



A qualidade na recuperação de dados governamentais: um estudo sobre dados de políticas públicas na internet

Fábio Mosso Moreira^I

<http://orcid.org/0000-0002-9582-4218>

Pedro Henrique Santos Bisi^{II}

<http://orcid.org/0000-0002-4513-9706>

Leonardo Castro Botega^{III}

<http://orcid.org/0000-0003-1495-5935>

José Eduardo Santarem Segundo^{IV}

<http://orcid.org/0000-0003-3360-7872>

Ricardo César Gonçalves Sant'Ana^V

<http://orcid.org/0000-0003-1387-4519>

^I Universidade Estadual Paulista, SP, Brasil.
Doutor em Ciência da Informação.

^{II} Universidade Estadual Paulista, SP, Brasil.
Mestre em Ciência da Informação.

^{III} Universidade Estadual Paulista, SP, Brasil.
Docente no Programa de Pós-graduação em Ciência da Informação.

^{IV} Universidade Estadual Paulista, SP, Brasil.
Docente no Programa de Pós-graduação em Ciência da Informação.

^V Universidade Estadual Paulista, SP, Brasil.
Docente no Programa de Pós-graduação em Ciência da Informação.

<http://dx.doi.org/10.1590/1981-5344/3994>

O uso da Internet para acessar dados governamentais é crescente e apresenta possibilidades para o cidadão

acompanhar a execução das políticas públicas. No contexto das políticas públicas para o setor da agricultura, destaca-se que estas constituem um fator preponderante no desenvolvimento social e econômico, principalmente dos pequenos produtores, que, por sua vez, necessitam desse conteúdo informacional para participarem e acompanharem as ações. O governo, mediante a Lei de Acesso à Informação, é responsável por garantir a ampla divulgação dos programas governamentais, entretanto, o que se observa muitas vezes é que os dados não são disponibilizados como deveriam, tendo principalmente problemas relacionados a qualidade das bases e dos serviços de recuperação. Assim, o objetivo da pesquisa consiste em identificar dimensões de qualidade envolvidas no acesso a dados governamentais na Internet e aplicá-las no estudo do processo de recuperação dos dados do Programa Nacional de Fortalecimento da Agricultura Familiar disponíveis no portal do Banco Central. Metodologicamente, realizou-se uma revisão teórica para identificar na literatura artigos que tratam sobre a questão da qualidade de dados no contexto governamental, e, posteriormente, uma Análise de Conteúdo para identificar categorias que pudessem representar tais dimensões de forma específica para este domínio. Estas dimensões foram aplicadas no estudo do processo de recuperação de dados de uma política pública agrícola direcionada ao desenvolvimento socioeconômico dos pequenos produtores, e os resultados obtidos estabelecem uma perspectiva sobre como a questão da qualidade de dados pode ser observada no cenário estudado, bem como suas implicações para o acesso e apropriação do conteúdo.

Palavras-chave: Recuperação de dados; qualidade; dados governamentais; política pública; PRONAF.

Quality in governmental data retrieval: a study of public policy data on the internet

The use of the Internet to access government data is growing and presents possibilities for citizens to monitor the implementation of public policies. In the context of public policies for the agriculture sector, it is highlighted that these constitute a major factor in social and economic development, especially for small producers, who, in turn, need this informational content to participate and monitor actions. The government, through the Access to Information Law, is responsible for ensuring the wide dissemination of government programs, however, what is often observed is that the data are not made available as they should, mainly having problems related to the quality of the databases and recovery services. Thus, the objective of the research is to identify dimensions of quality involved in accessing government data on the Internet and apply them to the study of the data recovery process of the Programa Nacional da Agricultura Familiar available on the Banco Central portal. Methodologically, a theoretical review was carried out to identify in the literature articles dealing with the issue of data quality in the government context, and, subsequently, a Content Analysis to identify categories that could represent such dimensions in a specific way for this domain. These dimensions were applied in the study of the data recovery process of a public agricultural policy aimed at the socio-economic development of small producers, and the results obtained to establish a perspective on how the issue of data quality can be observed in the studied scenario, as well as its implications. implications for access and ownership of content.

Keywords: *Data retrieval; data quality; governmental data; public policy; PRONAF.*

Recebido em 13.05.2019 Aceito em 13.05.2020

1 Introdução

O uso de tecnologias informacionais por parte dos pequenos produtores é influenciado pelas condições de conectividade no meio rural, pelo seu perfil cultural e socioeconômico, e pela presença de competências digitais desenvolvidas o suficiente para garantir o acesso a dados em

ambientes digitais. Como destaca Meitei e Devi (2009), a falta de conteúdo informacional compatível com a cultura destes produtores, a precariedade dos meios de comunicação, e o suporte financeiro inadequado para transferência de tecnologia nas pequenas comunidades rurais, são fatores que fazem com que muitas de suas necessidades informacionais não sejam atendidas.

Assim como a população urbana, e os produtores de médio e grande porte, os pequenos produtores também requerem diversos tipos de informação para a realização de suas atividades diárias que podem ser acessadas com o uso de tecnologias informacionais. Yusopet *al.* (2013) listam um conjunto de necessidades informacionais que se manifestam no cotidiano de pequenos produtores, categorizando-as nos seguintes grupos de informação: agrícola; médica e saúde; educação; religião; negócios; estilo de vida; entretenimento; modos familiares; esporte; carreiras e oportunidades; computadores e tecnologias; e governamental.

Com relação ao acesso a dados no domínio governamental, Sant'Ana (2009) ressalta o crescente interesse das administrações públicas em viabilizar a implantação de seus portais na Internet, seja por força da obrigatoriedade legal reforçada pela Lei de Acesso a Informação¹, seja pela necessidade de utilizar-se desta poderosa ferramenta de comunicação para fins políticos mais eficientes e transparentes. Para o autor, é importante que o governo disponibilize dados sobre suas ações, tais como àquelas relacionadas com a realização de políticas públicas, como os planos e programas governamentais.

As políticas públicas direcionadas ao desenvolvimento socioeconômico do meio rural são de grande importância para a evolução igualitária do setor agropecuário. Por exemplo, os programas governamentais de incentivos baseados no financiamento de crédito agrícola subsidiado garantem a viabilidade das produções principalmente para os pequenos produtores, que muitas vezes são descapitalizados, não dispõem de garantias adequadas, e não têm histórico de crédito (CONTI; ROITMAN, 2011).

Esta pesquisa parte do pressuposto de que, na dinâmica de execução das políticas públicas, o governo deve buscar tanto a eficiência na operacionalização dos programas governamentais quanto garantir que os dados resultantes de sua execução sejam disponibilizados de forma aderente ao perfil e necessidades de seus beneficiários.

¹ A Lei Federal nº 12.527/2011 – Lei de Acesso à Informação (LAI) – também é um instrumento importante, pois determinou novas obrigações às instituições públicas no que tange as condições de acesso à informação. Uma destas obrigações diz respeito à divulgação de informações em sites oficiais da rede mundial de computadores – Internet (BRASIL, 2011).

Para Farah (2008), a divulgação de programas governamentais envolve, muitas vezes, apenas a disseminação de conteúdos cujos quais as secretarias ou ministérios consideram como melhores práticas ou ações bem-sucedidas. Este fato reforça a relevância de se articular uma estratégia de disseminação ativa por parte das agências governamentais e a Internet figura como uma tecnologia que pode contribuir neste processo, uma vez que a mesma permite a ampliação da transparência pública e viabiliza a participação do cidadão de modo dialógico.

O governo é responsável por garantir a ampla divulgação das políticas públicas, contudo, observa-se que os dados não são disponibilizados como deveriam, tendo, principalmente, problemas relacionados à sua qualidade. Esta problemática motiva o objetivo da pesquisa, que consiste em identificar dimensões de qualidade envolvidas no acesso a dados governamentais na Internet e aplicá-las no estudo do processo de recuperação dos dados do Programa Nacional de Fortalecimento da Agricultura Familiar (PRONAF)² disponíveis no portal do Banco Central do Brasil (BC). Este estudo baseia-se no conceito de Ciclo de Vida dos Dados – CVD (SANT'ANA, 2016), uma vez que este permite estruturar uma perspectiva de análise que considera a qualidade como sendo um dos objetivos a serem atendidos nas fases de coleta, armazenamento, recuperação e descarte de dados (SANT'ANA, 2016).

Assim como é importante elaborar políticas públicas para o segmento dos pequenos produtores, também é fundamental sua ampla divulgação, que deve ser apoiada na disponibilização de dados qualificados. Desta forma, os resultados desta pesquisa podem estabelecer uma perspectiva sobre como a questão da qualidade de dados governamentais pode ser observada no cenário estudado, bem como suas implicações para o acesso e apropriação do conteúdo, uma vez que esta possui grande impacto na leitura, manipulação e esforço de desenvolvimento que o utilizador depende nos recursos informacionais disponibilizados (AGÊNCIA PARA MODERNIZAÇÃO ADMINISTRATIVA, 2016).

O artigo encontra-se organizado em sete seções, além desta introdução. Na segunda seção, apresenta-se uma revisão teórica acerca da temática que embasa o contexto desta pesquisa, que é o processo de recuperação de dados governamentais na Internet, dando ênfase para os aspectos da qualidade. A terceira seção do artigo descreve os

²O PRONAF é o principal programa governamental realizado para os pequenos produtores rurais e que possui o maior montante orçado de recursos pelo Governo Federal – 30 bilhões para o ano Safra da Agricultura Familiar 2017/2020 (SECRETARIA DE AGRICULTURA FAMILIAR E COOPERATIVISMO, 2017).

procedimentos metodológicos adotados na pesquisa. Na quarta seção, encontra-se a descrição dos recursos informacionais e das consultas contendo dados do PRONAF que estão disponíveis no portal do BC. A quinta seção traz uma sistematização das informações obtidas na literatura com o intuito de identificar e descrever categorias para representar dimensões das qualidades no domínio dos dados governamentais. Com a definição das dimensões de qualidade, a sexta seção discorre sobre as inferências obtidas a partir da aplicação destas nos recursos informacionais descritos na quarta seção. Por fim, a sétima seção destaca as considerações finais, contendo as conclusões e possíveis desdobramentos para a pesquisa.

2 Recuperação de dados governamentais na Internet

As possibilidades de uso da Internet para o acesso de dados governamentais pode motivar o desenvolvimento de mecanismos de serviços públicos mais transparentes para interação com a sociedade (MOREIRA; SANT'ANA, 2016). Para Sant'Ana (2009), a utilização desta tecnologia pelo governo pode ser dividida em: uso interno (auxiliando nos processos e fluxos informacionais de gestão da coisa pública) e uso externo (facilitando o compartilhamento de dados entre a organização pública e entidades externas).

Com relação ao uso externo da Internet, os órgãos governamentais podem adotar dois modelos para o desenvolvimento de mecanismos que viabilizem a transparência na disponibilização dos dados em seus portais: a forma ativa (tornando disponíveis conjuntos de dados para livre acesso à população); e a forma passiva (oferecendo mecanismos para receber solicitações dos cidadãos que requerem acesso aos conjuntos de dados ainda não disponíveis) (SANT'ANA, 2009).

No Brasil, uma das primeiras iniciativas que propuseram o uso da Internet para disponibilização de dados governamentais ocorreu em 1994, com o Portal da Transparência da Controladoria-Geral da União, "(...) trazendo informações sobre a aplicação de recursos do governo federal, abrangendo transferências e aplicações diretas" (PLATT NETO; CRUZ; VIEIRA, 2006, p.137).

Segundo Andersen e Henrikisen (2006), o Governo Eletrônico envolve questões tecnológicas (como a utilização dos recursos de informática) e organizacionais (como no tratamento e efetividade da administração pública). Para os autores, os objetivos perseguidos são: transparência, acessibilidade, prestação de contas, serviços públicos eficientes, políticas públicas e envolvimento dos usuários.

No ano de 2011, o governo brasileiro firmou um compromisso, em parceria internacional, de Governo Aberto (*Open Government Partnership*),

na qual o país é um dos estados fundadores. Para cumprimento desta parceria, os sítios oficiais do governo, que tomam acessíveis seus conjuntos de dados, devem atender aos oito princípios³ de dados abertos, utilizando recursos tecnológicos para ampliar a oferta, de modo que cidadãos tenham capacidade de recuperar dados que lhes sejam de interesse e atendam suas necessidades (RODRIGUES; SANT'ANA, 2012).

Apesar das possibilidades proporcionadas pela Internet para os serviços de disponibilização de dados governamentais, as características relacionadas à baixa carga semântica e à alta estruturação inerentes aos dados tornam seu processo de recuperação diferente dos processos de recuperação da informação via mecanismos de buscas (JANOWICZ *et. al.* 2012).

Segundo Van Rijsbergen (1979), na recuperação de dados busca-se por uma correspondência exata à necessidade, enquanto na recuperação da informação espera-se uma correspondência aproximada (a melhor correspondência possível); na recuperação de dados a inferência utilizada é a dedução lógica, enquanto na recuperação da informação é possível realizar uma inferência indutiva (possui graus de certeza ou incerteza); na recuperação de dados a linguagem utilizada é uma linguagem artificial restrita por sintaxes específicas, enquanto na recuperação da informação utiliza-se de uma linguagem natural que pode ser expressa apenas com as especificações necessárias; ainda sobre a linguagem de consulta utilizada, na recuperação de dados a linguagem possui maior sensibilidade ao erro (por ser amarrada a uma sintaxe padrão), enquanto na recuperação da informação pequenos erros podem ser contornados.

As características envolvidas no processo de recuperação de dados devem ser consideradas pelas agências governamentais ao optarem por incorporar os preceitos de Governo Eletrônico e os princípios do Governo Aberto, sendo que o aspecto da qualidade é um dos objetivos a serem perseguidos neste fluxo informacional.

O conceito de qualidade é variável e subjetivo, e considera as necessidades, ações e objetivos de cada domínio, caracterizando-se como parâmetros para que o conteúdo acessado esteja de acordo com os requisitos. Como a qualidade é um aspecto inerente a vários domínios de processo de recuperação de dados, seus problemas e suas dimensões podem ser mutáveis de acordo com o contexto (OLETO, 2006).

Estas dimensões são interpretações do problema de qualidade, são um ponto de vista, um parâmetro a ser considerado ao avaliar esta questão nos fluxos de dados, metadados ou informações, e devem ser

³Open Government Partnership. Disponível em: <https://www.opengovpartnership.org/>. Acesso em: 30 abr. 2020.

consideradas em relação a objetivos específicos do usuário e do domínio do conteúdo compartilhado. Existem múltiplas visões quanto à classificação das dimensões de qualidade da informação, gerando diferentes interpretações do conceito de qualidade, seus atributos e relações (KHALEGHI *et al.*, 2013).

Abordar as dimensões de qualidade no processo de recuperação de dados visa torná-los aderentes às necessidades do domínio cujos quais eles são estudados. Para Todoran *et al.* (2015), novas questões e problemas relacionados à complexidade, heterogeneidade e qualidade dos dados estão emergindo, tornando-se uma perspectiva cada vez mais relevantes em ambientes informacionais.

Gil-García e Pardo (2005) apresentam uma análise dos fatores de sucesso específicos da qualidade no domínio do Governo Eletrônico, e destacam que a garantia da conformidade dos dados é uma estratégia efetiva para lidar com os desafios relacionados à informação. Os autores abordam problemas de qualidade na disponibilização dos dados no contexto governamental, e citam trabalhos como Redman (1998), evidenciando dimensões como a acurácia, a consistência e a completude dos dados.

Barnes e Vidgen (2006) aplicaram a triangulação de métodos para avaliação da percepção da qualidade por parte dos usuários em um portal do Governo Eletrônico do Reino Unido, utilizando a técnica *eQual* 4.0, que possui três áreas principais – usabilidade, qualidade da informação e serviço de interação. Com relação a área da qualidade da informação, os autores abordam as seguintes dimensões neste estudo de caso: a acurácia, a confiabilidade, o tempo oportuno, a relevância, a clareza, o nível certo de detalhamento dos dados e o formato apropriado do arquivo disponibilizado.

Sadiq e Indulska (2017) abordam o crescimento do movimento de dados abertos, ressaltando que este tipo de informação tem o potencial de promover a inovação de negócios, empreendedores, organizações de pesquisa e agências do governo. Com ênfase dada para os dados abertos no contexto governamental, apresentam alguns exemplos como o caso do governo australiano, que aumentou em 900% o número de bases de dados abertas disponíveis entre 2013 e 2015, além de outros governos, como os da Nova Zelândia (mais de 3800 *datasets*), Reino Unido (mais de 23000 *datasets*), Estados Unidos (mais de 194000 *datasets*), e Canadá (mais de 240000 *datasets*), que também disponibilizam bases em seus portais oficiais. Os autores concluem que, no estudo de portais governamentais que disponibilizam dados abertos, existem evidências da ocorrência de alguns problemas relacionados aos aspectos da qualidade, como um “*time-to-value*” longo até que as bases sejam postas para uso, a

disponibilização de metadados deficientes, além do fato de que muitas bases de dados estão duplicadas, inconsistentes e faltando dados.

Rasyid e Alfina (2016) evidenciam a importância do Governo Eletrônico e, por meio de um estudo de caso aplicado em um portal oficial de uma agência do sistema de saúde, avaliam os aspectos da qualidade nos serviços governamentais de informação da Indonésia. Os autores desenvolvem uma análise baseada em um modelo contendo sete dimensões da qualidade, envolvidos neste processo, como o *design* do website; a confiabilidade; a responsividade; a segurança e privacidade; a personalização dos mecanismos de consulta, a informação (acurácia e atualização do conteúdo), e o fácil uso (forma do conteúdo e estrutura do website).

Lourenço (2015) avalia se a estrutura atual de sete portais do Governo Aberto está adequada para apoiar a transparência e a prestação de contas com a sociedade, definindo um conjunto de oito requisitos-chaves presentes na divulgação de dados. Um destes requisitos é a qualidade, que segundo o autor envolve aspectos como a acurácia, a consistência, e a confiabilidade. Além dos aspectos intrínsecos à qualidade, o autor também se refere a outras características-chave presentes neste processo que também se relacionam com a mesma, como a completude, a visibilidade, a compreensibilidade, a pontualidade, o valor e utilidade, a granularidade, e a comparabilidade.

A próxima seção deste artigo descreve os procedimentos metodológicos adotados para analisar os aspectos da qualidade no processo de recuperação de dados de políticas públicas na Internet, como parte do domínio dos dados governamentais.

3 Procedimento metodológico

O procedimento metodológico utilizado na pesquisa partiu de levantamento teórico a fim de obter informações para descrever aspectos envolvidos no acesso a dados governamentais na Internet, dando enfoque para resultados de estudos que tratam sobre a questão da qualidade do processo de recuperação.

A partir do conteúdo obtido com o levantamento teórico, definiram-se dimensões da qualidade específicas para o domínio dos dados governamentais. Para tanto, recorreu-se ao método de Análise de Conteúdo (BARDIN, 2009), e com o uso da técnica de análise categorial - *a posteriori* - foram definidas categorias para representar tais dimensões. O critério para definição das categorias baseou-se nas características apontadas pelos autores consultados no levantamento teórico como sendo requisitos para garantir a conformidade dos dados no domínio analisado. Estas características encontram-se sistematizadas no Quadro 1.

Após a definição das categorias, selecionou-se o portal do BC do Brasil <http://bcb.gov.br/> como objeto de estudo para aplicar as dimensões da qualidade no processo de recuperação de dados do PRONAF. A escolha por este portal teve como critério o fato de que esta instituição corresponde a principal agência regulamentadora das operações financeiras no país, segundo as orientações do Conselho Monetário Nacional, tendo como uma de suas funções exercer o controle de crédito e fiscalização das instituições financeiras (BANCO CENTRAL, 2017). Assim, compete ao BC centralizar todos os dados referentes aos financiamentos de crédito realizados pelas instituições financeiras no Brasil, incluindo as operações do PRONAF junto aos pequenos produtores.

Realizou-se, então, uma busca online no portal do BC aplicando o método de observação direta, realizada por meio da navegação e exploração deste ambiente informacional através dos menus e mecanismo de busca, onde foram identificados dois recursos informacionais contendo dados do PRONAF. As dimensões de qualidade definidas na pesquisa foram aplicadas nestes *data sets* de forma interpretativa, tomando como base a observação das características referentes às funcionalidades da interface de consultas, dos atributos de dados encontrados, da estrutura do conteúdo apresentado, e dos formatos de arquivos disponibilizados.

4 Descrição dos recursos informacionais e das bases de dados do PRONAF disponíveis no portal do Banco Central do Brasil (BC)

Os dois recursos informacionais contendo dados do PRONAF que estão disponíveis no portal do Banco Central foram: o "Anuário Estatístico do Crédito Rural"⁴ e a "Matriz de Dados do Crédito Rural - Contratações"⁵. Ambos os recursos informacionais continham *data sets* que puderam ser acessados por meio do item de menu na página inicial rotulado "Estabilidade Financeira", e seus submenus "Supervisão" e "Crédito Rural".

O conteúdo disponibilizado em ambos os recursos informacionais continha dados sobre a quantidade de contratos e valores contratados nos financiamentos do PRONAF, organizados em diversas categorias, como, por exemplo: período (ex: mensal e anual); atividade (ex: agricultura ou pecuária); finalidade (ex: custeio, investimento ou comercialização); linha

⁴Disponível em: https://www.bcb.gov.br/estabilidadefinanceira/anuario_estat_credrural. Acesso em 08/05/2019.

⁵ Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/micrrural>. Acesso em: 08/05/2019.

do programa (ex: PRONAF Jovem, PRONAF Floresta, etc.); fonte do recurso (ex: Tesouro, FAT, Fundos Constitucionais); região geográfica (ex: Norte, Sul, Sudeste,); unidade da federação (ex: São Paulo, Minas Gerais, Paraná, etc.); tipo de instituição operadora do crédito (ex: Bancos Públicos, Bancos Privados); tipo de beneficiário (ex: cooperativas ou produtores); faixas de financiamento (ex: em valores); área financiada (em hectares); e empreendimento (ex: frutas, hortaliças).

Anuário Estatístico do Crédito Rural

No Anuário Estatístico do Crédito Rural, encontram-se conjuntos de dados referentes à operacionalização do PRONAF entre os anos de 1999 até 2012 (**Figura 1**).

Figura 1 - Anuário Estatístico do Crédito Rural



Fonte: Autor adaptado de BC. Disponível em:

https://www.bcb.gov.br/estabilidadefinanceira/anuario_estat_credrural. Acesso em: 31 maio 2019.

Todas as consultas presentes neste recurso informacional possuem os mesmos serviços de visualização:

1.Apresentação:contendo um arquivo no formato *Portable Document Format*(.PDF) com dados não estruturados referentes à descrição do programa e apresentação dos dados.

2.Legislação: contendo um arquivo no formato PDF com dados estruturados no formato de tabela, contendo informações sobre o histórico da ocorrência de normativas relacionadas ao programa (ex: resoluções, circulares, cartas-circulares, comunicados), possuindo os seguintes atributos: número do documento e data.

3. Evolução dos Recursos: contendo dois arquivos, ambos no formato PDF, com informações sobre o histórico de evolução de recursos destinados ao programa. Um arquivo contendo dados estruturados no formato de tabela e outro no formato de gráfico.

4. Encargos Financeiros Praticados: contendo um arquivo no formato PDF, com dados estruturados no formato de tabelas, informando sobre a taxa dos encargos financeiros praticados. Possui os seguintes atributos: descrição da linha de crédito, indexador e taxa efetiva de juros.

5. Financiamentos Concedidos: contendo três subconsultas disponibilizando arquivos no formato PDF, que possuem dados estruturados tanto na forma de tabelas quanto de gráficos, e encontram-se organizados em três tipos de agrupamento: "contratos para produtos", "contratos para cooperativas" e "contratos totais". Os conjuntos de dados possuem os atributos "quantidade" e "valor", e encontram-se organizados segundo os parâmetros "atividade e finalidade", "fonte de recurso", "região geográfica e UF", "tipo de instituição", "faixa de financiamento, fonte de recurso e finalidade", "área financiada e empreendimento", e "unidade da federação, município, atividade e finalidade".

6. Instituições Participantes do Crédito Rural: contendo um arquivo no formato PDF com dados estruturados no forma de tabelas, listando as instituições que operacionalizam contratos do PRONAF. O conjunto de dados possui os atributos "ordem", "instituição" e "quantidade de municípios" atendidos por cada instituição.

7. Anuário Completo em Excel: contendo um arquivo compactado que agrupa todos os demais arquivos contendo dados que encontram-se disponíveis neste recurso informacional no formato Microsoft Excel (.XLS).

Matriz de Dados do Crédito Rural

O segundo recurso informacional identificado no portal do Banco Central contendo dados do PRONAF é a "Matriz de Dados do Crédito Rural" (Figura 2).

Figura 2 - Interface de uma das consultas presentes na Matriz de Dados do Crédito Rural



Fonte: Autor adaptado de BC. Disponível

em: <https://www.bcb.gov.br/estabilidadefinanceira/reportmicrrural/>. Acesso em: 31 maio 2019.

Neste segundo recurso informacional, encontram-se dez consultas para os dados do PRONAF referentes aos financiamentos executados a partir de 2013:

1. Definições, conceitos e esclarecimentos: arquivo no formato PDF com dados estruturados referentes aos aspectos gerais do programa, como, por exemplo, as definições, possíveis beneficiários e objetivos do crédito rural;

2. Evolução dos Recursos Financeiros: arquivo no formato XLSX contendo dados estruturados em tabelas, representando o histórico de evolução dos recursos aplicados no programa. O conjunto de dados possui atributos como: "ano", "valor nominal", "valor constante" e "variação anual";

3. Evolução dos Encargos Financeiros: arquivo no formato XLSX, contendo dados estruturados em tabela, que representam as taxas de financiamento das linhas do programa. O conjunto de dados possui atributos como: "período de vigência", "beneficiário/programa", "encargos financeiros" e "base normativa";

4. Visão por Região, UF e Município: serviços de visualização de dados organizados em dezessete tipos de agrupamento (ex: região e Brasil; Região e UF; Região, UF e Gênero; Região, UF e Tipo de Pessoa; Município; Custeio por Região, UF e Produto; Investimento por Região, UF e Produto; Comercialização por Região, UF e Produto; Custeio por

Município e Produto; Investimento por Município e Produto; Comercialização por Município e Produto; Região, UF, Segmento e IF; Ano Agrícola 2013/2014; Ano Agrícola 2014/2015; Ano Agrícola 2015/2016; Ano Agrícola 2016/2017, e; Ano Agrícola 2018/2018). A interface de recuperação permite a inserção de parâmetros como "intervalo mensal", "intervalo anual", "fonte de recursos", "programa" e "subprograma", e exibe as tabelas de dados em uma página no formato Active Server Page Extended(.ASPX), com hiperlinks para realizar o download dos dados em arquivos nos formatos Comma-SeparatedValues (.CSV), PDF e XLSX;

5.Visão por Segmento e IF: serviço de visualização de dados organizados em nove tipos de agrupamento (ex: Custeio Agrícola por Segmento, IF e Produto; Custeio Pecuário por Segmento, IF e Produto; Investimento Agrícola por Segmento, IF e Produto; Investimento Pecuário por Segmento, IF e Produto; Comercialização Agrícola por Segmento, IF e Produto; Comercialização Pecuária por Segmento, IF e Produto; Segmento, IF, Região e UF; e Segmento, IF e Tipo de Pessoa). A interface de recuperação permite a inserção de parâmetros como "intervalo mensal", "intervalo anual", "fonte de recursos", "programa" e "subprograma", e exibe as tabelas de dados em uma página ASPX, com hiperlinks para realizar o download dos dados em arquivos nos formatos CSV, PDF e XLSX;

6.Visão por Tipo de Pessoa e Gênero: serviço de visualização de dados organizados em quatro tipos de agrupamento (ex: Tipo de Pessoa, Região e UF; Gênero, Região e UF; Tipo de Pessoa, Segmento e IF, e; Faixa Etária e UF). A interface de recuperação permite a inserção de parâmetros como "intervalo mensal", "intervalo anual", "fonte de recursos", "programa" e "subprograma", e exibe as tabelas de dados em uma página ASPX, com hiperlinks para realizar o download dos dados em arquivos nos formatos CSV, PDF e XLSX;

7.Visão por Faixa de Valores: serviço de visualização de dados organizados em catorze tipos de agrupamentos (ex: Faixa de Valores; Faixa de Valores e UF; Custeio por Faixa de Valores, Região e UF; Investimento por Faixa de Valores, Região e UF; Comercialização por Faixa de Valores, Região e UF; Custeio por Faixa de Valores e Município; Investimento por Faixa de Valores e Município; Comercialização por Faixa de Valores e Município; Custeio Agrícola por Faixa de Valores, Segmento e IF; Custeio Pecuário por Faixa de Valores, Segmento e IF; Investimento Agrícola por Faixa de Valores, Segmento e IF; Investimento Pecuário por Faixa de Valores, Segmento e IF; Comercialização Agrícola por Faixa de Valores, Segmento e IF; e Comercialização Pecuária por Faixa de Valores, Segmento e IF). A interface de recuperação permite a inserção de parâmetros como "intervalo mensal", "intervalo anual", "fonte de

recursos”, “programa” e “subprograma”, e exibe as tabelas de dados em uma página ASPX, com hiperlinks para realizar o download dos dados em arquivos nos formatos CSV, PDF e XLSX;

8.Visão por Programa e Subprograma: serviço de visualização de dados organizados em quatro tipos de agrupamento (ex: Programa e Subprograma; Programa, Subprograma e UF; Programa, Subprograma e IF; Programa, e Subprograma e Produto). A interface de recuperação permite a inserção de parâmetros como “intervalo mensal”, “intervalo anual”, “fonte de recursos”, “programa” e “subprograma”, e exibe as tabelas de dados em uma página ASPX, com hiperlinks para realizar o download dos dados em arquivos nos formatos CSV, PDF e XLSX;

9.Visão por Fonte de Recursos: serviço de visualização de dados organizados em quatro tipos de agrupamento (ex: Fonte de Recursos; Fonte de Recursos e UF; Fonte de Recursos e IF; e Fonte de Recursos e Produto). A interface de recuperação permite a inserção de parâmetros como “intervalo mensal”, “intervalo anual”, “fonte de recursos”, “programa” e “subprograma”, e exibe as tabelas de dados em uma página ASPX, com hiperlinks para realizar o download dos dados em arquivos nos formatos CSV, PDF e XLSX, e;

10.Visão por Tipo de Beneficiário: serviço de visualização de dados organizados em quatro tipo de agrupamento (ex: Tipo de Beneficiário; Tipo de Beneficiário e UF; Tipo de Beneficiário e IF, e; Tipo de Beneficiário e Produto). A interface de recuperação permite a inserção de parâmetros como “intervalo mensal”, “intervalo anual”, “programa”, “subprograma” e “fonte de recursos”, e exibe as tabelas de dados em uma página ASPX, com hiperlinks para realizar o download dos dados em arquivos nos formatos CSV, PDF e XLSX.

A recuperação de dados do PRONAF no portal do BC permitiu observar características distintas na forma de disponibilização do conteúdo ao comparar os dois recursos informacionais analisados. Ambos continham o mesmo tipo de conteúdo (atributos), mas o recurso informacional “Matriz de dados do crédito rural” possui uma interface com maiores funcionalidades, que permite a parametrização de filtros específicos, manipulação dos dados na forma de tabela dinâmica, e download dos dados em arquivos com formatos distintos (.PDF, .XLSX, .CSV). Esse fato indica uma evolução na forma de disponibilizar os dados ao longo do tempo por parte da instituição.

A próxima seção do artigo apresenta uma sistematização das dimensões de qualidade na recuperação de dados governamentais definidas com a Análise de Conteúdo para, posteriormente, discutir sobre como as mesmas podem ser verificadas no cenário dos recursos informacionais e bases de dados descritas nesta seção.

5. Dimensões de qualidade na recuperação de dados governamentais na Internet

Para representar as dimensões da qualidade a serem aplicadas na análise do processo de recuperação de dados do PRONAF encontrado nos recursos informacionais disponibilizados no portal do BC, elaborou-se o Quadro 1 para sistematizar as categorias definidas e os respectivos autores que trataram sobre este aspecto no domínio dos dados governamentais. Foram identificadas treze categorias, sendo estas: Acurácia, Consistência, Completude, Confiabilidade, Pontualidade, Compreensibilidade, Granularidade, Formato de arquivo, Presença de Metadados, Design, Responsividade, Segurança e Personalização.

Quadro 1 – Dimensões de qualidade nos processos de recuperação de dados governamentais

Dimensões de Qualidade	GIL-GARCIA; PARDO (2005)	BARNES; VIDGEN (2006)	SADIQ; INDULSKA, (2017)	RASYID; ALFINA (2016)	LOURENÇO (2015)
Acurácia	"data quality problems include inaccuracies , inconsistencies, and incompleteness of data."	"Provides accurate information"			"quality is sometimes simply associated with accuracy "
Consistência	"data quality problems include inaccuracies, inconsistencies , and incompleteness of data."		"many open datasets have duplicate, inconsistent , and missing data"		" Consistency is also associated"
Completude	"data quality problems include inaccuracies, inconsistencies, and incompleteness of data."		"many open datasets have duplicate, inconsistent , and missing data"		" completeness means that citizens should have access to the legally required and permissible (...) information"
Confiabilidade		"Provides believable information"		"(...) model consists of seven dimensions: web design, reliability , responsiveness, privacy and security, personalization, information, and ease of use."	" Reliability , authenticity or validity is another important aspect of data quality"

Pontualidade		<i>"Provide timely information"</i>	<i>"there is also some evidence that the time-to-value from these datasets remains prohibitively long"</i>		<i>"timeliness might be considered as a key characteristic to allow for disclosed data"</i>
Compreensibilidade		<i>"Provides easy to understand information"</i>			<i>"Comprehensibility or understandability aims to ensure that users can understand and interpret appropriately the provided financial information"</i>
Granularidade		<i>"Provides information at the right level of detail"</i>			<i>"Granularity is an important characteristic of information disclosure as it is related to the level of detail with which data is made available"</i>
Formato de arquivo		<i>"Presents the information in an appropriate format"</i>		<i>"(...) model consists of seven dimensions: web design, reliability, responsiveness, privacy and security, personalization, information, and ease of use."</i>	
Presença de Metadados			<i>"the metadata, as well as the underlying data quality for these datasets, is known to be deficient."</i>		
Design da interface				<i>"(...) model consists of seven dimensions: web design, reliability, responsiveness, privacy and security, personalization, information, and ease of use."</i>	
Responsividade				<i>"(...) model consists of seven dimensions: web design, reliability, responsiveness, privacy and security, personalization, information, and ease of use."</i>	
Segurança				<i>"(...) model consists of seven dimensions: web design, reliability, responsiveness, privacy and security, personalization, information, and</i>	

				<i>ease of use."</i>	
Personalização				<i>"(...)model consists of seven dimensions: web design, reliability, responsiveness, privacy and security, personalization, information, and ease of use."</i>	

Fonte: Dados da Pesquisa – 2018.

Para a aplicação destas dimensões no objeto de estudo da pesquisa, considerou-se que o processo de recuperação de dados analisado pode ser organizado em uma perspectiva multinível de acesso (MOREIRA; SANT'ANA; JORENTE, 2016): o **primeiro nível** é composto pelas atividades de localização e exploração dos portais pelos usuários, utilizando-se dos recursos de navegação e sistemas de busca; o **segundo nível** compreende o arranjo dos elementos informacionais que constituem os objetos do ambiente informacional, suas relações e representações, como os *data sets*, conteúdos multimídias, documentos públicos, entre outros; e o **terceiro nível** envolve os elementos estruturais dos conjuntos de dados, seus registros e recursos de visualização, como formato de arquivo, atributos dos registros, elementos de visualização, entre outros.

A **acurácia** é uma dimensão de qualidade que impacta principalmente no terceiro nível do processo de recuperação de dados, e pode ser definida como sendo o cuidado, exatidão, grau de conformidade do dado com o padrão estabelecido como exatidão (SILVA *et al.*, 2006; WEBER *et al.*, 1999). Um problema prático de acurácia poderia ser observada na divulgação de dados que tiveram algum viés de erro no momento de sua medição devido a calibragem do instrumento de registro, ou, por exemplo, em um relatório de estoque gerado por um sistema de informação cujas movimentação de produtos sofreram alguma inconsistência de lançamento e os saldos não refletem a quantidade física real (DROHOMERETSKI; FAVARETTO, 2010).

A **consistência** é uma dimensão de qualidade que manifesta-se no terceiro nível do processo de recuperação de dados, e pode ser definida como o grau de coerência lógica de determinada variável em função do valor de outra variável complementar (ABATH *et al.*; 2014). Segundo Abath *et al.* (2014), um exemplo de problema de consistência poderia ser observado no registro de uma ocorrência em que uma pessoa encontra-se indicada como sendo do sexo masculino mas é caracterizada como vítima

de violência sexual com penetração vaginal, o que seria considerado um dado conflitante em questão de coerência.

A **completude** é uma dimensão de qualidade que também impacta no terceiro nível do processo de recuperação de dados, e refere-se ao grau de preenchimento dos campos analisados, mensurado pela proporção entre a quantidade de campos preenchidos e os não preenchidos. Na prática, a não completude do conjunto de dados pode ser observada quando o preenchimento da tabela no banco de dados tem algum campo ignorado, recebe numeral zero, ausência de preenchimento ou possui valor que indica que o dado está incompleto (CEPAL, 2003).

A **confiabilidade** é uma dimensão de qualidade que implica no primeiro e no terceiro nível do processo de recuperação de dados, podendo ser definida como sendo o estado de confiança que existe em relação aos dados, considerando que os valores do conjunto são completos o suficiente e sem erros de convicção para seu objetivo e contexto, bem como foram gerados e disponibilizados por ator autêntico e competente (MORGAN; WARING, 2004). Na prática, problemas de confiabilidade podem surgir quando a fonte que disponibiliza o conjunto é duvidosa, e com relação ao conteúdo, pode derivar de erros e sub notificações de dados no momento da coleta.

A **pontualidade** é uma dimensão de qualidade que impacta no terceiro nível de recuperação de dados, e pode ser definida como o lapso de tempo que ocorre entre uma mudança de um estado real e seu registro no sistema de informação ou base de dados (WANG; STRONG, 1996). A pontualidade pode influenciar principalmente na disponibilização de dados mercadológicos, como exemplificado por Barcellos (2013), que apresenta uma análise das implicações desta dimensão da qualidade na divulgação dos dados de demonstrações contábeis referentes às companhias financeiras listadas na BM&FBOVESPA para a tomada de decisão e estabelecimento de estratégia por parte dos investidores.

A **compreensibilidade** é uma dimensão de qualidade verificada no terceiro nível do processo de recuperação de dados, e pode ser definida como o nível de compreensão e de clareza que o dado se apresenta para os usuários, agregando características que permitam entender e assimilar sua utilização com o atendimento de uma necessidade informacional (POLZER; SCHNEIDER; WILSON, 2011). No contexto prático, a compreensibilidade poderia ser observada em um conjunto de dados quando observamos os rótulos adotados para representar as colunas de uma tabela, onde é comum encontrar ambiguidade de termos ou o uso de abreviações/siglas que possuem significado para quem disponibiliza, mas não transmite semântica para um usuário que não está habituado com expressões técnicas daquele domínio.

A **granularidade** é uma dimensão de qualidade que impacta principalmente no terceiro nível de recuperação de dados, e baseia-se nos possíveis subconjuntos que possam ser identificados e abstraídos a partir de determinado critério de dissimilaridade com uma ou mais dimensões de análise. Assim, a granularidade de um conjunto de dados está vinculada ao número de atributos que compõem o conjunto e a diversidade de seus conteúdos (SANTOS; SANT'ANA, 2013).

O **formato de arquivo** é uma dimensão de qualidade que impacta no segundo nível de recuperação de dados, e está relacionado com a tecnologia que abarca os conjuntos dados disponibilizados e sua condição de ser manipulado por ferramentas de armazenamento e processamento (WANG; STRONG, 1996). Quando trata-se de disponibilização de dados, é comum encontrar alguns formatos que são utilizados com mais frequência, como os formatos de planilha eletrônica (ex: XLS, XLSX), arquivos separados por vírgula (ex: CSV), documentos de texto (ex: DOC, DOCX, ODF, TXT), imagens (ex: PNG, JPEG), formatos de intercâmbio (ex: XML, JSON), entre outros.

A **presença de metadados** é uma dimensão de qualidade que também pode ser observada no segundo nível do processo de recuperação de dados, uma vez que os elementos de metadados podem auxiliar na representação de recursos informacionais por meio da convergência de formatos de codificação e organização de conhecimento, seguindo estruturas padrões para representara informação rotulando conteúdos a fim de facilitar sua localização por meio de sistemas de busca (MARTINES USERO, 2007). No contexto dos dados governamentais, destaca-se o uso do Padrão Metadados do Governo Eletrônico (e-PMG), que, segundo Comitê Executivo do Governo Eletrônico (CEGE, 2014), é um padrão composto por vinte elementos considerados necessários para recuperação e gerenciamento de informações e conteúdos no domínio público, sendo quinze deles baseados no Padrão Dublin Core (DC), e outros cinco elementos adicionais identificados como necessários para o contexto do Governo Eletrônico brasileiro.

O **design** também pode ser considerado uma dimensão de qualidade que impacta no primeiro e no segundo nível do processo de recuperação de dados, uma vez que esta dimensão consiste do modo como é feito a preparação dos elementos e objetos informacionais para que o ambiente informacional possa ser utilizado por seres humanos de forma eficiente e eficaz, utilizando-se de recursos distintos como palavras, imagens, tabelas, gráficos, mapas e desenhos (JACOBSON, 1999). Na contexto desta pesquisa, o *design* na qualidade de dados poderia ser analisado ao se observar as funcionalidades presentes na interface do

portal analisado, bem como na estrutura do serviço de visualização para os conjuntos de dados disponíveis.

A **responsividade** é uma dimensão de qualidade que impacta no primeiro nível do processo de recuperação de dados, e pode ser definida como a capacidade de adaptação da interface do ambiente informacional ao contexto dos dispositivos móveis que realizam o acesso (OLIVEIRA, 2014). De forma prática, a responsividade poderia ser observada ao se utilizar o mesmo serviço de recuperação de dados por meio de um smartphone ou de um notebook sem que haja prejuízos significativos quanto as funcionalidades da interface ou do conteúdo apresentado.

A **segurança** é uma dimensão de qualidade que impacta no segundo nível do processo de recuperação de dados, podendo ser definida como a extensão na qual o acesso aos conjuntos de dados pode ser restrito, privativo e/ou assegurados (WANG; STRONG, 1996). Os conjuntos de dados podem estar disponíveis de maneira aberta ou pode ser requerido uma credencial (ou cadastro) para efetivar o acesso aos arquivos e/ou serviços de consulta. Esta camada de controle pode ser considerado um aspecto da segurança aplicada na recuperação de dados.

Por fim, a **personalização** é uma dimensão de qualidade que impacta no segundo nível de recuperação de dados, podendo ser observada a partir das possibilidades e flexibilidades que o serviço de recuperação proporciona em uma consulta de acordo com preferências relatadas pelos usuários (KOUTRIKA; IOANNIDIS, 2005). De forma prática, esta dimensão de qualidade poderia se manifestar nas funcionalidades do serviço de recuperação, como a possibilidade da configuração de parâmetros e filtros específicos.

6 Análises e discussões

As análises e discussões apresentam resultados obtidos a partir da aplicação das dimensões da qualidade definidas na seção anterior deste artigo, e discorre sobre como estas podem se manifestar a partir das características observadas no processo de recuperação de dados do PRONAF disponíveis no portal do BC.

No processo de recuperação de dados do PRONAF no portal do BC, problemas de **acurácia** poderiam se manifestar caso houvesse um atraso no cadastramento de contratos executados nos últimos dias do mês (ex: dia 31), fazendo com que o mesmo seja contabilizado no mês subsequente. As consultas disponíveis no portal do BC permitem a configuração de filtros temporais (mês, ano), e caso atrasos como estes ocorram, os dados apresentados não refletiriam com precisão o montante financiado no período de tempo escolhido para a visão. Para mitigar este problema, as instituições financeiras operadoras do programa devem

adotar procedimentos e rotinas bem definidas para que haja o registro imediato dos contratos no sistema.

Por sua vez, a **consistência** poderia ser observada caso houvesse problemas de coerência entre a linha de crédito informada no contrato e o perfil do beneficiário, uma vez que existem requisitos específicos para enquadrar-se nos grupos de financiamento; ou, também, em casos onde houvesse a associação de um contrato classificado como sendo de custeio da atividade, mas seu objeto refere-se a um passivo que deve ser financiado por contratos classificados como de investimento (ex: compra de máquinas agrícolas). Em ambos os casos, a inconsistência ocasionaria problemas para o uso dos dados, podendo, inclusive, gerar inferências precipitadas na avaliação da política pública. Para tanto, destaca-se a necessidade de auditoria constante dos dados fornecidos pelas instituições financeiras a fim de evitar que hajam inconsistências no conteúdo disponibilizado.

A **completude** é uma dimensão da qualidade que poderia ser observada com a verificação de campos em brancos nas consultas disponíveis. Se na visão referente aos contratos executados por município, em determinado registro os campos 'quantidade de contratos' e 'valor de contratos' fossem exibidos em branco ao invés de um valor zerado (no caso deste município não ter tido nenhum beneficiário do programa), este indicador poderia gerar dúvidas nas análises. Problemas de completude podem ser tratados no momento da coleta dos dados, onde considera-se fundamental a eleição consciente dos campos definidos como obrigatórios e opcionais para o cadastramento dos contratos.

A **confiabilidade** na recuperação dos dados do PRONAF disponíveis no portal do BC pode ser condicionada a dois aspectos principais: o primeiro refere-se a autenticidade da fonte que está disponibilizando os dados, uma vez que o BC é a principal instituição responsável por coordenar as operações de crédito no Brasil, considera-se que a mesma corresponde a uma instituição oficial e confiável, e como se trata de um braço do Ministério da Economia, ela é regida por normas e leis específicas de conduta que definem sua idoneidade; e o segundo aspecto tem relação com a obrigatoriedade do cadastramento dos contratos para liberação do financiamento, uma vez que não há possibilidade de fornecimento de crédito sem o devido registro pela instituição financeira, são mínimas as chances de haver subnotificações das operações realizadas, o que também afetaria na confiabilidade dos dados divulgados.

A **pontualidade** é uma dimensão da qualidade que pode ser observada com a frequência de disponibilização dos dados no portal do BC. Como pode ser observado na simulação de consultas utilizando os filtros de temporalidade (mês/ano) presentes na interface, os dados só

são disponibilizados quando encerra-se o mês por completo, ou seja, não há atualização diária dos valores, e para consultar dados referentes a um mês completo é necessário realizá-la no mês subsequente.

A **compreensibilidade** poderia ser observada a partir de algumas características que representam os dados disponibilizados nos recursos informacionais encontrados no portal do BC, como, por exemplo, formatação adequada para apresentação do valor do dado (ex: no valor total financiado do PRONAF, utilizando formato de moeda em reais), termos utilizados no rótulo das colunas (ex: abreviação de Unidade Federativa como UF), e o uso de linguagem técnica nos campos que permitem conjuntos de caracteres (ex: título das fontes de recursos). Como trata-se de uma política pública cujos beneficiários são pequenos produtores, e que este público geralmente apresenta um perfil com características socioculturais tornam o uso de tecnologia digital não tanto receptivo (ex: nível de escolaridade, faixa etária), ou que possui competências digitais ainda de forma incipiente (MOREIRA, 2020), esta dimensão merece uma atenção especial por parte da instituição para garantir as condições mínimas para compreensão do conteúdo considerando estas questões.

A questão da **granularidade** no processo de recuperação de dados do PRONAF no portal do BC poderia ser observada quanto ao nível da amplitude de detalhamento dos dados e quantidade de atributos, como, por exemplo, a regionalidade dos contratos (ex: regiões – Sudeste; estados – São Paulo; e municípios – Marília) e as características dos projetos financiados (ex: atividade – Agricultura; finalidade – Custeio; área financiada – Reais por hectare; e tipo de empreendimento – Café).

No processo de recuperação de dados do PRONAF no portal do Banco Central, identificaram-se conjuntos de dados disponibilizados em diversos **formatos de arquivo** (ex: PDF, CSV, XLS, HTML e .ASPX). Este aspecto também pode ser considerado na análise da qualidade dos dados, sendo que alguns destes formatos podem ser processados apenas por tecnologias proprietárias (ex: XLS), enquanto outros dificultam o reuso do conteúdo (ex: PDF). Na contra mão destes problemas, destaca-se a utilização do formato CSV, pois o mesmo permite uma certa flexibilidade na transferência de dados entre aplicações.

Quanto a disponibilização de **metadados** no processo de recuperação de dados do PRONAF no portal do BC, observou-se que esta caracteriza-se pela existência de elementos que representam os atributos dos conjuntos de dados disponíveis nos recursos informacionais, entretanto não foi verificado a utilização de padrões específicos de metadados para representar os *data sets*.

O **design** da informação é uma dimensão de qualidade que pode ser observada a partir do arranjo dos elementos informacionais que constituem a interface do portal do BC, incluindo a estruturados serviços de visualização presentes nos recursos informacionais. Assim como foi discutido anteriormente na dimensão da compreensibilidade, a questão do *design* também é um aspecto que deve receber atenção especial da instituição, considerando as características intrínsecas do perfil sociocultural que define os beneficiários do PRONAF e suas competências digitais, fornecendo um produto com linguagem visual aderente a este usuário.

No processo de recuperação de dados do PRONAF, a **responsividade** é uma dimensão que também poderia ser analisada a partir de uma perspectiva baseada na interface do portal do BC, bem como de seus recursos informacionais e consultas. Ao acessar as mesmas páginas e serviços em dois suportes digitais distintos (ex: smartphone e notebook), deve haver uma adaptação da interface do ambiente informacional sem que haja distorções ou alterações das funcionalidades. Considerando que os beneficiários deste programa são pequenos produtores e que, em sua maioria, possuem dispositivos móveis ao invés de computadores pessoais (PNAD, 2017), é importante que haja um grau de responsividade mínima para garantir sua utilização mediante este suporte.

A **segurança** é a dimensão da qualidade que poderia ser observadas na necessidade ou não de realização de cadastros para obter acesso os recursos informacionais e suas consultas, bem como a solicitação de informações pessoais no momento do acesso (ex: e-mail). No contexto do processo analisado, não se verificou nenhum destes aspectos, sendo que todos os recursos e consultas possuíam acesso aberto.

Por fim, a dimensão da qualidade definida como **personalização** poderia ser observada na recuperação de dados do PRONAF no portal do PC, analisando-se as funcionalidades dos serviços de consulta fornecidos na interface dos recursos informacionais disponíveis. Neste sentido, destacam-se algumas características da Matriz de Dados do Crédito Rural, uma vez que este serviço possibilita uma plataforma passível da configuração de parâmetros específicos para refinamento e filtragem dos dados (ex: mês início e fim, fonte de recurso, programa e subprograma).

7. Considerações finais

A pesquisa buscou analisar aspectos da qualidade considerando todos os níveis do processo de recuperação de dados da operacionalização do PRONAF no website do BC, principal programa governamental

executado para o segmento composto pelos pequenos produtores no Brasil.

Entende-se que, a ação de compartilhamento de dados atendendo a critérios de qualidade, pode proporcionar maiores possibilidades para os beneficiários do programa, beneficiários em potencial, pesquisadores que buscam por fontes de dados para realizar análises de políticas públicas, e de maneira geral para a sociedade, considerando os atores engajados com a transparência governamental e o exercício da cidadania.

Compreender como as dimensões da qualidade podem se manifestar na oferta de dados governamentais pode contribuir para os atores que detêm as bases de dados e disponibilizam o conteúdo, adotarem estratégias mais eficientes de disseminação. O governo, considerando a Lei de Acesso a Informação e os princípios da transparência ativa, deve executar seu papel na concepção e distribuição dos dados, garantindo as condições mínimas de qualidade, e possibilitando um melhor uso e reuso do conteúdo. Assim, aumentariam as possibilidades de surgirem novas demandas para a obtenção das informações e participação do cidadão nas políticas públicas.

A pesquisa foi fundamentada pelo ciclo de vida dos dados (SANT'ANA, 2016), que propõe uma visão para o fluxo de dados baseada em fases e objetivos específicos, sendo que a recuperação foi a fase na qual a pesquisa deu enfoque, e a qualidade foi o objetivo considerado.

Ressalta-se que esta pesquisa não teve como objetivo analisar questões referentes a encontrabilidade e usabilidade das informações (outros conceitos de ampla relevância para a Ciência da Informação), nem a pretensão de apontar finalidades de usos e manipulações para os dados encontrados, mas teve como intuito principal a observância da conformidade dos dados no cenário estudado para fornecer uma estrutura de análise que possa ser replicada por outros estudos.

Alguns desdobramentos podem ser elencados, como a aplicação das dimensões da qualidade definidas neste artigo em outras fontes mantidas por outras instituições que também disponibilizam dados do PRONAF, bem como serem aplicadas na investigação de processos de recuperação de dados de outras políticas públicas.

Referências

ABATH, M. de B. *et al.* Avaliação da completude, da consistência e da duplicidade de registros de violências do Sinan em Recife, Pernambuco, 2009-2012. *Epidemiol. Serv. Saúde [online]*, v.23, n.1, 2014. p.131-142.

AGÊNCIA PARA MODERNIZAÇÃO ADMINISTRATIVA. *Guia Dados Abertos*. 2016. Disponível em:

http://www.dados.gov.pt/media/46120/guia_dados_abertos_ama.pdf.

Acesso em: 07maio2019.

ANDERSEN, K. V.; HENRIKISEN. H. Z. E-Government Maturity Models: Extension of the Layne and Lee Model. *Government Information Quarterly*, v.23, n.1, p.236-248, 2006. Disponível em:

<http://www.sciencedirect.com/science/article/pii/S0740624X05000973>.

Acesso em: 07 maio 2019.

BANCO CENTRAL DO BRASIL [BC]. *O Banco Central do Brasil – Bacen. Banco Central do Brasil*. 2014. Disponível em:

<http://www.bcb.gov.br/pre/portalCidadao/bcb/bcFaz.asp?idpai=LAIINSTITUCIONAL>. Acesso em: 07 maio 2019.

BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 2009.

BARNES, S. J.; VIDGEN, R. T. Data triangulation and web qualitymetrics: A case study in e-government. *Information and Management*, n. 34, 2006. 767-777. Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S03787206060053X>. Acesso em: 07 maio 2019.

BARCELLOS, L. P. Timeliness no Brasil: um estudo dos determinantes do prazo de divulgação das demonstrações contábeis das companhias não financeiras listadas na BM&FBOVESPA. 2013. *Dissertação* (Mestrado em Ciências Contábeis), Universidade Estadual do Rio de Janeiro - UERJ, Rio de Janeiro, 2013. Disponível em:

http://bdtd.ibict.br/vufind/Record/UERJ_380af28a66e544af763080a2c212ad19. Acesso em: 02 maio 2020.

BRASIL. *Lei número 12.527*, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. Disponível em:

http://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/lei/l12527.htm. Acesso em: 07 maio2019.

COMISIÓN ECONÓMICA PARA AMÉRICA LATINA Y EL CARIBE (CEPAL). *Gestión orientada a asegurar la calidad de los datos en los institutos nacionales de estadística*. Santiago de Chile: CEPAL; 2003.

COMITÊ EXECUTIVO DO GOVERNO ELETRÔNICO (CEGE). *Padrão de metadados do Governo Eletrônico (e-PMG 1.1)*. CEGE, 2014. Disponível em: <http://www.governoeletronico.gov.br>. Acesso em: 02 mar. 2020.

CONTI, B. M.; ROITMAN, F. B. PRONAF: uma análise da evolução das fontes de recursos utilizadas no programa. *Revista do BNDES*, v.1, n.1, p.131-168, 2011. Disponível em: <https://web.bndes.gov.br/bib/jspui/handle/1408/3364>. Acesso em: 07 maio 2019.

DROHOMERETSKI, E.; FAVARETTO, F. Um levantamento das causas e efeitos da falta de acurácia nos estoques: um estudo exploratório. *Revista Gestão Industrial*, Ponta Grossa, v.6, n.2, p. 142-158, 2010.

FARAH, M. F. S. Disseminação de políticas públicas e programas governamentais no nível subnacional de governo. *Revista de Administração e Diálogo*, v.2, n.11, p.69-89, 2008. Disponível em: <http://revistas.pucsp.br/index.php/rad/article/view/2714>. Acesso em: 07 maio 2019.

GIL-GARCÍA, J. R.; PARDO, T. A. E-government success factors: Mapping practical tools to theoretical foundations. *Government Information Quarterly*, n.22, 2005. p.187-216. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0740624X05000158>. Acesso em: 07 maio 2019.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA [IBGE]. *Pesquisa Nacional por Amostra de Domicílios Contínua – PNAD*. IBGE, 2017. Disponível em: <https://www.ibge.gov.br/estatisticas/multidominio/condicoes-de-vida-desigualdade-e-pobreza/17270-pnad-continua.html>. Acesso em: 20 abr. 2020.

JACOBSON, R. *Information Design*. Cambridge, Massachusetts: The MIT Press, 1999.

JANOWICZ, K. et al. Geospatial semantics and linked spatiotemporal data: past, present, and future. *Semantic Web*, v. 3, n. 4, p. 321-332, 2012.

KHALEGHI, B.; KHAMIS, A.; KARRAY, F. O.; RAZAVI, S. N. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*.v.14, n.1, p. 28-44, 2013.

KOUTRIKA, G.; IOANNIDIS, Y. Personalized Queries under a Generalized Preference Model. *In: ICDE*, 21. Tokyo. *Anais [...]*, p.841-852, 2005.

LOURENÇO, R. P.; An analysis of open government portals: A perspective of transparency for accountability, *Government Information Quarterly*, v. 32, n. 3, pp. 323-332, 2015.

MARTINES USERO, J. A. *Nuevas tecnologías para nuevas bibliotecas: desarrollo de servicios de información electrónica*. Buenos Aires: Alfagrama, 2007. 189 p.

MEITEI, L. S.; DEVI, T. P. Farmers Information Needs in Rural Manipur: An Assessment. *Annals of Library and Information Studies*, v.56, n.1, p.35-40, 2009. Disponível em: <https://pdfs.semanticscholar.org/b3f5/95677754e76d1386f01032632032b12be0f6.pdf>. Acesso em: 07 maio 2019.

MOREIRA, F. M. Recuperação de dados de políticas públicas na Internet pelo pequeno produtor: elementos constitutivos do campo informacional. 2020. Tese (Ciência da Informação) – Faculdade de Filosofia e Ciências (FFC/UNESP), Marília/SP, 2020.

SECRETARIA DE AGRICULTURA FAMILIAR E COOPERATIVISMO. *Plano Safra da Agricultura Familiar 2017/2020*. 2017. Disponível em: http://www.mda.gov.br/sitemda/sites/sitemda/files/user_img_1684/3Baixa_Cartilha_Plano_Safra_2017.pdf. Acesso em: 07 maio 2019.

MOREIRA, F. M.; SANT'ANA, R. C. G. Aspectos da disseminação no processo de recuperação de dados de programas governamentais: Um estudo do PRONAF. In: ENANCIB, 16, João Pessoa, 2016. *Anais [...]2016*. Disponível em: <http://www.ufpb.br/evento/lti/ocs/index.php/enancib2015/enancib2015/aper/viewFile/2910/1179>. Acesso em: 07 maio 2019.

MOREIRA, F. M.; SANT'ANA, R. C. G.; JORENTE, M. J. A complexidade na disponibilização e acesso a dados governamentais na Web. *Perspectivas em Ciência da Informação*, Belo Horizonte, v.21, n.3, 2016. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2540>. Acesso em: 07 maio 2019.

MORGAN, S. L.; WARING, C. G. *Guidance on Testing Data Reliability*. 2004. Disponível em: <https://www.gao.gov/assets/80/77213.pdf>. Acesso em: 10 agosto 2018.

OLETO, R. R. Percepção da qualidade da informação. *Ciência da Informação*, Brasília, v.35, n.1, p. 57-62. 2006.

OLIVEIRA, H. P. C. de. Arquitetura da informação pervasiva: contribuições conceituais. 2013 [i.e. 2014]. 202 f. Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências de Marília, 2013 [i.e. 2014]. Disponível em: <http://hdl.handle.net/11449/110387>. Acesso em: 07 maio 2019.

PLATT NETO, O. A. P.; CRUZ, F.; VIERA, A. L. Transparência das contas públicas: um enfoque no uso da Internet como instrumento de publicidade na UFSC. *Revista Contemporânea de Contabilidade*, v.1, n.5, p.135-146, 2006. Disponível em: <https://periodicos.ufsc.br/index.php/contabilidade/artide/view/1143>. Acesso em: 07maio2019.

POLZER, H.; SCHNEIDER, T.; WILSON, B. *US DoD Net-Centric Attributes*. 2011. Disponível em: http://www.ncoic.org/images/technology/whitepapers/NCA_Invited_Review_v2.0_20110216A.pdf. Acesso em: 07maio2019.

RASYID, A.; ALFINA, I. E-Services Quality Evaluation on E-Government Website: Case Study BPJS Kesehatan Indonesia. *In: INTERNATIONAL CONFERENCE ON COMPUTING AND APPLIED INFORMATICS*, 2016. *IOP Conf. Series: Journal of Physics*, n.801, 2017, Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/801/1/012036>. Acesso em: 07 maio 2019.

RODRIGUES, F. de A.; SANTANA, R. C. G. Restrições tecnológicas e de acesso a dados disponíveis sobre destino de repasse financeiros federais para a saúde pública em ambientes informacionais digitais. In: Encontro Nacional de Pesquisa em Ciência da Informação, 13. Rio de Janeiro. *Anais...*, 2012. Disponível em: <http://www.eventosecongressos.com.br/metodo/enancib2012/areas-trita/pdfs/19435.pdf>. Acesso em: 07 maio 2019.

SADIQ, S.; INDULSKA, M. Open Data: Quality over quantity. *International Journal of Information Management*, n.37, 2017. p.150-154. Disponível em: <http://dx.doi.org/10.1016/j.ijinfomgt.2017.01.003>. Acesso em: 07 maio 2019.

SANT'ANA, R. C. G. Ciclo de Vida dos Dados: uma perspectiva a partir da Ciência da Informação. *Informação & Informação*, v.21, n.2, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/artide/view/27940>. Acesso em: 07 maio 2019.

SANT'ANA, R. C. G. *Tecnologia e Gestão Pública Municipal: mensuração da interação com a sociedade*. São Paulo: Cultura Acadêmica Editora - UNESP, 2009. 181p.

SANTOS, P. L. A. C.; SANT'ANA, R. C. G. Dado e granularidade na perspectiva da informação e tecnologia: uma interpretação pela ciência da informação. *Ciência da Informação*, Brasília, v.42, n.2, 2013. Disponível em: revista.ibict.br/ciinf/article/view/1382. Acesso em: 07 maio 2019.

SILVA, J. P. L.; TRAVASSOS, C. M. R.; VASCONCELLOS, M. M. de ; CAMPOS, L. M. Revisão sistemática sobre encadeamento ou linkage de bases de dados secundários para uso em pesquisa em saúde no Brasil. *Cadernos Saúde Coletiva (UFRJ)*, v.14, p. 197-224, 2006.

TODORAN, I.; HUSSAIN, Z.; GROMOV, N. SOA Integration modeling: na evaluation of how soa ML Completes UML modeling. *In: IEEE INTERNATIONAL ENTERPRISEDISTRIBUTED OBJECT COMPUTING CONFERENCE WORKSHOPS*, 15, [S.l.], 2011. p. 57-66.

VAN RIJSBERGEN, C. J. *Information Retrieval*. 2.ed. Newton: Butterworth-Heinemann, 1979. 224p.

WANG, R.; STRONG, D. Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*. p. 5-33. 1996.

WEBER, E.; ANZOLCH, R.; LISBOA FILHO, J.; COSTA, A. C.; IOCHPE, C. *Qualidade de dados geoespaciais*. UFRGS, 1999. Disponível em: <http://www.dpi.ufv.br/~jugurta/papers/RELPESQUAL.PDF>. Acesso em: 27 abr. 2020.

YUSOP, N. I.; IBRAHIM, H.; YUSOF, S. A. M.; AJI, Z. M.; DAHALIN, Z. D.; GHAZALI, O.; SAAD, M. N.; ABU, M. A. Information Needs of Rural Communities. *The Journal of Community Informatics*.v.9, n.1, 2013. Disponível em: <http://ci-journal.net/index.php/ciej/article/view/586/965>. Acesso em: 07 maio 2019.