dados publicados. Para isso é preciso avançar nas questões levantadas da literatura neste trabalho. A padronização da publicação com tecnologias Web, extensível para necessidades de qualquer ente da federação, é um caminho para diminuir a desigualdade na publicação e permitir a interoperabilidade e integração de dados sem a necessidade de milhares de implementações de sistemas de captura de dados. De fato, uma mesma solução poderia ser compartilhada por diversos entes, reduzindo o custo global nas despesas públicas.

As organizações XBRL International, Dublin Core Metada Initiative e W3C desenvolveram respectivamente o padrão XBRL para relatórios empresariais, o conjunto Dublin Core de metadados e a recomendação XSLT para conversões baseadas em XML. Esses padrões, todos abertos, foram utilizadas neste trabalho de modo que publicadores de dados possam fornecer os mesmos dados em outros formatos de forma automática, através das alternativas por XSLT. A implementação observou-se que esses padrões são bastante maduros e atendem as necessidades atuais de publicação de dados governamentais abertos.

Esta proposta de publicação abre caminho para novos estudos sobre formatos processáveis por máquina, novos sistemas, qualidade de dados governamentais, sendo sugerido ao longo do texto o estudo de como formar uma rede de dados com as possibilidades de ligação da proposta deste trabalho. Um formato que pode ser republicado de outras formas automaticamente, auxiliará no fornecimento de dados para máquinas semânticas, e é uma importante abordagem para o futuro da Web e em como os usuários irão interagir com esses dados. Nesse sentido, os metadados propostos neste trabalho têm grande importância para o (re)uso dos dados em sistemas de informação.

ANEXO A

Instância em XBRL

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <!-- conversões pré-definidas , por padrão os navegadores vão
   escolher XHTML -->
3 <?xml-stylesheet type="text/xsl" href="xhtml.xsl" title="XHTML" ?>
4 <?xml-stylesheet type="text/xsl" href="csv.xsl" title="CSV"
   alternate="yes" ?>
5 <?xml-stylesheet type="text/xsl" href="json.xsl" title="JSON"
   alternate="yes" ?>
6 <!-- relatório XBRL -->
7 <!-- xmlns (XML Namespace) descreve o caminho dos padrões/
   vocabulários para o documento XML -->
8 <xbrli:xbrl
9   xmlns="http://www.cgu.gov.br/orcamento/2013"
10  xmlns:dc="http://purl.org/dc/elements/1.1/"
11  xmlns:xbrli="http://www.xbrl.org/2003/instance"
12  xmlns:iso4217="http://www.xbrl.org/2003/iso4217"
13  xmlns:link="http://www.xbrl.org/2003/linkbase"
14  xmlns:xlink="http://www.w3.org/1999/xlink"
15  targetNamespace="http://www.cgu.gov.br/orcamento/2013">
16  <link:schemaRef xlink:type="simple" xlink:href="http://www.cgu.
   gov.br/orcamento/2013/esquema.xsd"/>
17  <!-- contexto dos fatos -->
18  <xbrli:context id="cx1">
19    <xbrli:entity scheme="http://orcamento.orgao.uf.gov.br/">

```

```

20  <xbrli:identifier scheme="http://orcamento.orgao.uf.gov.br/">
    Página de Publicação de Orçamento do Órgão</
    xbrli:identifier>
21 </xbrli:entity>
22 <xbrli:period>
23   <xbrli:startDate>2012-01-01</xbrli:startDate>
24   <xbrli:endDate>2012-12-31</xbrli:endDate>
25 </xbrli:period>
26 <xbrli:scenario>
27   <dc:dcterms>
28     <dc:type>Dataset</dc:type>
29     <dc:title>Orçamento do Governo Federal</dc:title>
30     <dc:creator>Controladoria Geral da União</dc:creator>
31     <dc:publisher>Governo Eletrônico</dc:publisher>
32     <dc:created>2012-01-02</dc:created>
33     <dc:modified>2013-08-18</dc:modified>
34     <dc:source>http://orcamento.orgao.uf.gov.br/2013/uo252</
        dc:source>
35     <dc:license>Constituição de 1988, Art 5º; Lei 12.527 de 2011;
        Decreto 7.724 de 2012</dc:license>
36     <dc:provenance>download de arquivo</dc:provenance>
37   </dc:dcterms>
38 </xbrli:scenario>
39 </xbrli:context>
40 <!-- unidade monetária -->
41 <xbrli:unit id="brl">
42   <xbrli:measure>iso4217:BRL</xbrli:measure>
43 </xbrli:unit>
44 <!-- fatos da execução do orçamento -->
45 <orcamento>
46   <receita>
47     <categoria id="ct1">
48       <nome>Receita Corrente</nome>
49       <origem id="or1">
50         <nome>Receita Tributária</nome>
51         <especie id="ep1">

```



```
52      <nome>Impostos</nome>
53      <rubrica id="rb2">
54        <nome>Impostos sobre o Patrimônio e a Renda</nome>
55        <alinea id="al4">
56          <nome>Impostos sobre a Renda e Proventos de Qualquer
            Natureza</nome>
57          <subalinea id="sa10">
58            <nome>Pessoas Físicas</nome>
59            <fonte id="fd100">
60              <nome>Recursos do Tesouro – Exercício Corrente; Recursos
                Ordinários</nome>
61            </fonte>
62            <resultado id="rs0">
63              <nome>P</nome>
64            </resultado>
65            <prevista>
66              <data contextRef="cx1">2012-02-01</data>
67              <valor contextRef="cx1" unitRef="br1">1000.5</valor>
68            </prevista>
69            <lancada>
70              <data contextRef="cx1">2012-03-01</data>
71              <valor contextRef="cx1" unitRef="br1">1000.5</valor>
72            </lancada>
73            <recolhida>
74              <data contextRef="cx1">2012-04-01</data>
75              <valor contextRef="cx1" unitRef="br1">1000.5</valor>
76            </recolhida>
77          </subalinea>
78        </alinea>
79      </rubrica>
80    </especie>
81  </origem>
82  </categoria>
83  </receita>
84  <despesa>
85    <esfera id="ef10">
```

<nome>Orçamento Fiscal</nome>
 <orgao id="og39">
 <nome>Ministério dos Transportes</nome>
 <uo id="uo252">
 <nome>Departamento Nacional de Infraestrutura de Transportes
 – DNIT</nome>
 <ug id="ug252">
 <nome>Departamento Nacional de Infraestrutura de
 Transportes – DNIT</nome>
 </ug>
 <funcao id="fn26">
 <nome>Transporte</nome>
 <subfuncao id="sf782">
 <nome>Transporte Rodoviário</nome>
 <programa id="pg2075">
 <nome>Transpores Rodoviário</nome>
 <objetivo>O que será feito?</objetivo>
 <iniciativa>O que será entregue?</iniciativa>
 <acao id="ac7M64">
 <nome>Construção de Trecho Rodoviário</nome>
 <descricao>O que é feito? Para que é feito?</descricao>
 <implementacao>Como é feito?</implementacao>
 <produto>Qual o resultado?</produto>
 <unidade>Como mensurar?</unidade>
 <subtitulo id="sb35">
 <nome>Rio Grande do Sul</nome>
 </subtitulo>
 <idoc id="io9999">
 <nome>Outros recursos</nome>
 </idoc>
 <iduso id="iu0">
 <nome>Recursos não destinados à contrapartida</nome>
 </iduso>
 <fonte id="fr100">
 <nome>Recursos do Tesouro – Exercício Corrente;
 Recursos Ordinários</nome>

```
119      </fonte>
120      <natureza id="nt4490">
121        <nome>Despesas de Capital; Investimentos; Aplicação
          Direta</nome>
122      </natureza>
123      <elemento id="el51">
124        <nome>Obras e Instalações</nome>
125      </elemento>
126      <resultado id="rt2">
127        <nome>Primária Discricionária</nome>
128      </resultado>
129      <orcado>
130        <data contextRef="cx1">2012-01-01</data>
131        <valor contextRef="cx1" unitRef="brl">1000.5</valor>
132      </orcado>
133      <licitacao id="li12345">
134        <numero>12345</numero>
135        <bemServico>bem ou serviço fornecido</bemServico>
136        <credor>pessoa física ou jurídica beneficiária do
          pagamento</credor>
137        <procedimento>tipo de licitação</procedimento>
138        <empenhado>
139          <data contextRef="cx1">2012-02-01</data>
140          <valor contextRef="cx1" unitRef="brl">1000.5</valor>
141        </empenhado>
142        <liquidado>
143          <data contextRef="cx1">2012-08-01</data>
144          <valor contextRef="cx1" unitRef="brl">1000.5</valor>
145        </liquidado>
146        <pago>
147          <data contextRef="cx1">2012-10-01</data>
148          <valor contextRef="cx1" unitRef="brl">1000.5</valor>
149        </pago>
150      </licitacao>
151    </acao>
152  </programa>
```

```
153         </subfuncao>
154     </funcao>
155 </uo>
156 </orgao>
157 </esfera>
158 </despesa>
159 </orcamento>
160 </xbrli:xbrl>
```

ANEXO B

Taxonomia em XBRL

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <xs:schema elementFormDefault="qualified"
3   xmlns:xs="http://www.w3.org/2001/XMLSchema"
4   xmlns="http://www.cgu.gov.br/orcamento/2013"
5   xmlns:dc="http://purl.org/dc/elements/1.1/"
6   xmlns:xbrli="http://www.xbrl.org/2003/instance"
7   xmlns:iso4217="http://www.xbrl.org/2003/iso4217"
8   xmlns:link="http://www.xbrl.org/2003/linkbase"
9   xmlns:xlink="http://www.w3.org/1999/xlink"
10  targetNamespace="http://www.cgu.gov.br/orcamento/2013">
11  <xs:import namespace="http://purl.org/dc/elements/1.1/"
12    schemaLocation="dc.xsd"/>
13  <xs:import namespace="http://www.w3.org/1999/xlink"
14    schemaLocation="xlink.xsd"/>
15  <xs:import namespace="http://www.xbrl.org/2003/instance"
16    schemaLocation="xbrli.xsd"/>
17  <xs:import namespace="http://www.xbrl.org/2003/linkbase"
18    schemaLocation="link.xsd"/>
19  <xs:element name="orcamento">
20    <xs:complexType>
21      <xs:choice minOccurs="1" maxOccurs="2">
22        <xs:element ref="receita"/>
23        <xs:element ref="despesa"/>
24      </xs:choice>

```

```
21 </xs:complexType>
22 </xs:element>
23 <xs:element name="receita">
24   <xs:complexType>
25     <xs:sequence>
26       <xs:element ref="categoria" maxOccurs="unbounded" />
27     </xs:sequence>
28   </xs:complexType>
29 </xs:element>
30 <xs:element name="categoria">
31   <xs:complexType>
32     <xs:complexContent>
33       <xs:extension base="nome">
34         <xs:sequence>
35           <xs:element ref="origem" maxOccurs="unbounded" />
36         </xs:sequence>
37         <xs:attribute name="id" use="required" type="xs:NCName" />
38       </xs:extension>
39     </xs:complexContent>
40   </xs:complexType>
41 </xs:element>
42 <xs:element name="origem">
43   <xs:complexType>
44     <xs:complexContent>
45       <xs:extension base="nome">
46         <xs:sequence>
47           <xs:element ref="especie" maxOccurs="unbounded" />
48         </xs:sequence>
49         <xs:attribute name="id" use="required" type="xs:NCName" />
50       </xs:extension>
51     </xs:complexContent>
52   </xs:complexType>
53 </xs:element>
54 <xs:element name="especie">
55   <xs:complexType>
56     <xs:complexContent>
```

```

57     <xs:extension base="nome">
58         <xs:sequence>
59             <xs:element ref="rubrica" maxOccurs="unbounded" />
60         </xs:sequence>
61         <xs:attribute name="id" use="required" type="xs:NCName" />
62     </xs:extension>
63 </xs:complexContent>
64 </xs:complexType>
65 </xs:element>
66 <xs:element name="rubrica">
67     <xs:complexType>
68         <xs:complexContent>
69             <xs:extension base="nome">
70                 <xs:sequence>
71                     <xs:element ref="alinea" maxOccurs="unbounded" />
72                 </xs:sequence>
73                 <xs:attribute name="id" use="required" type="xs:NCName" />
74             </xs:extension>
75         </xs:complexContent>
76     </xs:complexType>
77 </xs:element>
78 <xs:element name="alinea">
79     <xs:complexType>
80         <xs:complexContent>
81             <xs:extension base="nome">
82                 <xs:sequence>
83                     <xs:element ref="subalinea" />
84                 </xs:sequence>
85                 <xs:attribute name="id" use="required" type="xs:NCName" />
86             </xs:extension>
87         </xs:complexContent>
88     </xs:complexType>
89 </xs:element>
90 <xs:element name="subalinea">
91     <xs:complexType>
92         <xs:complexContent>

```

```
93     <xs:extension base="nome">
94       <xs:sequence>
95         <xs:element ref="fonte" />
96         <xs:element ref="resultado" />
97         <xs:choice maxOccurs="unbounded">
98           <xs:element ref="prevista" />
99           <xs:element ref="lancada" />
100          <xs:element ref="recolhida" />
101        </xs:choice>
102      </xs:sequence>
103      <xs:attribute name="id" use="required" type="xs:NCName" />
104    </xs:extension>
105  </xs:complexContent>
106 </xs:complexType>
107 </xs:element>
108 <xs:element name="prevista">
109   <xs:complexType>
110     <xs:sequence>
111       <xs:element ref="data" />
112       <xs:element ref="valor" />
113     </xs:sequence>
114   </xs:complexType>
115 </xs:element>
116 <xs:element name="lancada">
117   <xs:complexType>
118     <xs:sequence>
119       <xs:element ref="data" />
120       <xs:element ref="valor" />
121     </xs:sequence>
122   </xs:complexType>
123 </xs:element>
124 <xs:element name="recolhida">
125   <xs:complexType>
126     <xs:sequence>
127       <xs:element ref="data" />
128       <xs:element ref="valor" />
```



```
129     </xs:sequence>
130   </xs:complexType>
131 </xs:element>
132 <xs:element name="despesa">
133   <xs:complexType>
134     <xs:sequence>
135       <xs:element ref="esfera" maxOccurs="unbounded" />
136     </xs:sequence>
137   </xs:complexType>
138 </xs:element>
139 <xs:element name="esfera">
140   <xs:complexType>
141     <xs:complexContent>
142       <xs:extension base="nome">
143         <xs:sequence>
144           <xs:element ref="orgao" maxOccurs="unbounded" />
145         </xs:sequence>
146         <xs:attribute name="id" use="required" type="xs:NCName" />
147       </xs:extension>
148     </xs:complexContent>
149   </xs:complexType>
150 </xs:element>
151 <xs:element name="orgao">
152   <xs:complexType>
153     <xs:complexContent>
154       <xs:extension base="nome">
155         <xs:sequence>
156           <xs:element ref="uo" maxOccurs="unbounded" />
157         </xs:sequence>
158         <xs:attribute name="id" use="required" type="xs:NCName" />
159       </xs:extension>
160     </xs:complexContent>
161   </xs:complexType>
162 </xs:element>
163 <xs:element name="uo">
164   <xs:complexType>
```

```
165 <xs:complexContent>
166   <xs:extension base="nome">
167     <xs:sequence>
168       <xs:element ref="ug" minOccurs="0" />
169       <xs:element ref="funcao" maxOccurs="unbounded" />
170     </xs:sequence>
171     <xs:attribute name="id" use="required" type="xs:NCName" />
172   </xs:extension>
173 </xs:complexContent>
174 </xs:complexType>
175 </xs:element>
176 <xs:element name="ug">
177   <xs:complexType>
178     <xs:complexContent>
179       <xs:extension base="nome">
180         <xs:attribute name="id" use="required" type="xs:NCName" />
181       </xs:extension>
182     </xs:complexContent>
183   </xs:complexType>
184 </xs:element>
185 <xs:element name="funcao">
186   <xs:complexType>
187     <xs:complexContent>
188       <xs:extension base="nome">
189         <xs:sequence>
190           <xs:element ref="subfuncao" maxOccurs="unbounded" />
191         </xs:sequence>
192         <xs:attribute name="id" use="required" type="xs:NCName" />
193       </xs:extension>
194     </xs:complexContent>
195   </xs:complexType>
196 </xs:element>
197 <xs:element name="subfuncao">
198   <xs:complexType>
199     <xs:complexContent>
200       <xs:extension base="nome">
```

```

201     <xs:sequence>
202       <xs:element ref="programa" maxOccurs="unbounded" />
203     </xs:sequence>
204     <xs:attribute name="id" use="required" type="xs:NCName" />
205   </xs:extension>
206 </xs:complexContent>
207 </xs:complexType>
208 </xs:element>
209 <xs:element name="programa">
210   <xs:complexType>
211     <xs:complexContent>
212       <xs:extension base="nome">
213         <xs:sequence>
214           <xs:element ref="objetivo" />
215           <xs:element ref="iniciativa" />
216           <xs:element ref="acao" maxOccurs="unbounded" />
217         </xs:sequence>
218         <xs:attribute name="id" use="required" type="xs:NCName" />
219       </xs:extension>
220     </xs:complexContent>
221   </xs:complexType>
222 </xs:element>
223 <xs:element name="objetivo" type="xs:string" />
224 <xs:element name="iniciativa" type="xs:string" />
225 <xs:element name="acao">
226   <xs:complexType>
227     <xs:complexContent>
228       <xs:extension base="nome">
229         <xs:sequence>
230           <xs:element ref="descricao" minOccurs="0" />
231           <xs:element ref="implementacao" minOccurs="0" />
232           <xs:element ref="produto" minOccurs="0" />
233           <xs:element ref="unidade" minOccurs="0" />
234           <xs:element ref="subtitulo" minOccurs="0" />
235           <xs:element ref="idoc" />
236           <xs:element ref="iduso" />

```

```
237     <xs:element ref="fonte" />
238     <xs:element ref="natureza" />
239     <xs:element ref="elemento" />
240     <xs:element ref="resultado" />
241     <xs:element ref="orcado" maxOccurs="unbounded" />
242     <xs:element ref="licitacao" maxOccurs="unbounded" />
243   </xs:sequence>
244   <xs:attribute name="id" use="required" type="xs:NCName" />
245 </xs:extension>
246 </xs:complexContent>
247 </xs:complexType>
248 </xs:element>
249 <xs:element name="descricao" type="xs:string" />
250 <xs:element name="implementacao" type="xs:string" />
251 <xs:element name="produto" type="xs:string" />
252 <xs:element name="unidade" type="xs:string" />
253 <xs:element name="subtitulo">
254   <xs:complexType>
255     <xs:complexContent>
256       <xs:extension base="nome">
257         <xs:attribute name="id" use="required" type="xs:NCName" />
258       </xs:extension>
259     </xs:complexContent>
260   </xs:complexType>
261 </xs:element>
262 <xs:element name="idoc">
263   <xs:complexType>
264     <xs:complexContent>
265       <xs:extension base="nome">
266         <xs:attribute name="id" use="required" type="xs:NCName" />
267       </xs:extension>
268     </xs:complexContent>
269   </xs:complexType>
270 </xs:element>
271 <xs:element name="iduso">
272   <xs:complexType>
```

```
273 <xs:complexContent>
274   <xs:extension base="nome">
275     <xs:attribute name="id" use="required" type="xs:NCName"/>
276   </xs:extension>
277 </xs:complexContent>
278 </xs:complexType>
279 </xs:element>
280 <xs:element name="natureza">
281   <xs:complexType>
282     <xs:complexContent>
283       <xs:extension base="nome">
```

ANEXO C

Glossário

fonte: Wikipédia wikipedia.org

CSV - é acrônimo de Comma-Separated Values, um formato de arquivo de computador que pode conter valores separados por algum delimitador, ponto e vírgula (;) por exemplo, pode ser criado em qualquer editor de texto e lido em uma planilha de textos, onde cada linha/coluna será as linhas do arquivo , separados por ; (ponto-e-vírgula).

Dublin Core - esquema de metadados que visa descrever objetos digitais, tais como, videos, sons, imagens, textos e sites na web, desenvolvido pela Dublin Core Metadata Initiative (DCMI) (Iniciativa Dublin Core Metadados), uma organização dedicada a promover a adoção de padrões de interoperabilidade de metadados e desenvolver vocabulários especializados para descrever fontes que tornem mais inteligentes sistemas de descobrimento de informações.

JSON - é acrônimo de JavaScript Object Notation, um formato leve para intercâmbio de dados computacionais.

PDF - é acrônimo de Portable Document Format, um formato de arquivo, desenvolvido pela Adobe Systems em 1993, para representar documentos de maneira independente do aplicativo, do hardware e do sistema operacional usados para criá-los

PURL - é um acrônimo de Persistente Uniform Resource Locator (URL) que é usado para redirecionar ou localizar um recurso da Web solicitado, podendo ser páginas Web ou conjuntos de dados

XBRL - é acrônimo de eXtensible Business Reporting Language, um padrão no XML para relatórios de negócios. Foi criado por um consórcio internacional sem fins lucrativos (XBRL International Incorporated) de mais de 300 organizações, entre entidades reguladoras, agências governamentais e empresas de software.

XML - é acrônimo de eXtensible Markup Language, uma recomendação da W3C para gerar linguagens de marcação para necessidades especiais. Seu propósito principal é a facilidade de compartilhamento de informações através da internet.

XML Signature - também chamado XMLDSig, XML-DSig ou XML-Sig, é uma recomendação W3C que define uma sintaxe XML para assinaturas digitais, permitindo a verificação de integridade.

XSLT - é acrônimo de eXtensible Stylesheet Language for Transformation - linguagem extensível para folhas de estilo de transformações), é uma linguagem de marcação XML usada para criar documentos que definem a apresentação dos documentos XML nos navegadores outros aplicativos que a suportem.

Referências

- 1 OPEN GOVERNMENT PARTNERSHIP. *Declaração de Governo Aberto*. Set. 2011. 2–3 p. Disponível em: <http://www.opengovpartnership.org/sites/www.opengovpartnership.org/files/page_files/Declaracao_de_Governo_Aberto.pdf>.
- 2 CONTROLADORIA GERAL DA UNIÃO. *Controle Social: Orientações aos cidadãos para participação na gestão pública e exercício do controle social, Coleção Olho Vivo*. Brasília: [s.n.], 2010. 12,16 p. Disponível em: <<http://www.cgu.gov.br/Publicacoes/CartilhaOlhoVivo/Arquivos/ControleSocial.pdf>>.
- 3 CÔMITE GESTOR DA INTERNET. *Pesquisa sobre o uso das tecnologias da informação e da comunicação no Brasil: 2005-2009*. São Paulo: [s.n.], 2010. 76 p. Disponível em: <<http://www.cetic.br/tic/5anos/index.htm>>.
- 4 CRAVEIRO, G. S.; SANTANA, M. T.; ALBUQUERQUE, J. P. Assessing open government budgetary data in brazil. In: *ICDS 2013, The Seventh International Conference on Digital Society*. [s.n.], 2013. p. 20–27. Disponível em: <http://www.thinkmind.org/download.php?articleid=icds_2013_1_40_10183>.
- 5 BRASIL; MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO. *Modelo de Acessibilidade em Governo Eletrônico - e-MAG*. Ago. 2011. Disponível em: <<http://emag.governoeletronico.gov.br/emag/>>.
- 6 BERTOT, J. C.; JAEGER, P. T.; GRIMES, J. M. Using icts to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, Elsevier, v. 27, n. 3, p. 264–271, 2010.
- 7 BRASIL. *Lei Complementar nº 131*. Mai. 2009. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp131.htm>.
- 8 CORRÊA, I. A experiência brasileira na promoção da transparência. In: *Seminário Internacional sobre Acesso à Informação: Desafios de Implementação*. Brasília: CGU, 2010. Disponível em: <<http://www.cgu.gov.br/acessoainformacao/eventos/seminario-internacional/>>

arquivos/Izabela-Correa-A-experiencia-brasileira-na-promocao-da-\transparencia-ativa-1.pdf>.

9 BRASIL. *Lei Complementar nº 101*. Mai. 2000. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm>.

10 W3C ESCRITÓRIO BRASIL. *Melhorando o acesso ao governo com o melhor uso da web*. São Paulo: Comitê Gestor da Internet no Brasil, 2009. 28–29,39,46–47,52,54–55 p. Disponível em: <<http://www.w3c.br/pub/Materiais/PublicacoesW3C/gov-web.pdf>>.

11 MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO. *Padrões de Interoperabilidade de Governo Eletrônico - e-PING*. Nov. 2012. 4 p. Disponível em: <<http://www.governoeletronico.gov.br/biblioteca/arquivos/documento-da-e-ping-versao-2013/>>.

12 BAUER, F.; KALTENBÖCK, M. *Linked Open Data: The Essentials*. Viena: edition mono, 2012. 4,23 p. Disponível em: <<http://www.semantic-web.at/LOD-TheEssentials.pdf>>.

13 BRASIL; MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO. *Manual técnico de orçamento MTO - Versão 2013*. 2012. 14,17–19,33 p. Disponível em: <http://www.orcamentofederal.gov.br/informacoes-orcamentarias/manual-tecnico/MTO_2013_110UT2012_terceira-versao.pdf>.

14 BRASIL. *Constituição da República Federativa do Brasil de 1988*. Out. 1998. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>.

15 BRASIL. *Decreto nº 7.185*. Mai. 2010. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2010/Decreto/D7185.htm>.

16 BRASIL. *Lei nº 12.527*. Nov. 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>.

17 BRASIL. *Decreto nº 7.724*. Mai. 2012. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/Decreto/D7724.htm>.

18 HOUAISS, A. *Dicionário da língua portuguesa*. 2009.

19 OPEN GOVERNMENT WORKING GROUP. *8 Principles of Open Government*. Dez. 2007. Disponível em: <<http://www.opengovdata.org/home/8principles>>.

- 20 SUNLIGHT FOUNDATION. *10 principles of Open Government Data*. Ago. 2010. Disponível em: <<http://sunlightfoundation.com/policy/documents/ten-open-data-principles>>.
- 21 SHERIDAN, J.; TENNISON, J. Linking uk government data. In: *Linked Data on the Web Workshop at WWW2010*. [S.l.: s.n.], 2010.
- 22 MADNICK, S. E. et al. Overview and framework for data and information quality research. *Journal of Data and Information Quality (JDIQ)*, ACM, v. 1, n. 1, p. 2, 2009.
- 23 NATIONAL INFORMATION STANDARDS. *Understanding metadata*. [S.l.]: NISO Press, 2004.
- 24 ZUIDERWIJK, A.; JEFFERY, K.; JANSSEN, M. The necessity of metadata for open linked data and its contribution to policy analyses. In: *Conference on E-Democracy and Open Government (CeDEM 2012)*. [S.l.: s.n.], 2012. p. 281–294.
- 25 POWELL, L. Building a strong foundation with dublin core. In: *Proceedings of the 2009 International Conference on Dublin Core and Metadata Applications*. [S.l.]: Dublin Core Metadata Initiative, 2009. (DCMI '09), p. 87–95.
- 26 MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO. *Padrão de Metadados do Governo Eletrônico - e-PMG*. Jan. 2010. 4 p. Disponível em: <<http://www.governoeletronico.gov.br/biblioteca/arquivos/e-pmg>>.
- 27 CAMPOS, M. L. de A.; GOMES, H. E. *Taxonomia e Classificação: o princípio de categorização*. 2010. 9 p.
- 28 RICCIO, E. et al. *Introdução ao XBRL—nova linguagem para a divulgação de informações empresariais pela internet*. [S.l.]: SciELO Brasil, 2006. 166–182 p.
- 29 SILVA, P. C. da. *A Gestão Eletrônica da Informação Financeira no Banco Central do Brasil Apoiada por XBRL e Regras de Inferências*. [S.l.]: Banco Central do Brasil.
- 30 INTERNATIONAL, X. *Extensible Business Reporting Language (XBRL) 2.1*. Dez. 2012. Disponível em: <<http://www.xbrl.org/specification/xbrl-recommendation-2003-12-31+corrected-errata-2012-01-25.htm>>.
- 31 KOBASHI, N. Y. *Vocabulário controlado: estrutura e utilização*. [S.l.]: Escola Nacional de Administração Pública, Dez. 2008. 1–2 p.

- 32 MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃOS. *Vocabulário Controlado do Governo Eletrônico*. Mar. 2011. Disponível em: <<http://www.governoeletronico.gov.br/biblioteca/arquivos/vocabulario-controlado-do-governo-eletronico-vcge/download>>.
- 33 KÄFER, T. et al. *Towards a Dynamic Linked Data Observatory*. 2012. Disponível em: <<http://events.linkedata.org/ldow2012/papers/ldow2012-paper-14.pdf>>.
- 34 HARTIG, O. *Provenance Information in the Web of Data*. 2009. Disponível em: <http://events.linkedata.org/ldow2009/papers/ldow2009_paper18.pdf>.
- 35 OPEN KNOWLEDGE FOUNDATION. *Beyond Access: The Right to (Re)Use Public Information*. 2011. 8–9 p.
- 36 ABITEBOUL, S.; BUNEMAN, P.; SUCIU, D. *Data on the Web: from relations to semistructured data and XML*. [S.l.]: Morgan Kaufmann, 2000.
- 37 ANTONIOU, G.; HARMELEN, F. van. *A semantic web primer*. EUA: the MIT Press, 2008. 3 p.
- 38 BERNERS-LEE, T. *Linked Data*. Jun. 2009. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>.
- 39 O'RIAIN, S.; CURRY, E.; HARTH, A. Xbrl and open data for global financial ecosystems: A linked data approach. *International Journal of Accounting Information Systems*, Elsevier, v. 13, n. 2, p. 141–162, 2012.
- 40 GARCÍA, R.; GIL, R. Publishing xbrl as linked open data. In: CITESEER. *CEUR Workshop Proceedings*. [S.l.], 2009. v. 538.
- 41 FERNANDES, P. O.; CAMPOS de L. F. B. *Fundamentos para uma central de balanços brasileira*.
- 42 ALONSO, G. et al. *Web services*. [S.l.]: Springer, 2004.