

# Qualidade em Cadastros Públicos de Pessoas: uma proposta de guia para avaliação

**Luisa Helena Santos Franco**

---

Prof. Dr. Rafael Silveira e Silva

Coletânea de Pós-Graduação  
**Políticas Públicas**



REPÚBLICA FEDERATIVA DO BRASIL  
TRIBUNAL DE CONTAS DA UNIÃO

---

**MINISTROS**

Ana Arraes (Presidente)  
Bruno Dantas (Vice-Presidente)  
Walton Alencar Rodrigues  
Benjamin Zymler  
Augusto Nardes  
Aroldo Cedraz  
Raimundo Carreiro  
Vital do Rêgo  
Jorge Oliveira

**MINISTROS-SUBSTITUTOS**

Augusto Sherman Cavalcanti  
Marcos Bemquerer Costa  
André Luis de Carvalho  
Weder de Oliveira

**MINISTÉRIO PÚBLICO JUNTO AO TCU**

Cristina Machado da Costa e Silva (Procuradora-Geral)  
Lucas Furtado (Subprocurador-Geral)  
Paulo Soares Bugarin (Subprocurador-Geral)  
Marinus Eduardo de Vries Marsico (Procurador)  
Júlio Marcelo de Oliveira (Procurador)  
Sérgio Ricardo Costa Caribé (Procurador)  
Rodrigo Medeiros de Lima (Procurador)



**DIRETORA-GERAL**

Ana Cristina Melo de Pontes Botelho

**DIRETORA DE RELAÇÕES INSTITUCIONAIS,  
PÓS-GRADUAÇÃO E PESQUISAS**

Flávia Lacerda Franco Melo Oliveira

**CHEFE DO DEPARTAMENTO  
DE PÓS-GRADUAÇÃO E PESQUISAS**

Clemens Soares dos Santos

**CONSELHO ACADÊMICO**

Maria Camila Ávila Dourado

Tiago Alves de Gouveia Lins e Dutra

Marcelo da Silva Sousa

Rafael Silveira e Silva

Pedro Paulo de Moraes

**COORDENADORA ACADÊMICA**

Renata Miranda Passos Camargo

**COORDENADORES PEDAGÓGICOS**

Pedro Paulo de Moraes

Flávio Sposto Pompêo

**COORDENADORA EXECUTIVA**

Maria das Graças da Silva Duarte de Abreu

**PROJETO GRÁFICO E CAPA**

Núcleo de Comunicação – NCOM/ISC

# **Qualidade em Cadastros Públicos de Pessoas: uma proposta de guia para avaliação**

**Luisa Helena Santos Franco**

Monografia de conclusão de curso submetida ao Instituto Serzedello Corrêa do Tribunal de Contas da União como requisito parcial para a obtenção do grau de especialista.

**Orientador(a):**

Prof. Dr. Rafael Silveira

**Banca examinadora:**

Prof. Me. Wesley Vaz

## REFERÊNCIA BIBLIOGRÁFICA

FRANCO, Luisa Helena Santos. **Qualidade de Cadastros Públicos de Pessoas: uma proposta de guia para avaliação**. 2022. Monografia (Especialização em Avaliação de Políticas Públicas) – Instituto Serzedello Corrêa, Escola Superior do Tribunal de Contas da União, Brasília DF. 79 fls.

## CESSÃO DE DIREITOS

NOME DO(A) AUTOR(A): Luisa Helena Santos Franco

TÍTULO: Qualidade de Cadastros Públicos de Pessoas: uma proposta de guia para avaliação

GRAU/ANO: Especialista/2022

É concedido ao Instituto Serzedello Corrêa (ISC) permissão para reproduzir cópias deste Trabalho de Conclusão de Curso e emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Do mesmo modo, o ISC tem permissão para divulgar este documento em biblioteca virtual, em formato que permita o acesso via redes de comunicação e a reprodução de cópias, desde que protegida a integridade do conteúdo dessas cópias e proibido o acesso a partes isoladas desse conteúdo. O autor reserva outros direitos de publicação e nenhuma parte deste documento pode ser reproduzida sem a autorização por escrito do autor.

---

Luisa Helena Santos Franco  
luisahf@tcu.gov.br

## FICHA CATALOGRÁFICA

L131a Franco, Luisa Helena Santos

Qualidade em Cadastros Públicos de Pessoas: uma proposta de guia para avaliação / Luisa Helena Santos Franco. – Brasília: ISC/TCU, 2022.  
79 fls. (Monografia de Especialização)

1. Avaliação de Políticas Públicas. 2. Tema 2. 3. Tema 3. I. Título.

CDU 02  
CDD 020

# Qualidade em Cadastros Públicos de Pessoas: uma proposta de guia para avaliação

**Luisa Helena Santos Franco**

Trabalho de conclusão do curso de pós-graduação lato sensu em Avaliação de Políticas Públicas realizado pelo Instituto Serzedello Corrêa como requisito para a obtenção do título de especialista.

Brasília, 8 de junho de 2022.

**Banca Examinadora:**

---

Prof. Rafael Silveira e Silva, Dr.  
Orientador  
Senado Federal

---

Prof. Wesley Vaz, Me.  
Avaliador  
Tribunal de Contas da União

#### DAS UTOPIAS

Se as coisas são inatingíveis... ora!  
Não é motivo para não querê-las...  
Que tristes os caminhos, se não fora  
A presença distante das estrelas!

Mário Quintana (Espelho Mágico)

Apesar da verdade ser exata e clara em si própria, quando chega até nós se torna vaga, pois é tecnicamente invisível. O bom é que a verdade chega a nós como um sentido secreto das coisas. Nós terminamos adivinhando, confusos, a perfeição.

Clarice Lispector (A descoberta do Mundo)

## Resumo

Cadastros públicos de pessoas são importantes instrumentos para a implementação de políticas públicas. Baixa qualidade de dados nesses cadastros impacta a efetividade de políticas públicas de diversas formas: pode comprometer a elegibilidade em programas sociais, aumentar o custo de investimentos públicos, deixar excluídos pessoas no público-alvo de ações do Estado, entre outros exemplos. Por isso, é importante avaliar a qualidade desses cadastros e promover sua melhoria contínua. Para isso, é importante definir como se entende qualidade nesse contexto, identificar os elementos que a compõem e organizar meios para medi-los. Neste trabalho, são apresentadas as dimensões de qualidade aplicáveis a cadastros públicos sobre pessoas e apresentado um guia de como avaliar essa qualidade.

**Palavras-chave:** Cadastros públicos; qualidade de dados; avaliação; indicadores.



## Abstract

Government registers of people are important instruments for the implementation of public policies. Low quality of data in these registers impacts the effectiveness of public policies in several ways: it can compromise eligibility for social programs, increase the cost of public investments, exclude people in the target audience of State actions, among other examples. Therefore, it is important to assess the quality of these registers and promote their continuous improvement. For this, it is important to define how quality is understood in this context, identify the elements that compose it and organize means to measure them. In this work, the quality dimensions applicable to government registers of people are presented and a guide on how to assess this quality is presented.

**Keywords:** public registries; data quality; assessment; indicators.

## Lista de figuras

Figura 1 Gastos do governo federal em 2020 com covid-19 .....	15
Figura 2 Cadastros usados no auxílio emergencial .....	15
Figura 3 – Conceitos nas Dimensões de Qualidade de Dados .....	23
Figura 4 – Possíveis consumos para um cadastro social .....	25
Figura 5 Ciclo de vida de cadastros públicos sobre pessoas .....	26
Figura 6 – Completude de público-alvo em cadastros sociais (2017-2017) .	40
Figura 7 - Trecho de formulário para preenchimento do CadÚnico, com indicação de consistência de alguns campos.....	42
Figura 8 – Famílias sem atualização no Cadastro Único.....	49
Figura 9 - Qualidade de dados no contexto das organizações.....	65

## Lista de quadros

Quadro 1 – Dimensões de qualidade contabilizadas em levantamento de vários autores.....	20
Quadro 2 - Dimensões de qualidade para dados abertos em vários autores .....	21
Quadro 3 - Dimensões de qualidade em pesquisa ontológica em vários autores .....	22
Quadro 4 - Programas que utilizam dados do CadÚnico.....	33

## Sumário

<b>1.</b>	<b>Introdução .....</b>	<b>13</b>
<b>2.</b>	<b>Problema e justificativa.....</b>	<b>13</b>
<b>3.</b>	<b>Objetivos .....</b>	<b>16</b>
<b>4.</b>	<b>Metodologia .....</b>	<b>17</b>
<b>5.</b>	<b>Referencial Teórico .....</b>	<b>18</b>
5.1.	Sobre qualidade de dados.....	18
5.2.	Sobre bases de dados e políticas públicas.....	24
<b>6.</b>	<b>Desenvolvimento .....</b>	<b>29</b>
6.1.	Cadastros públicos sobre pessoas .....	29
6.1.1.	Contabilização de nascimento e morte dos cidadãos: o SIRC.....	31
6.1.2.	Cidadão-cliente do Estado Social: o Cadastro Único .....	32
6.1.3.	Eleitores: a base do TSE .....	34
6.1.4.	Contribuinte: a base do CPF.....	35
6.2.	Qualidade de cadastros públicos de pessoas .....	36
6.2.1.	Dimensões de qualidade dos dados.....	36
6.2.2.	Processos que afetam a qualidade dos cadastros .....	60
<b>7.</b>	<b>Síntese da proposta de uma ferramenta de avaliação .....</b>	<b>67</b>
<b>8.</b>	<b>Considerações finais.....</b>	<b>73</b>
	<b>Referências bibliográficas.....</b>	<b>76</b>

## 1. Introdução

O presente trabalho se propõe a apresentar, diante do problema da avaliação de qualidade dos cadastros públicos sobre pessoas, um instrumento para apoio a tal avaliação, com o objetivo de possibilitar o aprimoramento das informações desses cadastros tão importantes para a execução de políticas públicas.

## 2. Problema e justificativa

Conforme lembra Morton (2021), as relações entre os governos e suas populações podem variar bastante no tempo e no espaço (e têm variado) de acordo com a maneira e os valores que as sociedades adotam para se enxergar, tanto em relação a si mesmas, como em relação a este ator social chamado Estado. Essa mudança do papel do Estado ao longo da história resulta também na necessidade de diferentes instrumentos para execução das funções que lhe cabem.

Se nos estados absolutistas a função primordial do estado era a defesa territorial, as principais informações que interessavam aos governos diziam respeito às terras e ao poderio bélico. Ao longo dos séculos, progressivamente, cada sociedade foi instituindo políticas públicas para atendimento das demandas em Saúde, Educação, Trabalho, Transporte, Moradia e Assistência Social (JANUZZI, 2020). A formação dos conceitos modernos relacionados ao chamado “Estado de bem-estar social” com características mais solidárias ou mais individualizadas para provimento (e financiamento) dos serviços de proteção – com toda escala de implementação real que existe entre essas duas características – trouxeram o foco das necessidades de informação para o ator do outro lado dessa relação: o cidadão.

Atualmente, os governos e a execução das políticas públicas dependem de maneira estruturante da qualidade das informações sobre pessoas – os cidadãos – mantidas pela burocracia estatal. Januzzi (2020) lembra que

Políticas sociais são muito intensivas em informação no processo de seu desenho e implementação. Elas se estruturam, no caso brasileiro, como sistemas complexos, articulando programas de natureza universal com ações redistributivas em várias áreas setoriais, operados por agentes em diferentes níveis federativos de governo, em contextos desiguais em termos de capacidade de gestão e de perfil socioeconômico de públicos-alvo.

Assim, se é hoje entendido como dever do Estado garantir aos indivíduos condições mínimas nas áreas de saúde, educação, habitação, seguridade social, entre outras, bem como atuar em momentos de crise (como na pandemia da covid-19), isso não é possível sem bons cadastros públicos sobre pessoas.

De fato, não só o Brasil como muitos países oferecem uma infinidade de benefícios e serviços sociais para atender às diversas necessidades de suas populações: transferências de renda, serviços especiais para crianças, jovens, idosos ou pessoas com necessidades especiais, políticas de trabalho, regulação de serviços urbanos etc. Embora esses serviços sejam bem diferentes, possuem alguns instrumentos em comum, dentre os quais o cadastro de pessoas é de especial relevância, pois permite conhecer seu público-alvo e determinar sua elegibilidade (LEITE, 2017).

No entanto, não apenas as políticas sociais dependem dessas informações. De fato, os cadastros públicos de pessoas físicas suportam a divulgação, captação, registro e determinação de elegibilidade potencial para um ou mais programas ou políticas sociais, mas as várias lentes com as quais o Estado vê a pessoa – cidadão, contribuinte, assalariado, eleitor, aposentado, motorista, estudante, servidor público, político etc. – materializam-se nos cadastros públicos sobre pessoas.

É evidente, então, que a qualidade de tais cadastros impacta a execução dessas políticas públicas: uma informação errada de renda retira uma família elegível do Bolsa Família; a incorreta identificação de uma pessoa como trabalhador ao longo de um período impede o recebimento de um benefício previdenciário; dados incorretos de contribuinte geram cobranças indevidas, para mais ou para menos; um óbito não registrado permite que um benefício seja recebido indevidamente, aumentando os custos do programa associado ao benefício; a falha de uma carteira de motorista pode impedir a execução de atividade laboral.

Tomando como exemplo apenas as ações do governo em 2020 em função da pandemia de covid-19, a figura a seguir mostra que a dimensão dos investimentos no tema foi da ordem de R\$ 524 bilhões.

**Figura 1 Gastos do governo federal em 2020 com covid-19**

Fonte: tesouro transparente.

**Figura 2 Cadastros usados no auxílio emergencial****3º PASSO**

A DATAPREV cruza os dados recebidos com os mais de 33 bilhões de registros do CNIS.



Fonte: Dataprev

Somente o Auxílio Emergencial representou gasto de R\$ 293 bilhões. Esta ação fez uso intensivo de cadastros públicos, entre eles o CNIS, o Cadastro Único e a base CPF da Receita.

Acompanhamentos realizados pelo TCU no contexto do “Coopera TCU” apontaram indícios de que recebimento do auxílio emergencial por requerentes sem direito ao benefício pode ter atingido o montante aproximado de R\$ 45,7 bilhões

(TCU, 2021). Nestas fiscalizações, foram apontados vários problemas e inconsistências cadastrais, que resultaram em pagamentos efetuados para beneficiários falecidos, que acumulavam outros benefícios ou que tinham emprego regular. Além disso, identificou-se pagamentos feitos a beneficiários cujos CPF’s estavam cancelados, nulos ou suspensos na base da Receita (TCU, 2020). Tais ocorrências ilustram os problemas que as falhas nos cadastros públicos podem gerar.

### 3. Objetivos

Com o objetivo de organizar itens que componham instrumento de apoio a avaliação de qualidade dos cadastros públicos sobre pessoas no Brasil, são descritas as principais dimensões de qualidade, a partir de revisão bibliográfica sobre estudos já realizados sobre o tema, sempre pautando pela sua aplicação em cadastros de pessoas.

Além disso, a partir de análise sobre relevantes cadastros públicos de pessoas disponíveis no TCU e que já foram objeto de auditorias, é realizado confronto com as dimensões de qualidade identificadas, e aplicadas questões delas decorrentes. Vale dizer que tal confronto não se dará de modo exaustivo para todas as questões em todos os cadastros. Ao contrário, serão utilizados exemplos mais ilustrativos de questões de qualidade em cada cadastro, de tal forma a auxiliar na compreensão de cada dimensão da qualidade aqui abordada. O objetivo de tal estratégia é mostrar, de forma didática para o leitor, como falhas numa dada dimensão impactam no caso real de cadastros públicos.

Hoje, o Tribunal de Contas da União já faz avaliações de várias bases de dados com cadastros que são suporte para políticas públicas. Uma vez que é preciso verificar constantemente a qualidade desses cadastros no âmbito da avaliação das políticas públicas por eles suportadas, pretende-se que o conhecimento acadêmico aqui sistematizado subsidie tal avaliação e, conseqüentemente, contribua para a melhoria da qualidade desses cadastros.

Embora não exclusivamente, a perspectiva dessa avaliação será principalmente a do controle externo. O objetivo é dialogar com a prática já executada e o aprendizado organizacional existente, visando que proposituras de determinações e recomendações técnicas estejam sintonizadas com boas práticas e legitimada pela produção acadêmica do assunto.

Finalmente, vale explicitar que não será tratado aqui sobre o processo de prover e manter qualidade de dados em cadastros públicos de pessoas. Esse processo está ligado à governança dos dados e pode envolver ações em diversos níveis. Aqui, o foco será a avaliação da qualidade e em como fazer essa avaliação – que é uma parte desse processo de gerenciamento de qualidade e governança de dados.



## 4. Metodologia

Para formulação da proposta do checklist de qualidade, a estratégia foi a pesquisa de publicações de referência contendo dimensões de qualidade de dados combinada com a análise comparada dessas dimensões aplicadas a cadastros públicos de pessoas no Brasil, partindo das experiências do próprio TCU em auditorias sobre o tema.

Segundo Fachin (2006), “o método comparativo consiste em investigar coisas ou fatos e explicá-los segundo suas semelhanças e suas diferenças”, ou seja: analisam-se dados concretos e deduzem-se os elementos constantes, abstratos e gerais. São investigações de caráter indireto com as quais se almeja algum nível de generalizações úteis para futuras investigações. Por isso, considerou-se adequado que a análise das dimensões de qualidade fosse confrontada com alguns exemplos de cadastros públicos de pessoas no Brasil, permitindo observar sua aplicabilidade e utilidade, embasando sua inclusão na proposta do checklist.

Para a busca de artigos publicados em revistas científicas foram consultadas principalmente as bases de produção científica do Scielo, Capes e O’Reilly, incluindo publicações sobre estudos de cadastros específicos disponíveis, como o Cadastro Único, bases de eleitores e CPF. Adicionalmente, foram verificadas as referências bibliográficas dos artigos identificados por essas fontes.

Fonte importante de pesquisa também foram as auditorias do próprio TCU sobre cadastros públicos, especialmente da última década, nas quais se pôde observar os achados comuns e seus impactos nas políticas públicas a eles relacionadas.

Os cadastros que exemplificam as dimensões de qualidade foram escolhidos pela disponibilidade de acesso às bases e às auditorias já realizadas sobre eles pelo TCU, bem como sua abrangência e relevância em políticas públicas no Brasil hoje.

## 5. Referencial Teórico

### 5.1. Sobre qualidade de dados

O primeiro desafio teórico deste trabalho consiste na própria definição de qualidade – e especialmente qualidade de dados armazenados em bases, normalmente informatizadas.

Vale lembrar que não há consenso na literatura especializada sobre as definições e dimensões que constituem a qualidade de dados. No campo da ciência da informação, há mesmo críticas e dúvidas sobre a operacionalidade desse conceito e até uma certa recorrência em afirmar que as definições de qualidade de dados são ambíguas, vagas ou subjetivas”. (PAIM et al, 1996).

Há abordagens que percebem o dado informacional como um objeto, de tal forma que é possível definir um conjunto de características intrínsecas à informação para representar sua qualidade (CICHY & RASS, 2019). Essa abordagem é bastante aderente quando se trata de informação armazenada em bases de dados e permite, de forma razoavelmente objetiva, mensurar aspectos ou dimensões de qualidade dos dados dessas bases. Com essa perspectiva, estudos sobre gestão e governança de dados ancoram-se na premissa de que é possível controlar e alavancar eficazmente o uso desses ativos informacionais para atender às necessidades das organizações ao prover disponibilidade, segurança e qualidade aos dados. (BARBIERI, 2013).

Uma síntese conhecida de estudos e práticas de governança de dados é mantida pela Global Data Management Community – DAMA, uma organização internacional sem fins lucrativos, constituída por profissionais e técnicos dedicados a promover os conceitos e melhores práticas da governança de dados e gestão da informação. A DAMA é responsável pelo DMBok (*Data Management Body of Knowledge*), um compilado organizado em 2009 (e revisado em 2017) com a participação de mais de 120 profissionais de todo o mundo que agrupa padrões, terminologias e boas práticas em gerenciamento de dados, e coloca a gestão da qualidade da informação como um dos processos críticos para esse gerenciamento (DAMA, 2009). Nesse contexto, apresenta seis dimensões principais para a qualidade de dados: completude, unicidade, tempestividade, validade, acurácia (ou precisão) e consistência (DAMA-UK,2013), além de usabilidade, flexibilidade, valor e confiança – estas últimas não chamadas de dimensões mas de “características que impactam na

qualidade” (DAMA-UK, 2013). Já na versão revisada do DMBok (DAMA, 2015), observa-se que existem dimensões que podem ser verificadas intrinsecamente aos dados e outras, que dependem do contexto – a exemplo de reputação, confiança e usabilidade – e que não existe uma “lista única” sobre o tema, apesar da convicção de que alguma lista tem que ser definida como ponto de partida para a avaliação de qualidade.

De fato, há na literatura referência a outras dimensões variando entre estudos e autores, e ainda interpretações diferentes para uma mesma dimensão referenciada. Independente da nomenclatura, trata-se de esforço para tornar a qualidade de dados mais precisa, a fim de que se possa avaliá-la e aprimorá-la. Guedes et al (2021) sintetizam levantamento que utilizou protocolo PRISMA para revisão sistemática de literatura (livros e artigos), com um corpus final de 18 documentos, e apresentam o Quadro 1 de dimensões de qualidade identificadas. Neste quadro, as marcações indicam a repetição dos conceitos nos documentos, destacando-se então: precisão, completude e confiabilidade.

Estudo similar realizado por Slibar et al (2021), no contexto de dados abertos – isto é, dados amplamente disponíveis na Internet para consumo, desenvolvimento e produção de conhecimento – chegou a outro conjunto de dimensões para qualidade, com pontos em comum. O Quadro 2 apresenta as dimensões encontradas, a partir de revisão sistemática em um corpus de 34 documentos. Observa-se que, nesse contexto, se destacam completude, acurácia, consistência e acessibilidade.

**Quadro 1 – Dimensões de qualidade contabilizadas em levantamento de vários autores**

INDICADORES		Acessibilidade	Atualização	Cobertura	Comparabilidade	Completude	Contabilidade	Conformidade	Consistência	Credibilidade	Desempenho	Detalhamento	Escalabilidade	Formato adequado	Integração de dados	Compreensibilidade	Manutenção	Navegabilidade	Objetividade	Pontualidade	Precisão	Relevância	Representação	Segurança	Semântica	Singularidade	Tolerância a falha	Usabilidade	Volatilidade	
		AUTORES ANALISADOS	Ayyash 2017				*													*		*		*						
Azeroual et al 2020					*															*	*	*								
Behkamal et al 2014			*		*																*									
Bibi, Jannisar e Inayet, 2016						*			*																		*			
Bizer 2007					*	*			*													*		*	*					
Harting e Lewoniewski 2020			*		*																*									
Josang e Ismail 2006						*															*									
Knight e Burn 2005					*	*			*	*					*						*	*								
Manzoor e Truong e Dustdar 2014						*															*	*						*		
Mohammadi, Abrizah e Nazari 2015					*	*															*	*						*		
Nehmy e Paim 1998			*	*																	*	*								
Pinto, Guerreiro Quesada, Granell, 2014	*		*												*				*											
Reiter et al. 2014						*	*			*											*	*								
Reznik e Lyshevski 2015					*	*															*	*							*	
Rukkas e Zhoitkevych 2015						*															*	*								
Sá e Martins 2016	*					*						*		*				*						*				*		
Saracevic 1996	*				*	*							*		*		*				*	*								
Sugahara e Januzzi 2005					*														*		*				*					

Fonte: Guedes et al, 2021, p.

Quadro 2 - Dimensões de qualidade para dados abertos em vários autores

Reference	Completeness	Accuracy	Consistency	Accessibility	Timeliness	Usage	Retrievability	Openness	Transparency	Understandability
Batini and Scannapieco (2016)	✓	✓	✓	✓	✓					
Bellini and Nesi (2013)	✓	✓	✓							
Caracas et al. (2018)					✓					
Corsar and Edwards (2017)			✓						✓	
Csáki and Prier (2018)				✓		✓				
De Blasio and Selva (2018)									✓	
Fan and Zhao (2017)	✓									
Rani et al. (2018)					✓					
Hjalmarsson et al. (2015)			✓	✓						
Jarolimek and Martinec (2016)								✓		
Karafili et al. (2019)	✓	✓	✓							
Knap (2016)							✓			
Kučera et al. (2013)	✓	✓	✓		✓					
Langer et al. (2018)				✓						
Wang et al. (2013)	✓									
Martin et al. (2017)			✓							
Matamoros et al. (2018)	✓									
Neumaier et al. (2018)		✓				✓	✓	✓		
Neumaier et al. (2016)	✓	✓								
Ren and Glissmann (2012)	✓		✓	✓	✓					
Sadiq and Indulska (2017)	✓		✓							
Sanabria et al. (2018)	✓	✓				✓	✓	✓		
Sayogo et al. (2014)	✓									
Sheppard and Terveen (2011)		✓	✓							
Strozyna et al. (2018)				✓						
Kubler et al. (2018)	✓									
Torchiano et al. (2017)	✓	✓	✓							✓
Trifonov et al. (2017)				✓						
Utamachant and Anutariya (2018)										✓
Vetrò et al. (2016)		✓								
Viscusi et al. (2014)	✓				✓					
Yang et al. (2015)				✓						
Yi (2019)		✓								
Zheng and Gao (2016)					✓	✓		✓		

Fonte: Slibar et al, 2021, p.

Observa-se que, se não há concordância entre os autores sobre qual é a “lista definitiva” de dimensões para qualidade de dados, o ponto em comum é o esforço em definir mais precisamente o que é qualidade, bem como a consciência de que tal definição sempre dependerá do contexto e do objetivo de quem usa a informação.

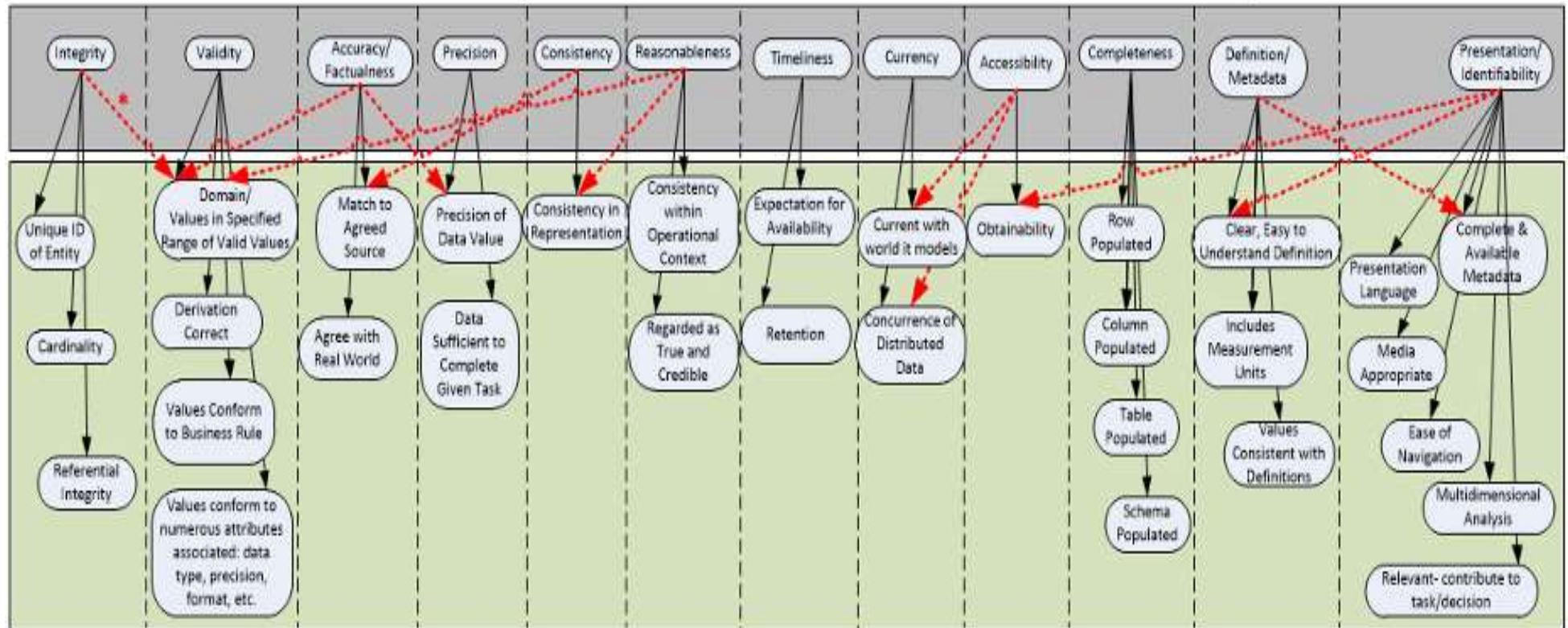
### Quadro 3 - Dimensões de qualidade em pesquisa ontológica em vários autores

<b>Dimension</b>	<b># cited</b>	<b>Dimension</b>	<b># cited</b>	<b>Dimension</b>	<b># cited</b>
Accuracy	25	Format	4	Comparability	2
Reliability	22	Interpretability	4	Conciseness	2
Timeliness	19	Content	3	Freedom from bias	2
Relevance	16	Efficiency	3	Informativeness	2
Completeness	15	Importance	3	Level of detail	2
Currency	9	Sufficiency	3	Quantitativeness	2
Consistency	8	Usableness	3	Scope	2
Flexibility	5	Usefulness	3	Understandability	2
Precision	5	Clarity	2		

Fonte: Wand & Wang, 1996, p.

Assim, há autores que chegam a propor, em outra linha metodológica, que qualidade não é algo intrínseco aos dados, mas totalmente dependente do contexto (Mohammadi et al. 2017; Cichy e Rass 2020; Härting e Lewoniewski 2020). Tais autores argumentam que um dado pode ser bom para uma organização ou uso, e ser ruim para outro, e assim definem qualidade da informação em função da adequabilidade para quem a consome – ou, de forma mais pragmática, informação de qualidade é aquela que é útil para quem dela precisa e no momento quando precisa. Nessa linha, dimensões como relevância, usabilidade, eficácia, disponibilidade e sigilo são destacadas. Myers (2013), em um esforço de compilação sobre o tema, discutiu a sobreposição de vários sentidos atribuídos a cada uma das dimensões, sintetizadas na Figura 3.

Figura 3 – Conceitos nas Dimensões de Qualidade de Dados



\*Arrows in red dashed line signify overlap between concepts referenced in more than one data quality dimension

Fonte: Myers, 2013, p.

De fato, é importante compreender o quanto as definições das dimensões são, na verdade, sobrepostas e interdependentes. Assim, por exemplo, quando falta completude e um campo não está preenchido, há também uma falta de aderência com a realidade (acurácia); quando um atributo não está semanticamente consistente com outro, pode ser um problema de desatualização (tempestividade) de um dos atributos ou registros; um valor de identificador fora dos parâmetros válidos para ele pode implicar também em falta de unicidade, além de inconsistência. A divisão em dimensões é apenas um instrumento para melhor definir e apontar falhas de qualidade.

Pelo que se vê até aqui, qualidade de dados pode ser entendida como uma construção multidimensional, e a definição de quais dimensões e níveis são necessários é impactada pelos usos dos dados. No caso de cadastros públicos sobre pessoas, importantes suportes na execução de políticas públicas, propõe-se aqui aplicar as dimensões apontadas pelo DMBok como as principais dentro da perspectiva intrínseca aos dados – completude, unicidade, tempestividade, validade, acurácia e consistência –, e aquelas dependentes de contexto e mais diretamente relacionadas ao uso dos cadastros de pessoas como suporte a políticas públicas: usabilidade, legalidade e sigilo. As dimensões selecionadas são explicadas e exemplificadas no item 6.2.1 deste trabalho.

## 5.2. Sobre bases de dados e políticas públicas

Seguindo uma linha de pensamento importante sobre qualidade de dados, Olson (2003) chama a atenção que

Não se pode separar dados de seus usos. Para avaliar a qualidade dos dados, deve-se primeiro coletar uma especificação completa dos usos pretendidos e, em seguida, julgar os dados quanto à sua adequação para esses usos. (tradução livre)

Em estudo sobre cadastros sociais, Leite et al (2017) apontam que tais cadastros são ao mesmo tempo sistemas de informação e instrumentos de operacionalização da política, porque viabilizam a identificação e seleção dos elegíveis a benefícios sociais. Nesse estudo, conduzido pelo Banco Mundial e que



envolveu casos de 20 países (e que no Brasil tomou como exemplo o Cadastro Único), foram identificados alguns tipos de consumidores dos dados:

- a) Os cidadãos a quem os dados se referem, e que tipicamente interagem no momento do cadastramento (e eventuais atualizações) para, depois, consumir os resultados da política;
- b) Os gestores, que estão focados em coletar e gerenciar os dados do cadastro para produzir as informações necessárias para a política; e
- c) Outras políticas e cadastros que vão, também, consumir (e talvez, alimentar) informações nesses cadastros – uma vez que as políticas são interligadas e existe uma ampla variedade de ações que o Estado pode oferecer a partir de um cadastro social. Esta variedade está exemplificada na Figura 4.

**Figura 4 – Possíveis consumos para um cadastro social**



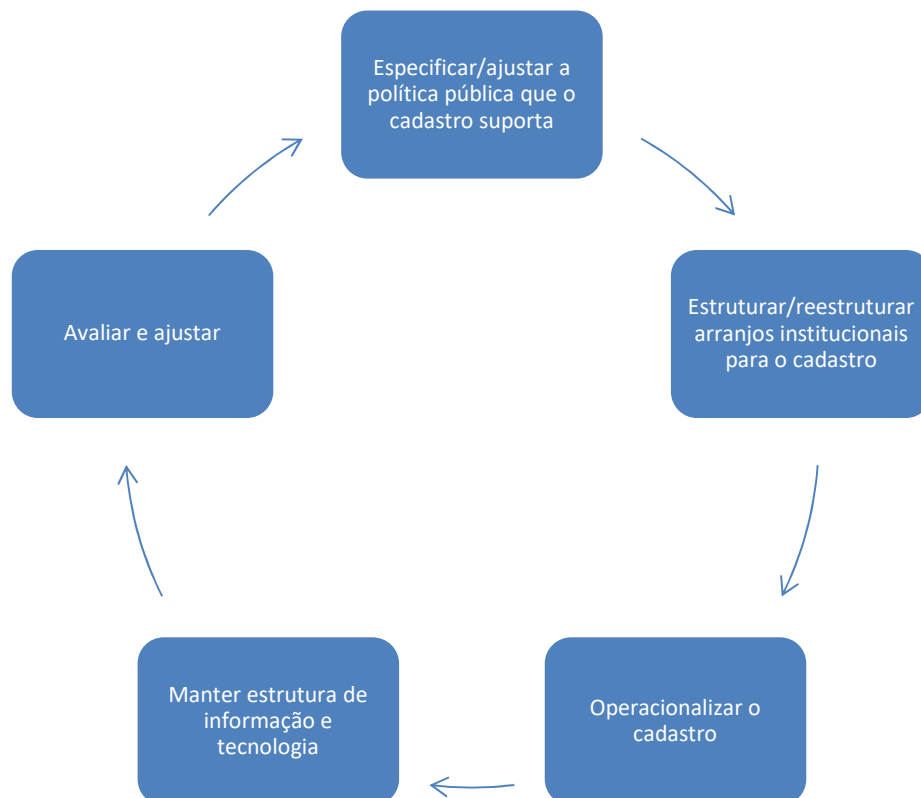
Fonte: Leite et al, 2017, p.

Olson (2003) atenta que sempre existem os consumidores para os quais são planejados os sistemas de informação (clientes e gestores), e os “usos que não eram conhecidos ou definidos no momento em que os bancos de dados foram projetados e implementados”, já que apenas no mundo ideal é possível planejar antecipadamente todos os usos para os dados armazenados. De fato, esses tipos de atores podem ser generalizados para cadastros públicos: o cadastro sobre motoristas, gerenciado pelo

Detran, também é utilizado pelo TCU para identificar, por exemplo, o endereço de pessoas; o cadastro de eleitores, gerenciado pelo TSE, tem previsão de uso do atributo de identificação biométrica para composição de cadastro nacional de identificação única. É sob o ponto de vista de todos os atores que se pode avaliar a qualidade dos cadastros e as informações obtidas a partir deles para as políticas públicas a que se destinam.

Leite et al (2017) destaca que os cadastros públicos sobre pessoas com enfoque para suporte a políticas sociais têm algumas etapas em comum, ilustradas na Figura 5. Note-se que o ciclo descrito está bastante ligado à política pública a que dará suporte, e não pode ser desvinculado dela. De fato, a própria definição dos campos que comporão a base, a forma de coleta e atualização dessas informações bem como os requisitos de qualidade para elas estão relacionados às necessidades da política.

**Figura 5 Ciclo de vida de cadastros públicos sobre pessoas**



Fonte: a autora, com base em *Structure of Social Registries* (LEITE et al, 2017).

Pode-se dizer, assim, que a relevância de possíveis problemas nos cadastros é proporcional à importância das políticas que tal cadastro suporta – e essa importância pode ser avaliada de várias formas: materialidade, abrangência, urgência, entre outros. A partir da visão de que “políticas públicas são um conceito abstrato que se materializa por meio de instrumentos variados” (SECHI, 2016), podemos considerar que os cadastros públicos de pessoas são instrumentos para políticas públicas.

Do ponto de vista de avaliação das políticas públicas, e considerando o modelo de ciclo de políticas públicas comum na literatura – apesar de todas as ressalvas sobre a não linearidade e clareza dessas fases no mundo real – a implementação é o momento em que “são produzidos os resultados concretos da política pública” e “regras, rotinas e processos sociais são convertidos de intenções em ações” (SECCHI, 2016). Nessa visão, a operacionalização dos cadastros públicos sobre pessoas ocorre principalmente na etapa de implementação da política (LEITE, 2017). E, como lembra Januzzi (2020), avaliações de implementação de políticas são especialmente relevantes para identificação de problemas críticos de ações complexas pelo porte, objetivos e contexto de operação das políticas e programas no Brasil. Nesse contexto, a avaliação da implementação de políticas suportadas por cadastros de pessoas passa pela avaliação de seus cadastros: a aderência do ciclo de vida do cadastro à política e a qualidade dos dados desse cadastro.

Além disso, vale lembrar que, conforme preconizado no “Referencial para Avaliação de Governança Multinível em Políticas Públicas Descentralizadas” e no “Referencial de Controle de Políticas Públicas” (ambas publicações do TCU), avaliações de políticas – e dos cadastros públicos que lhe dão suporte – precisam ser feitas no contexto dos próprios órgãos implementadores e, também, por órgãos de controle.

Entre exemplos dessas avaliações já feitas pelo TCU, temos:

- Fiscalização contínua de benefícios, feita pelo TCU desde 2015 sobre os benefícios assistenciais e previdenciários;
- Análise de atos de pessoal, feita pelo TCU de maneira automatizada desde 2014, que verifica a legalidade de atos de pessoal sobre os servidores públicos da esfera federal (cadastro SIAPE);
- Qualidade no cadastro único para programas sociais
- Qualidade de dados do cadastro nacional de informações sociais – CNIS
- Qualidade de dados na base do CPF/CNPJ
- Qualidade de dados nas bases do SISOB e do SIRC
- Qualidade de dados no Infoseg (Sistema de Informações de Segurança Pública)
- Qualidade de dados no cadastro de alunos do SisUAB (Sistema para Universidade Aberta do Brasil)
- Avaliação na qualidade da fiscalização da Anatel sobre a utilização dos cadastros de usuários de telefones pré-pagos.
- Qualidade no sistema de seguro pescador (cadastro seguro defeso)
- Qualidade no Cadastro Inclusão da Pessoa com Deficiência
- Qualidade no cadastro de Identidade estudantil

Por isso, neste documento, espera-se que o guia para avaliação de qualidade dos cadastros, apesar de ter como foco a avaliação feita pelo TCU, também possa apoiar gestores no processo de avaliação e aprimoramento de qualidade dos seus cadastros.

Complementarmente, para que apoiar a avaliação das dimensões da qualidade do cadastro por todos os responsáveis, é fundamental indicar as fontes que apontam o estado de cada dimensão, ou seja, as unidades de informação que sintetizam ou realçam características de um sistema, simplificando seus fenômenos: os indicadores (VAN BELLEN, 2000).

Do ponto de vista de políticas públicas, os indicadores são instrumentos que permitem identificar e medir aspectos relacionados a um determinado conceito, fenômeno, problema ou resultado de uma intervenção na realidade (SOUSA, 2013). O indicador é, então, uma mensuração/representação da realidade. Ainda de acordo

com Sousa (2013), são características importantes dos indicadores, dentre outras, em especial:

- a) Validade: capacidade de representar, com a maior proximidade possível, a realidade que se deseja medir;
- b) Confiabilidade: ter origem em fontes confiáveis e usar metodologias reconhecidas de coleta, processamento e divulgação.
- c) Simplicidade/Inteligibilidade: ser de fácil obtenção, construção, manutenção, comunicação e entendimento pelo público interno ou externo.
- d) Sensibilidade: refletir tempestivamente as mudanças decorrentes das intervenções realizadas.

No caso dos cadastros públicos de pessoas, as dimensões de qualidade podem ser consideradas os indicadores, que devem atender aos requisitos supracitados. Para eles, procurou-se no guia oferecer sugestões de fontes confiáveis e valores esperados que permitam avaliar com simplicidade e sensibilidade adequadas a qualidade buscada para o cadastro.

## 6. Desenvolvimento

### 6.1. Cadastros públicos sobre pessoas

No contexto brasileiro, há cadastros sobre pessoas mantidos por organizações públicas e implementados em milhões de registros. São vários os órgãos responsáveis pelos cadastros públicos de pessoas físicas no Brasil, com visões diferentes em função da política ou programa que suportam – e essa multiplicidade impacta a qualidade dos cadastros. Temos vários arranjos de bases cadastrais:

- a) Com gestão centralizada federal e abrangência também no país todo, como a base de contribuintes ou de eleitores. São bases grandes e complexas;
- b) Com gestão descentralizada em órgãos diversos, ainda que com abrangência nacional. É o caso dos registros gerais de identificação (RG's), expedidos por institutos de identificação (como as secretarias de segurança estaduais) que atualmente ainda usam bases sem integração entre si;
- c) Com gestão descentralizada e abrangência local, mas que acabam tendo impacto nacional, como é o exemplo das bases de servidores públicos

municipais e estaduais. Nesse caso, regras de previdência e gestão de pessoas (como a vedação à acumulação de cargos) se aplicam sobre esses servidores, sem que haja controle de integridade entre elas;

- d) Com gestão centralizada mas operação local, como é o caso do Cadastro Único, no qual a inclusão de dados se dá sob responsabilidade de entes subnacionais, que também são consumidores desses dados. Como programas sociais estaduais e municipais utilizam o Cadastro Único como base, há inclusive a possibilidade de replicação local de informações, com mais risco de desatualização de informações.

Aliado a isso, o Brasil ainda não possui um registro de identidade único para seus cidadãos. O Registro de Identidade Civil - RIC é o documento de identificação civil brasileiro, criado junto com o Cadastro Nacional de Registro de Identificação Civil pela Lei nº 9.454/1997, alterada pela Lei nº 12058/2009. Apesar da sua criação em 1997, somente em 2010 o Ministério da Justiça publicou as especificações técnicas do novo documento (por meio da Resolução nº 2/2010). Paralelo a isso, em 2017 foi promulgada a Lei nº 13.444/2017, que dispõe sobre a Identificação Civil Nacional (ICN), de responsabilidade do TSE.

Já recentemente, em 23/02/2022, o Decreto nº 10977/2022 criou uma carteira de identidade válida para todo o país. O número usado para o registro nacional será o do Cadastro de Pessoas Físicas (CPF), com mesma identidade visual para o documento em todo país. O decreto entrou em vigor em 1º de março e os institutos nacionais de identificação têm até 6 de março de 2023 para se adequarem à mudança. Até lá, continuam sendo emitidos os RG's locais normalmente.

O CPF já vem sendo usado, na prática, como principal número de referência, inclusive em iniciativas como a inclusão do número do Cadastro de Pessoa Física (CPF) nos registros e nas certidões de nascimento, casamento e óbito pelos Cartórios de Registro Civil de todo o Brasil (Provimento CNJ nº 63/2017).

Adicionalmente, esforços de integração entre os cadastros públicos já existentes têm sido promovidas, a exemplo do Decreto nº 10.046/2019, que institui o cadastro base cidadão e o comitê central de governança de dados. Esse Decreto, aliás, é atualmente objeto de ação direta de inconstitucionalidade (ADI 6649/DF) pela OAB, que alega que o Decreto extrapola limites legais do presidente e segue em oposição à Lei Geral de Proteção de Dados.

Tais ações do poder público, algumas vezes sobrepostas ou mesmo conflitantes entre si, resultam em grande dificuldade integração entre os diversos cadastros públicos, bem como na qualificação dos dados ali presentes. Além disso, a coerência dos outros atributos comuns a várias políticas – a exemplo do conceito de família e os dados de endereço, dentre outros – demanda grande esforço técnico e institucional. Frequentemente verificam-se deficiências, sobreposições e conflitos nos arranjos que suportam cadastros públicos de pessoas físicas no Brasil, e de modo especial naqueles que são instrumentos para políticas sociais.

O TCU já avaliou vários cadastros públicos sobre pessoas, como citado anteriormente neste trabalho. Em algumas dessas avaliações, a vinculação a políticas públicas importantes – como aquelas ligadas às funções assistenciais e previdenciárias, por exemplo – fez surgir a necessidade de avaliações reiteradas ao longo do tempo. Além disso, o crescente desenvolvimento da análise de dados como ferramenta para avaliação de políticas públicas e a necessidade de verificar a identidade dos cidadãos em políticas avaliadas, tornou interessante a internalização de várias bases de dados públicos – dentre as quais algumas bases sobre pessoas.

A fim de dialogar com o guia aqui proposto e exemplificar a sua aplicação, foram selecionados quatro desses cadastros, descritos a seguir. Tal aplicação da ferramenta não será de forma exaustiva para todas as dimensões de qualidade analisadas mas, como recurso didático, serão apontados exemplos mais emblemáticos de aplicação de verificação de certas dimensões, em certos cadastros, e de problemas encontrados, considerando resultados de auditorias já realizadas pelo TCU.

### **6.1.1. Contabilização de nascimento e morte dos cidadãos: o SIRC**

O Sistema Nacional de Informações de Registro Civil – Sirc, de acordo com os dados em seu site:

é uma base de governo que tem por finalidade captar, processar, arquivar e disponibilizar dados relativos a registros de nascimento, casamento, óbito e natimorto, produzidos pelos cartórios de registro civil das pessoas naturais [...] com o apoio de uma plataforma digital, em um fluxo que conecta os cartórios aos ambientes de governo eletrônico do Estado brasileiro.

[...]

Com a gestão integrada e com segurança da base formada pelas informações enviadas por cartórios, será possível qualificar outras bases de dados governamentais, subsidiar políticas públicas e ajudar a coibir fraudes na concessão de benefícios e crimes como falsificação e tráfico de pessoas. (SIRC,2015)

O SIRC é gerido por um comitê gestor, de regimento próprio, coordenado de forma alternada, em períodos anuais, pelos representantes do Ministério da Previdência Social e da Secretaria de Direitos Humanos da Presidência da República.

Os dados e eventos sobre os cidadãos registrados nessa base são bastante relevantes para várias políticas públicas, a exemplo de benefícios concedidos por filhos dependentes, pensões por morte, benefícios para a família, dentre outros. Já o Instituto Brasileiro de Geografia e Estatística (IBGE) utiliza esses dados para circularização de informações em relação às suas próprias informações do censo, que por sua vez subsidiam desde a formulação de inúmeras políticas até a distribuição do orçamento nacional a cada município (IBGE, 2022).

Em termos tecnológicos, a base, administrada de maneira centralizada pelo INSS e mantida pela Dataprev, é alimentada por serventias (cartórios) em todo o país. De acordo com dados do site oficial (<http://www.sirc.gov.br/>), o SIRC conta hoje com mais de 41,6 milhões de registros, sendo 23,2 milhões de nascimentos, 8,6 milhões de casamentos e 9,8 milhões de óbitos.

O SIRC substitui as funções do antigo Sistema de Óbitos (Sisobi), definitivamente desativado em dezembro de 2021. No entanto, os dados do antigo sistema ainda não foram totalmente migrados. Para se ter uma ideia, o Sisobi possui cerca de 26,8 milhões de óbitos em sua base. Por isso, na prática, os dois sistemas ainda são consultados pelas políticas públicas para verificação de óbito.

### **6.1.2. Cidadão-cliente do Estado Social: o Cadastro Único**

O Cadastro Único para Programas Sociais do Governo Federal (CadÚnico) foi instituído por meio do Decreto nº 3.877, de 24 de julho de 2001. Em 2003, por meio da Medida Provisória n. 132/2003, posteriormente transformada na Lei 10.836/2004, o cadastro passou a suportar o então principal programa federal de transferência de renda: o Programa Bolsa Família. Em 2007, por meio do Decreto 6.135/2007, regras de negócio do CadÚnico foram formalizadas e aprimoradas, definindo-se explicitamente como seu objetivo:

a obtenção de unicidade na identificação do público-alvo a ser cadastrado, a consistência das informações cadastrais, a integração dos programas e das políticas públicas que utilizam essa ferramenta, e a racionalização do processo de cadastramento realizados pelos diversos órgãos. (exposição de motivos do decreto)



Entre 2003 e 2005, o CadÚnico passou a ser oficialmente um “social registry” – ou cadastro de pessoas para programas assistenciais – com gestão centralizada e operação compartilhada entre a União, os estados, o Distrito Federal e os municípios. Ao aderirem ao Cadastro Único, todos os entes federados assumiram compromissos e atribuições específicos, que devem ser executados de forma articulada (Leite, 2017).

#### Quadro 4 - Programas que utilizam dados do CadÚnico

<u>Programas que usam o CadÚnico:</u>
<ul style="list-style-type: none"><li>• Auxílio Brasil</li><li>• Casa Verde Amarela</li><li>• Bolsa Verde – Programa de Apoio à Conservação Ambiental</li><li>• Programa de Erradicação do Trabalho Infantil – PETI</li><li>• Fomento – Programa de Fomento às Atividades Produtivas Rurais</li><li>• Carteira do Idoso;</li><li>• Aposentadoria para pessoa de baixa renda;</li><li>• Programa Brasil Carinhoso;</li><li>• Programa de Cisternas;</li><li>• Telefone Popular;</li><li>• Carta Social;</li><li>• Pro Jovem Adolescente;</li><li>• Tarifa Social de Energia Elétrica;</li><li>• Passe Livre para pessoas com deficiência;</li><li>• Isenção de Taxas em Concursos Públicos.</li></ul>
Os Estados e municípios também utilizam os dados do Cadastro Único como base para seus programas sociais.

Fonte: Caixa Econômica Federal sobre o Cadastro Único. Disponível em <https://www.caixa.gov.br/servicos/cadastro-unico/paginas/default.aspx>. Acessada em 22/05/2022

Conforme atenta Direito & Koga (2020):

Em função do seu elevado grau de cobertura da população pobre do país, pela ampla variedade de informações sobre suas condições de vida e por contar com nome e endereço dessa população, o CadÚnico figura certamente entre as mais importantes fontes de informação sobre a população pobre. As informações sobre renda podem ser utilizadas de variadas formas, além de servir diretamente para a seleção dos beneficiários do PBF.

O Cadastro Único enquadra-se no que Olsen (2003) chama de “ampliação de usos” para um cadastro. De fato, além dos dados de identificação, o questionário que os agentes levantam junto ao candidato contém dados sobre características do domicílio, composição da família, características de populações específicas (indígenas e quilombolas), dados de pessoas com deficiência na família, escolaridade, trabalho e remuneração. Além disso, há formulários para coleta específica de dados para população em situação de rua (Ministério da Cidadania, 2019).

Direito e Koga (2020) argumentam que, se por um lado, essas possibilidades de uso amplo de tais informações tornam mais crítica a manutenção da qualidade do cadastro, inclusive em campos não obrigatórios e não essenciais inicialmente para a elegibilidade de programas, é possível utilizar a própria informação do cadastro para identificar possíveis inconsistências e, com isso, aprimorar sua qualidade e identificar distorções.

O Cadastro Único atualmente contempla mais de 25 milhões de famílias em seus registros.

### **6.1.3. Eleitores: a base do TSE**

Conforme atenta PRATA (1998), “eleitor é talvez o aspecto mais evidente do cidadão numa democracia representativa”. De fato, no Brasil, nossa Constituição estabelece, em seu Art. 14., que a “soberania popular será exercida pelo sufrágio universal e pelo voto direto e secreto, com valor igual para todos”, estabelecendo o voto como obrigatório a todo brasileiro alfabetizado entre 18 e 70 anos.

A Constituição também institui a Justiça Eleitoral e, complementarmente, a Lei nº 4.737, de 15 de julho de 1965, estabelece o código eleitoral, incluindo aí o alistamento e registro de eleitores. Já em 1985, a Lei nº 7.444/1985 estabeleceu a implantação do processamento eletrônico de dados no alistamento eleitoral, quando se procedeu à revisão dos eleitores inscritos, bem como à conferência e à atualização dos respectivos registros, que passaram a ser “cadastros mantidos em computador”. Além disso, a mesma lei de 1985 estabeleceu que

Art. 4º - Para a conferência e atualização dos registros eleitorais a que se refere o art. 2º desta Lei, a Justiça Eleitoral poderá utilizar, também, informações pertinentes, constantes de cadastros de qualquer natureza, mantidos por órgãos federais, estaduais ou municipais.

Complementarmente, a operacionalização da gestão do Cadastro Eleitoral está definida na Resolução TSE nº 23.659, de 26 de outubro de 2021. É o TSE quem define procedimentos e normas complementares para o cadastro de eleitores, que em 2018 contava com 147.306.294 registros eleitorais – número esse que deverá ser atualizado com os cerca de 500 mil novos títulos de eleitores regularizados até maio de 2022, segunda informa o site do TSE.

Considerando a importância das eleições para o Estado democrático de direito, é fácil perceber a criticidade da qualidade do cadastro de eleitores do Brasil. Não bastasse a importância inerente, no caso brasileiro esse cadastro se reveste de importância adicional, por possuir uma ampla base de impressões digitais em formato eletrônico, constituindo relevante base biométrica de identificação única.

#### **6.1.4. Contribuinte: a base do CPF**

Em sua análise dos diversos papéis do cidadão frente ao Estado, PRATA (1998) lembra que contribuinte é a perspectiva “mais antiga do cidadão: muito antes de poder votar, ele já era obrigado a pagar tributos”.

O Cadastro de Pessoas Físicas (CPF) é mantido pela Secretaria da Receita Federal e regido pelo Decreto-Lei nº 401/1968 e a Instrução Normativa nº 1548/2015. Sua destinação é o controle dos contribuintes para o imposto federal.

Inicialmente, o CPF possuía cartão físico próprio. O contribuinte que apresentou declaração do exercício de 1969, ano-calendário de 1968 recebeu, no início de 1970, juntamente com o Manual de Orientação e formulários, duas vias do Cartão de Identificação do Contribuinte (CIC), emitidos eletronicamente e com prazo de validade. A declaração de rendimentos do imposto de renda das pessoas físicas passou a ter um campo para indicação do número de inscrição no Cadastro das Pessoas Físicas. (SRFB, 2022)

Entre 1980 e 90 foi feito grande investimento na informatização dos processos da Secretaria da Receita e, em 1996, já havia programa para declaração do imposto disponível no sistema operacional Windows (SRFB,2022). Do lado das bases de dados, foi feito grande esforço também para a melhoria de qualidade. Na prática, hoje, o CPF vem sendo usado como o registro de identificação dos cidadãos – embora inicialmente não tenha sido concebido para esse fim, e isso gera alguns impactos.

Os campos exatos de compõem a base do CPF não são de domínio público, mas um subconjunto deles é compartilhado entre órgãos da administração pública, com a função de identificação, tendo por fundamento o Decreto nº 10.046/2019. A Instrução Normativa nº 1548/2015 descreve campos de identificação da base do CPF, bem como procedimentos típicos de controle de pagamento de imposto, como a indicação de pendência de regularização.

A base do CPF possui atualmente quase 240 milhões de registros ativos. O número excede a população estimada no Brasil, e se justifica pelo fato de incluir CPF's em situações de óbito cujo espólio ainda não foi concluído.

## **6.2. Qualidade de cadastros públicos de pessoas**

### **6.2.1. Dimensões de qualidade dos dados**

Uma dimensão da qualidade de dados é um recurso ou característica mensurável dos dados que funciona como suporte para definir os requisitos de qualidade para esses dados. Podem ser entendidas como indicadores para traduzir os resultados da avaliação da qualidade dos dados, e sua variação ao longo do tempo pode apontar melhorias aplicadas à qualidade (WAND & WANG, 1996). A capacidade de medir e rastrear essa qualidade é fundamental para gerenciar e melhorar a produção e distribuição de informações e para a criação do conhecimento (GUEDES et al, 2021).

As dimensões da qualidade de dados são fortemente dependentes do contexto de uso desses dados (MYERS, 2013) e, no contexto desse estudo, devem refletir os requisitos mais importantes para cadastros sobre pessoas usados como instrumentos para implementação de políticas públicas. Nos tópicos seguintes, estão listadas dimensões de qualidade relevantes sob essa perspectiva.

#### **6.2.1.1. Completude**

É possível atribuir à completude, no contexto de qualidade de bases de cadastros públicos, dois sentidos principais.

O primeiro sentido, associado a bases de dados em geral, diz respeito ao quão completos estão os dados de acordo com as definições daquela base. Essa visão,

bastante comum nos estudos de governança de dados no contexto da tecnologia da informação, procura identificar se todos os atributos obrigatórios estão preenchidos e se os não obrigatórios preenchidos são suficientes para a execução do processo de negócio associado, considerando uma referência que explicita tais requisitos (DAMA, 2009). Essa referência pode ser um normativo que explicita requisitos de completude para o conjunto de dados, mas o controle fica facilitado quando também está registrada num formato próximo da implementação do banco de dados, a fim de evitar falhas e ambiguidades. O dicionário de dados é um metadado que contém várias informações importantes sobre os dados da base, inclusive os requisitos de completude dos campos (HEUSER, 1998).

A completude evidencia estreita relação entre o cadastro e os sistemas de informação que operam sobre ele (OLSON, 2003), os quais devem estar aderentes à política pública que apoiam. Uma vez que os campos sejam definidos como obrigatórios, por exemplo, sua ausência pode resultar em comportamentos inesperados das aplicações que os manipulam, e regras de negócio podem deixar de ser cumpridas. Por outro lado, a falta de completude também pode indicar falha nas aplicações e nos controles que alimentam os dados da base.

Uma vez que exista um dicionário de dados e que os campos que são obrigatórios sejam claramente definidos, a avaliação de completude pode ser feita confrontando essas definições e o efetivo preenchimento dos registros da base.

Adicionalmente, Olson (2003) chama atenção sobre a importância de especificar a distinção, para um campo, entre “nenhum valor” e “eu não sei o valor correto”. De acordo com o autor, a ausência dessa distinção na descrição e nos procedimentos de inserção de dados na base é uma fonte comum de falha de qualidade nos dados – e bastante difícil de detectar.

O Cadastro Único, por exemplo, de acordo com o Manual do Entrevistador (MCID, 2022), apresenta campos obrigatórios que visam identificar, localizar e definir perfil para a família candidata a benefícios assistenciais, como o Bolsa Família (atualmente substituído pelo Auxílio Brasil), Auxílio Gás ou Tarifa Social de Energia Elétrica, além de isenção de IPTU concedida por estados e municípios. Esses dados são coletados pelo entrevistador e, além disso, a família é acompanhada no local em que foi feito o cadastro. Assim, dados da entrevista e do entrevistador também são mantidos, para operacionalização de atividades da política.

Em auditorias continuadas executadas pelo TCU, nas quais se teve acesso à base de dados, observou-se que vários desses campos não estavam preenchidos. Como exemplo, trazemos aqui um extrato relativo à completude de campos, de acordo com dicionário de dados fornecido pela Secretaria Nacional de Renda de Cidadania ao órgão de controle à época.

**Tabela 1 – Incompletudes no Cadastro Único**

Nome do campo	Quantidade de ocorrências	% em relação ao total de registros
DTA_ENTREVISTA_FAM	109	0,00
NOM_LOCALIDADE_FAM	506.186	1,87
NOM_TIP_LOGRADOURO_FAM	506.196	1,87
NOM_LOGRADOURO_FAM	506.186	1,87
NUM_CEP_LOGRADOURO_FAM	506.187	1,87
NOM_ENTREVISTADOR_FAM	400.847	1,48
NUM_CPF_ENTREVISTADOR_FAM	400.847	1,48
COD_LOCAL_DOMIC_FAM	96	0,00
COD_ESPECIE_DOMIC_FAM	305.348	1,13
COD_PARENTESCO_RF_PESSOA	295.325	0,37
COD_LOCAL_NASCIMENTO_PESSOA	210.147	0,26
COD_CERTIDAO_REGISTRADA_PESSOA	9.709	0,01

Fonte: a autora, com base nos dados do Relatório FCB 2016-2017

A falta de informações como endereço poderia dificultar ou impossibilitar tanto visitas de averiguação quanto o mapeamento de necessidades da família. A falta da informação de parentesco com o responsável familiar, de forma similar, complica ainda mais o mapeamento da família, e dá espaço para fraudes nessa composição familiar. E o controle fica dificultado se os dados do entrevistador não são informados. O conjunto dessas incompletudes abre espaço para falhas e fraudes.

Outro exemplo, ainda mais emblemático pelas suas consequências é o caso do Sisobi e do SIRC. Para a identificação dos alvos de várias políticas públicas pode ser conveniente a utilização do número do Cadastro de Pessoas Físicas da Receita Federal do Brasil (CPF). Entretanto, tal obrigação é recente nos registros de

nascimento e óbito (2015) e, como se observa na Tabela 2, abaixo, tanto no SISOBI quanto no SIRC há grande quantidade de registros nulos.

**Tabela 2 - Registros no SISOBI sem campo de identificação de CPF**

Base de Dados	Registros (milhões)	Registros nulos (milhões)	% de nulos
SISOBI	28,7	11,0	38%
SIRC- Certidões de óbito	10,4	6,0	58%
SIRC - Certidões de óbito oriundas do Sisobi	19,0	9,3	49%

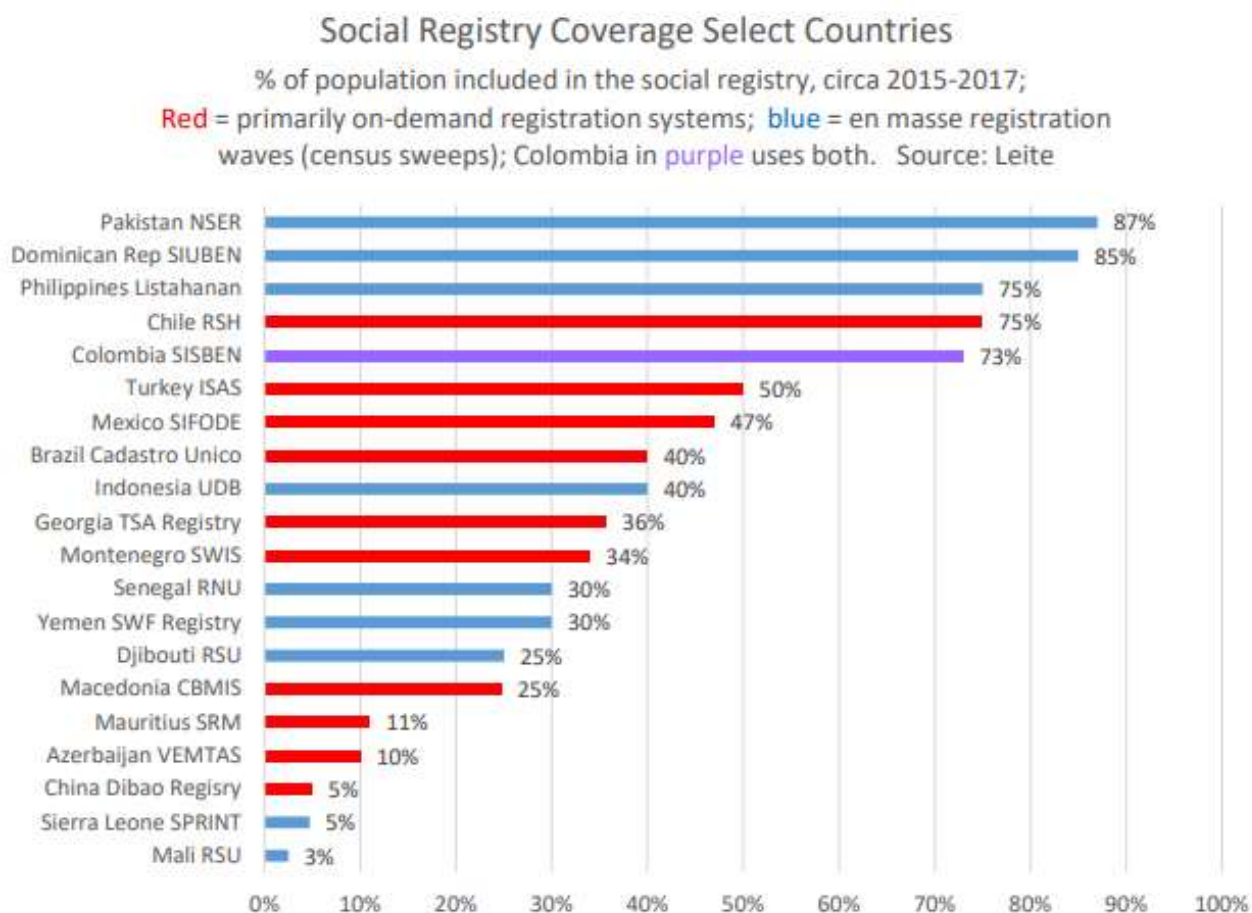
Fonte: dados do LabContas – laboratório de bases de dados do TCU (2022)

O segundo sentido associado à completude, especialmente voltado para o caso de cadastros de pessoas, refere-se à proporção de registros armazenados em relação ao total do universo definido para a base. Em termos práticos, busca dimensionar se o público-alvo do cadastro está ali representado em sua totalidade (LEITE, 2017).

Em cadastros públicos sobre pessoas, a completude é fundamental para garantir que o público-alvo das políticas públicas está representado em quantidade suficiente para a meta definida pela política pública associada e que a ação do Estado se estende a todos aqueles que, de fato, se propõe a alcançar. Este tipo de avaliação é um pouco mais complexo, mas pode ser feita confrontando o total de registros com outras bases de referência e/ou com tratamento estatístico sobre os dados.

Um exemplo são os registros civis de nascimento e óbito, cujo direito é previsto na Constituição como fundamental e gratuito. Assim, no caso do SIRC, que contém tais registros, o IBGE realiza sistematicamente um pareamento entre essa base e bases do Ministério da Saúde (Sistema de Informações sobre Nascidos Vivos - SINASC e Sistema de Informações sobre Mortalidade - SIM) para obter estimativas da incompletude do SIRC – o que representa subnotificação de nascimentos e óbitos. Tal ação visa apoiar a “melhora da enumeração dos eventos vitais, o aprimoramento dos sistemas de informações e a erradicação do sub-registro” (IBGE, 2022).

Leite et al (201x), destaca a importância da completude em cadastros públicos sobre pessoas e compara essa situação em vários dos países do estudo.

**Figura 6 – Completude de público-alvo em cadastros sociais (2017-2017)**

Fonte: Leite, 2019, p.

### 6.2.1.2. Consistência

Consistência é a ausência de diferença ao comparar duas ou mais representações de um objeto, frente a uma definição (DAMA, 2003). É, assim, um indicador do quanto os parâmetros e valores de uma base de dados são coerentes entre si e não se contradizem (GUEDES et al, 2021). Em se tratando de cadastros de pessoas, referem-se à coerência de campos repetidos e à integridade cruzada entre atributos (HEUSER, 1998). Nota-se que dependem, essencialmente, da semântica dos dados – descrita no dicionário de dados.

Para avaliar a consistência, é preciso confrontar diferentes campos do mesmo banco de dados, cujos resultados devem ser coerentes. Isto porque, por uma série de motivos, cadastros contêm informações total ou parcialmente redundantes. Um exemplo comum em cadastros sobre pessoas são os campos de endereço, que incluem CEP e estado/cidade: como só há um CEP por localidade, não seria



necessário o segundo campo mas, se ele existir (e isso é comum), deve ser coerente com o primeiro.

Em outros casos, a semântica ou tipo de preenchimento de um campo depende de outro, relacionado. Por exemplo, se há a informação de “pai desconhecido”, não deve haver preenchimento do nome do pai em outro campo; ou, se o trabalhador informa que perdeu o emprego, a data de encerramento do contrato empregatício tem que estar preenchida. Nesses casos, mais uma vez, há forte dependência de um metadado que contenha informações de semântica suficientes para explicitar essas relações – e tornar possível a verificação de eventuais inconsistências (OLSON, 2003). A Figura 7 ilustra possíveis campos que possuem consistência: algumas questões só devem ser preenchidas se o domicílio é permanente, ou se existem ao menos um banheiro; além disso, o número de quartos não pode ser maior do que o número de cômodos total da casa. Embora esses itens não impeçam o cadastramento de um candidato a beneficiário, imprecisões podem sugerir falhas no processo de coleta de dados e comprometem a confiabilidade das informações.

**Figura 7 - Trecho de formulário para preenchimento do CadÚnico, com indicação de consistência de alguns campos**

Formulário Principal de Cadastro - CADÚNICO - F1

2.01 - O local onde está situado o seu domicílio tem, na maioria, características:

1 - Urbanas     2 - Rurais

2.02 - Qual é a espécie do seu domicílio?

1 - Particular permanente  
 2 - Particular improvisado  
 3 - Coletivo

2.03 - Quantos cômodos tem seu domicílio?

2.04 - Quantos cômodos estão servindo, permanentemente, de dormitório para os moradores do seu domicílio?

2.05 - Qual é o material predominante no piso do seu domicílio?

1 - Terra  
 2 - Cimento  
 3 - Madeira aproveitada  
 4 - Madeira aparelhada  
 5 - Cerâmica, lajota ou pedra  
 6 - Carpete  
 7 - Outro material

2.06 - Qual é o material predominante na construção das paredes externas do seu domicílio?

1 - Alvenaria/tijolo com revestimento  
 2 - Alvenaria/tijolo sem revestimento  
 3 - Madeira aparelhada  
 4 - Taipa revestida  
 5 - Taipa não-revestida  
 6 - Madeira aproveitada

2.08 - Qual é a forma de abastecimento de água utilizada no seu domicílio?

1 - Rede geral de distribuição  
 2 - Poço ou nascente  
 3 - Cisterna  
 4 - Outra forma

2.09 - No seu domicílio ou na propriedade existe banheiro ou sanitário?

1 - Sim     2 - Não → Passe ao 2.11

2.10 - De que forma é feito o escoamento do banheiro ou sanitário?

1 - Rede coletora de esgoto ou pluvial  
 2 - Fossa séptica  
 3 - Fossa rudimentar  
 4 - Vala a céu aberto  
 5 - Direto para um rio, lago ou mar  
 6 - Outra forma

2.11 - O lixo do seu domicílio:

1 - É coletado diretamente  
 2 - É coletado indiretamente  
 3 - É queimado ou enterrado na propriedade  
 4 - É jogado em terreno baldio ou logradouro (rua, avenida, etc.)  
 5 - É jogado em rio, lago ou mar  
 6 - Tem outro destino

2.12 - Qual é a forma de iluminação utilizada no seu domicílio?

1 - Elétrica com medidor próprio  
 2 - Elétrica com medidor comunitário  
 3 - Elétrica sem medidor  
 4 - Óleo, querosene ou gás

Handwritten annotations in blue ink: "Passe ao 3.01" with an arrow pointing to question 2.02; "Passe ao 2.11" with an arrow pointing to question 2.09; and a bracket connecting question 2.11 to question 2.12.

Fonte: Manual do Entrevistador do Cadastro Único, 2018.

Adicionalmente, quando o Estado faz uso dos dados de seus cadastros para executar suas políticas, ou mesmo para aprimorar a formulação, a implementação, a avaliação e o monitoramento dessas políticas públicas, realizar as operações em bases de dados de maneira isolada não é suficiente (TCU,2018). Muitas vezes é necessário que uma organização pública tenha acesso a dados que estão sob gestão de outras instituições. Nesse momento, torna-se crítico que uma informação de uma base esteja coerente também com a de outra base. A inconsistência, nesse caso, implica que alguma (ou ambas) as bases estão com problemas de acurácia – que veremos no item 6.2.1.6.

Em auditorias no TCU, é frequente o cruzamento de bases para verificação de consistência de dados entre elas, em função do impacto que essa diferença pode ter para as políticas públicas. Uma verificação comum de consistência é a indicação de

óbito. É do procedimento de várias políticas públicas a verificação do óbito para suspensão de pagamento de benefícios. No caso do auxílio emergencial, por exemplo: o beneficiário não poderia estar falecido e ser contemplado com o auxílio. A equipe de auditoria do TCU que acompanhou o auxílio então, cruzou informações, tendo o campo CPF como identificador de pessoa, entre os beneficiários selecionados para receber o auxílio e bases que continham informação de falecimento de pessoas. A partir daí, foram identificados um total de 17.084 beneficiários distintos que, em uma ou mais das bases, estavam apontados como falecidos. Além disso, a equipe verificou que nem todas as bases testadas continham informação coerente entre si para o mesmo beneficiário – isto é, um beneficiário poderia ser apontado como falecido em uma das bases e não em outra.

**Tabela 3 - Exemplo de inconsistência entre bases que indicam falecimento**

<b>Base onde foi verificado óbito</b>	<b>Número de beneficiários do auxílio emergencial localizados na base</b>
Base de Pessoas Físicas (CPF) – seleção daquelas com indicação de falecimento	9.590
Sistema Nacional de óbitos (Sisobi)	840
Sistema de Registro Civil (Sirc)	534
Base de beneficiários do INSS – seleção dos instituidores de pensão por morte	6.987
Base de Servidores Públicos (SIAPE) – seleção dos instituidores de pensão por morte	2
<b>Total de beneficiários distintos do auxílio, com indício de falecimento</b>	<b>17.084</b>

Fonte: dados da tabela “Beneficiários titulares do auxílio emergencial com indícios de falecimento” in “1o. Relatório de Acompanhamento do Auxílio Emergencial”, 2020

A base do Sisobi, aliás, tem tido sua consistência questionada já há algum tempo pelo TCU, tendo sido alvo de fiscalizações (TCU, 2017). Tal base teve suas funções substituídas pelo SIRC em dezembro de 2021. No entanto, os registros anteriores a esta data ainda precisam ser obtidos a partir dela e, atualmente, o TCU ainda avalia a migração de dados entre essas bases, para que possa vir a ser utilizada somente uma delas. Como parte dessas verificações, por exemplo, observou-se que dos 28,7 milhões de óbitos registrados na base Sisobi, mais de 4,4 milhões de registros (15%), possuem número de CPF nulo ou não encontrado na base de dados do Cadastro de Pessoa Física da Receita Federal.

Já dentre os números de CPF presentes no Sisobi e encontrados no Cadastro da Receita Federal (24,3 milhões), 2,5 milhões (10%) possuem o nome diferente ou

a data de nascimento diferente, sendo que cerca de 200 mil possuem as duas informações diferentes.

**Tabela 4 - Inconsistência de nome e data de nascimento, em bases diferentes**

<b>Mesmo nº de CPF no SISOBÍ e na Receita Federal</b>			
<b>Igualdade dos nomes</b>	<b>Total</b>	<b>Com data de nascimento iguais</b>	<b>Com data de nascimento diferentes</b>
Nomes iguais	22.606.555	21.835.726	652.433
Nomes diferentes	1.679.979	1.465.502	197.368
<b>TOTAL</b>	<b>24.286.534</b>	<b>23.301.228</b>	<b>849.801</b>

Fonte: a autora, com base nos dados do LabContas – **Laboratório** de dados do TCU, em 01/06/2022.

### 6.2.1.3. Unicidade

A unicidade é um tipo especial de consistência que visa a ocorrência de apenas um registro na base de dados para cada objeto do mundo real, de maneira não ambígua e precisamente identificada, de tal forma que o objeto possa ser distinguido de todos os outros do mesmo tipo (MAYDANCHIK, 2007). Em termos de modelagem de dados do mundo real para registros no cadastro, a dimensão da unicidade visa garantir uma correspondência biunívoca entre cada instância de uma entidade modelada e sua representação na base de dados (HEUSER, 1998).

Em geral, isso é implementado em bases de dados por meio de uma identificação única para cada registro (chave) e não deve haver mais de um registro com a mesma chave na base (DAMA, 2009). Em particular, quando falamos de cadastros públicos de pessoas, referimo-nos às formas de identificação única de uma pessoa e à não ocorrência de multiplicidade dessas identificações na base do cadastro (LEITE, 2017).

Cadastros armazenados em Sistemas Gerenciadores de Banco de Dados (SGBD) têm suporte à definição de chaves de maneira nativa na maioria das ferramentas de mercado. Uma vez definido um identificador como chave, a própria ferramenta limita a ocorrência de mais de um registro com o mesmo valor de chave (DATE, 2004). Nesse sentido, a garantia de unicidade passa a se referir à própria

escolha de qual campo do registro será a chave primária que identificará o registro. Conforme atenta HEUSER (1998), mais de um campo ou combinações de campos podem servir para distinguir um registro de outro; nesse caso, um deles, ou uma combinação mínima, é escolhida como chave primária, enquanto os outros são considerados “chaves alternativas”. As chaves podem ser mais “intuitivas” ou “naturais” (em geral, quando o valor é conhecido no “mundo real” fora do sistema, por exemplo, número de telefone ou nome) ou “artificiais” (por exemplo, um sequencial gerado por um programa para cada novo registro no cadastro).

No caso de cadastros públicos sobre pessoas, a escolha de qual campo será chave, e se ele será mais “natural” ou não, se relaciona com o problema da identificação do cidadão frente ao Estado. E este não é um problema simples (OLIVEIRA NETO, 2019; MENEZES, 2019).

De fato, no Brasil existem vários identificadores para os cidadãos: título de eleitor, CPF, NIS, NIT, carteira de motorista, RG... O problema da unicidade se desdobra, então, em:

- a) identificar que, para aquele cadastro, há uma definição clara de qual (ou quais) campo(s) identifica(m) unicamente uma pessoa, e que só há uma ocorrência desse campo na base do cadastro para cada indivíduo;
- b) identificar que outros cadastros que reconhecem a mesma chave ou outra chave candidata para o mesmo indivíduo estão consistentes entre si.
- c) identificar que aquele registro unicamente identificado diz respeito a uma única pessoa real, e vice-versa – ponto em que o problema tangencia a acurácia.

Na prática, a unicidade é um dos principais problemas nas bases de dados de cadastros públicos sobre pessoas. A falta de unicidade gera impactos sobre a política pública associada ao cadastro, dá margem a falhas e fraudes, e dificulta a interrelação entre políticas.

Para avaliar o nível de unicidade de uma base cadastral é necessário confrontar a definição da chave primária e verificar a ocorrência única dessa chave na base. Caso haja na mesma base outras chaves alternativas, vale verificar se elas também são únicas. Também podem ser aplicadas técnicas de análise, em geral com o apoio de ferramentas automatizadas, para verificar se existem registros que “parecem” ser a mesma pessoa (por exemplo, testando um conjunto de dados potencialmente identificadores, como nome, nome da mãe, data de nascimento, mas apenas a chave

diferente). Algumas vezes, tais ferramentas permitem “níveis de certeza” e comparam dados “quase iguais” (por exemplo, nomes abreviados). O resultado dessas análises pode apontar registros ou identificações duplicadas e, dependendo do contexto, indicam falha ou fraude no cadastro.

Além disso, partindo do princípio que mesmo em diferentes cadastros, a pessoa (cidadão) que se quer representar é o mesmo, é possível fazer validações entre cadastros, para melhorar a precisão das análises e identificação de duplicidade. Esse tipo de validação envolve técnicas e ferramentas para identificar, em caso de inconsistência, qual registro ou qual informação deve ser considerada como identificadora da pessoa.

Enfim, vale dizer que o problema da unicidade em cadastros está diretamente relacionado à identificação única do cidadão frente ao Estado. No Brasil, várias têm sido as tentativas de estabelecer uma identidade única, o que apontado no item **Erro! Fonte de referência não encontrada.** deste documento. Uma vez consolidada uma identificação nacional única, a ocorrência dessa identificação no cadastro passa a ser um fator de qualidade de unicidade. Por ora, entretanto, temos que conviver com uma multiplicidade de identificações.

Por exemplo, em auditoria realizada pelo TCU em 2019 foi possível verificar que dados do SIRC apresentavam inconsistência de identificação, resultando em risco de fraude não só para o próprio cadastro, mas tendo consequências para outros cadastros, já que a certidão de nascimento pode ser utilizada para obter outros documentos.

Campo chave	Quantidade de ocorrência múltipla	% em relação ao tamanho da base
Matrícula da certidão	0	0.0
Cadastro de pessoa física (CPF)	5.472	0.0
Declaração de nascido vivo (DNV)	141.867	1.1
Nome, data de nascimento e nomes das filiações	16.912	0.1

Fonte: adaptado de “Tabela 1: Resultado da verificação de unicidade e completude de certidões de nascimento”, em (TCU, 2019).

Já no caso do Sisobi, a chave declarada da certidão de óbito é a "Identificação do Cartório" - CNPJ, CEI etc, dependendo do tipo -, o "Número do Livro", o "Número da Folha" e o "Número do Termo". Os dados analisados para a presente amostra foram informados 219 vezes (a extração é de 09/2003 a 01/2022, exceto 02 e 03/2012). Ainda que se considere que todas as chaves fossem informadas a cada mês, ou seja, desprezando-se 219 repetições do que seria a identificação única (uma para cada ano e mês informado), verifica-se que existem várias repetições dessa mesma identificação, que deveria ser única, para pelo menos 21.514 registros, distribuídos conforme a Tabela 5:

**Tabela 5 - nível de repetição da mesma chave e suas partes, no Sisobi**

IDENTIFICAÇÃO QUE DEVERIA SER ÚNICA					REPETIÇÕES (pelo menos)
Tipo Ident. Cartório	Identificação Cartório	Número do Livro Registro Óbito	Folha Registro Óbito	Número Termo Óbito	
1	7735814000113	1	1	0	3.687
0	2172200126	1	1	0	1.474
3	38646331415	1	1	0	960
0	2173000014	1	1	0	938
0	2173000073	1	1	0	913
1	83549493000140	1	1	0	739
0	2070100366	1	1	0	732
Entre 600 e 699 repetições da mesma identificação					2.648
Entre 500 e 599 repetições da mesma identificação					500
Entre 400 e 499 repetições da mesma identificação					1.281
Entre 300 e 399 repetições da mesma identificação					1.785
Entre 200 e 299 repetições da mesma identificação					2.519
Entre 100 e 199 repetições da mesma identificação					1.742
Entre 1 e 99 repetições da mesma identificação					1.596
<b>TOTAL</b>					<b>21.514</b>

Fonte: dados obtidos do LabContas – Laboratório de bases de dados do TCU, obtido em 01/06/2022.

#### 6.2.1.4. Tempestividade

OLSON (2003) reporta que dados são “vivos” e é muito comum que uma mudança aconteça num objeto representado numa base de dados e a base de dados não reflita essa atualização. A dimensão temporal da qualidade visa que os dados da base representem o estado mais atualizado dentro dos requisitos definidos para as informações (DAMA, 2015). No que diz respeito a bases cadastrais de pessoas, é necessário que os dados sejam coletados, processados e atualizados no intervalo de tempo necessário para que a informação mais atual, dentro do propósito da base, esteja acessível quando for requisitada (LEITE, 2017).

De fato, as informações sobre pessoas em cadastros públicos frequentemente referem-se a estados ou situações que mudam ao longo do tempo: pessoas mudam de endereço, de emprego, aposentam-se, casam-se e separam-se, nascem e morrem – uma grande variedade de eventos que podem ser registrados nos cadastros. Assim, é fundamental identificar quais os atributos representam essas características mutáveis, e determinar o grau de sincronização com a realidade que se pretende para os dados da base. Essa necessidade está, em geral, descrita exteriormente à base, em documentos de negócio, manuais e modelos, e é fundamental que seja localizada para que se possa analisar esta dimensão da qualidade dos dados. LEITE (2017) pondera que a maneira e a periodicidade pela qual são coletados e atualizados os dados em cadastros sociais afeta diretamente a atualidade desses dados no cadastro.

Uma vez identificado o requisito, é o momento de escolher a forma pela qual se poderá avaliá-lo na base. Em geral, é boa prática o uso de controles (flags) de atualização para campos relevantes (HEUSER, 1998). Também é possível utilizar controles de atualização do próprio sistema ou da aplicação que suporta a base de dados. A verificação de atualidade deve ser feita em função da maneira que está registrada. O não registro já é um problema, pois não se tem como averiguar a qualidade nesse quesito (DAMA, 2015).

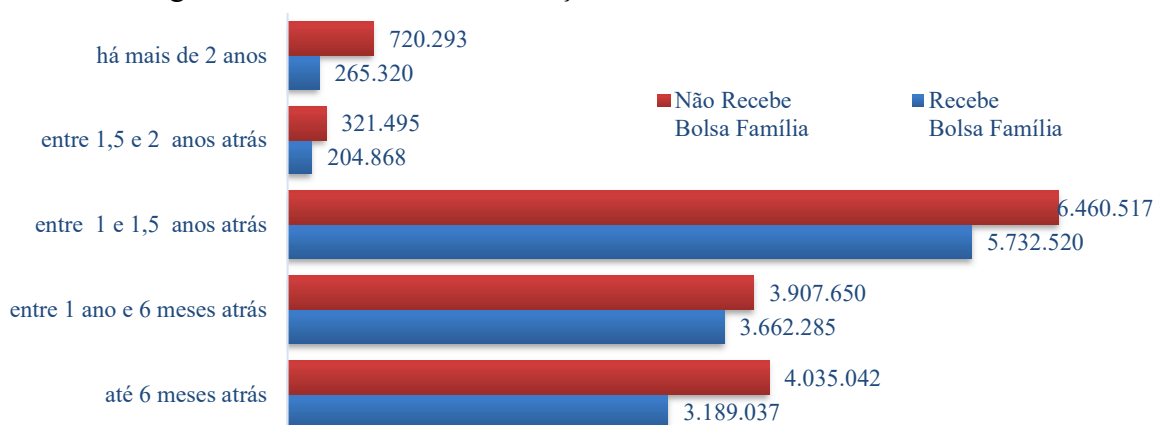
Assim, para uma base de dados como o Cadastro Único, que suporta políticas assistenciais voltadas para renda e composição familiar, a desatualização dos dados de renda, número de filhos ou separação pode impactar a elegibilidade da família. Nesse exemplo em particular, o recente Decreto nº 11.016, de 29 de março de 2022 define que os dados cadastrais devem ser atualizados no mínimo a cada dois anos mas, na prática, pode haver situações em que haja impacto direto em caso de desatualização bem menor. Foi o que aconteceu, por exemplo, durante a vigência do auxílio emergencial em função da pandemia de covid- 19. Como o cadastro foi



utilizado como base para elegibilidade, famílias que tiveram redução de renda ou que aumentaram seus membros poderiam ter recebido valor superior, mas não o fizeram pela desatualização. Tal *gap* temporal ficou ainda mais aumentado pelas limitações impostas pela própria pandemia.

O TCU acompanhou, por meio da ação “Coopera”, várias ações relacionadas ao enfrentamento da covid-19. No primeiro relatório de acompanhamento do auxílio emergencial, foram identificados riscos de exclusão ou inclusão indevida, relacionados à atualização do cadastro, conforme resumido na Figura 8. O gestor precisou adotar medidas para mitigar esses riscos, realizando verificações de renda e emprego por meio de cruzamento de dados intenso em outras bases da administração.

Figura 8 – Famílias sem atualização no Cadastro Único



Fonte: relatório TCU-2020, referente ao CadÚnico 03/2020

O grau de atualidade do cadastro depende bastante dos processos que lhe dão suporte, e isso pode ser investigado para apontar pistas das causas de eventuais desatualizações. A regularidade de inclusão de novas pessoas, bem como da verificação de mudanças (mecanismos como prova de vida ou eventos de recadastramento) e os meios de acesso para atualizações são fatores que podem impactar a qualidade temporal do cadastro.

#### 6.2.1.5. Validade

Um dado é válido se está conforme a sintaxe da sua definição, incluindo aqui formato, tipo, precisão numérica e domínio de valores. Um domínio de valores pode

ser um conjunto definido de valores válidos (como em uma tabela de referência), um intervalo de valores ou um valor que pode ser determinado por meio de regras (DAMA, 2003). Tal definição costuma estar em dicionário de dados ou documento similar.

Dados inválidos podem gerar comportamentos inesperados nos sistemas de informação que utilizam esses dados ou nas conclusões de análises de dados sobre o comportamento dos dados. (OLSEN, 2003).

A validade de um dado costuma ser uma característica relativamente fácil de prover e verificar, inclusive de forma automatizada. Há ferramentas de software especializadas na tarefa de “data cleaning” que executam a verificação de validade de campos, incluindo aí a compatibilização entre formatos diferentes para o mesmo campo em duas bases que precisam ser integradas (RAHM & DO, 2000). Por isso, falhas de validade, especialmente em campos importantes, pode indicar que os controles de entrada e atualização de campos estão muito fracos – o que pode abrir portas para fraudes. E isso é especialmente importante nos cadastros sobre pessoas.

No Sistema Nacional de Informações de Registro Civil (SIRC), que é o atual sistema para registro das certidões de óbito em substituição ao Sisobi, verifica-se que foram inseridos CPFs sem a correta validação do DV, conforme a Tabela 6 listada abaixo. A última ocorrência dessa situação foi em 2017.

**Tabela 6 - lista de CPF's com formatos inválidos no SIRC**

<b>Número do CPF</b>	<b>Data do Registro</b>
000000000000	04/10/2016
111111111111	11/06/1970
111111111111	31/12/2013
111111111111	13/05/2016
26038266504	27/03/2017
333333333333	27/05/1988
<b>Quantidade</b>	<b>6</b>

Fonte: dados obtidos no LabContas – laboratório de bases de dados do TCU, em 01/06/2022

Para verificar a validade dos dados, então é preciso compará-los com as restrições de domínio. A contabilização pode ser dar em função da quantidade de registros com erros e também em função da criticidade dos campos inválidos (por

exemplo, campos chave inválidos são mais impactantes do que um outro atributo). De todo modo, vale lembrar que os dados podem ser válidos e, mesmo assim, não terem acurácia.

#### 6.2.1.6. Acurácia

Em estatística, acurácia é o indicador da estimativa de um valor que traduz a proximidade da observação ao valor que se procura/espera atingir (GUIMARÃES, 2008) – ou seja, indica o quanto o valor “do modelo” expressa “o valor real”. Expandindo esse conceito, acurácia (ou exatidão) diz respeito ao quão precisamente os dados representam as entidades mapeadas na vida real. (DAMA 2003). Para Olson (2003) e Loshin (2011), é a maior e mais importante dimensão da qualidade de dados, uma vez que a sua insuficiência custa muito em termos de decisões erradas, perda de oportunidades e custos de correção posterior:

Especialistas estimam que o custo de dados de baixa qualidade é de 15 a 25% do lucro operacional. Em uma pesquisa recente com 599 empresas conduzida pela Pricewaterhouse Coopers, uma estimativa de má gestão de dados custa às empresas mais de US\$ 1,4 bilhão por ano apenas em problemas de faturamento, contabilidade e estoque. Grande parte desse custo é atribuído à dimensão acurácia da qualidade dos dados. Esta situação não se restringe às empresas: custos semelhantes também podem ser encontrados em organizações governamentais ou educacionais. (Olsen,2003) *tradução livre*.

Acurácia costuma ser difícil de verificar e de ajustar na base, quando não está satisfatória – ela não pode ter uma intervenção pontual, sem avaliar as causas da imprecisão, sob pena voltar a baixar a qualidade rapidamente (OLSON, 2003).

A comparação com o “mundo real” pode ser complexa e exigir a atuação em processos e recursos fora do universo da governança de dados. De fato, especialmente no caso de cadastro sobre pessoas, pode ser preciso repetir uma coleta de dados, confirmar informações junto ao cidadão ou realizar auditorias (presenciais ou não) (LEITE, 2017).

Por causa dessa dificuldade, muitas avaliações de acurácia acabam sendo feitas por meio de cruzamento com outra base de dados para a qual já foi verificada acurácia (isto é, uma base considerada “correta”) ou entre várias bases de dados disponíveis (quando é arbitrada uma “probabilidade” de acurácia a partir do nível de consistência entre essas bases) (DAMA, 2015; LOSHIN, 2011). É possível haver

muitas fontes potenciais de informações corretas: bancos de dados “mestres” ou “de referência”, um conjunto semelhante e corroborativo de valores de dados de outra tabela ou calculados dinamicamente/combinados.

Nas auditorias do TCU, é frequente a necessidade de verificação de acurácia como ponto de partida para avaliação da ação pública. Um exemplo acontece na Fiscalização Contínua de Benefícios, na qual o TCU avalia, a cada período (em geral, anual) a base de pessoas que recebeu pagamento previdenciário, assistencial ou trabalhista. Tomando como exemplo a fiscalização do ano de 2020, para os benefícios previdenciários, um dos primeiros passos da avaliação é responder à pergunta: “A pessoa para quem está registrado o pagamento desse benefício existe realmente?”. Trata-se, claramente, de um problema de acurácia. Não é viável ao TCU promover um recadastramento ou uma “prova de vida” para cada beneficiário ou mesmo uma amostra deles – como o faz, sistematicamente, o INSS, exatamente para verificação da acurácia do registro. Sendo assim, o TCU opta por coletar um conjunto de atributos que identifica cada pessoa que consta na base de pagamentos, e verificar a consistência em relação a outras bases a que tem acesso e que contém informações cadastrais: a base de CPF, da Receita Federal, e do Título de Eleitor, do Tribunal Superior Eleitoral. Ao comparar as três bases, utilizando-se de programação em linguagem para acesso a bancos de dados (SQL) e ferramentas automatizadas de comparação de campos de texto e data, a avaliação compara os campos que são chaves possíveis como identificadores únicos (CPF e título de eleitor) e dados descritivos da pessoa (nome, data de nascimento, nome da mãe, nome do pai – se houver). Quando todos os campos estão consistentes entre si, o critério de acurácia é satisfeito totalmente; a partir daí, há uma “escala” de aceitabilidade, considerando a consistência de parte dos campos em parte das bases, até um nível considerado “inconclusivo” (quando a incerteza não permite afirmar que se trata de uma pessoa real) ou até “não combinado” (os campos são tão inconsistentes entre si, inclusive os campos chave, que parece tratar-se de pessoas diferentes e uma base refuta a outra).

**Tabela 7 - Resultado da comparação de pessoas nas bases de pagamentos, de CPF e de Título de eleitor**

Avaliação da Acurácia para os titulares de benefício (referência 12/2020)		
Resultado do critério aplicado	Total	
Confirmado	35.297.565	89,95%
Inconclusivo	148.935	0,38%
Não aceito	3.794.336	9,67%
<b>Total</b>	<b>39.240.836</b>	<b>100,00%</b>

Quanto às pessoas titulares de pagamentos previdenciários em 12/2020, a base de pagamentos possui 89,95% de acurácia.

Fonte: adaptado de “Relatório de Acompanhamento do Instituto Nacional de Seguro Social” TCU, 2020.

Observamos que a acurácia final da base de pagamentos é de 89,95%.

Aplicando teste similar nos dados do Sisobi, conforme se observa na Tabela 8, considerando a comparação com a base CPF, a acurácia foi de 86,4%

**Tabela 8 - Registros nos quais não se consegue confirmar pessoa no Sisob**

Mesmo nº de CPF no Sisobi e na base CPF			
Igualdade dos nomes	Total	Com datas de nascimento iguais	Com datas de nascimento diferentes
Nomes iguais	16.008.966	15.236.542	653.784
Nomes diferentes	1.647.549	1.435.002	195.682
<b>TOTAL</b>	<b>17.656.515</b>	<b>16.671.544</b>	<b>849.466</b>

Fonte: LabContas – Laboratório de Bases de dados do TCU (2022)

### 6.2.1.7. Legalidade (Compliance)

A legalidade é a dimensão da qualidade que explicita se a posse e tratamento de um conjunto de dados concorda com os marcos regulatórios mais recentes (ENAP, 2019). É uma dimensão dependente do contexto dos dados, e pode se aplicar a cada

campo, registro ou à base como um todo. Inclui como referência regras internas da organização, normativos legais e infra legais (Pipino, 2002).

No caso de cadastros públicos sobre pessoas no Brasil, há pelos menos os marcos regulatórios:

- a) próprios da política que implementa;
- b) Lei Geral de Proteção de Dados;
- c) Lei de Acesso à Informação.

A exigência e compromisso de qualidade de dados nos cadastros públicos de pessoas físicas por parte da administração está materializada em alguns normativos principais na Lei de Acesso à Informação (Lei 12527/2011):

Art. 4º (...)

V - tratamento da informação: conjunto de ações referentes à produção, recepção, classificação, utilização, acesso, reprodução, transporte, transmissão, distribuição, arquivamento, armazenamento, eliminação, avaliação, destinação ou controle da informação;

VI - disponibilidade: qualidade da informação que pode ser conhecida e utilizada por indivíduos, equipamentos ou sistemas autorizados;

VII - autenticidade: qualidade da informação que tenha sido produzida, expedida, recebida ou modificada por determinado indivíduo, equipamento ou sistema;

VIII - integridade: qualidade da informação não modificada, inclusive quanto à origem, trânsito e destino;

IX - primariedade: qualidade da informação coletada na fonte, com o máximo de detalhamento possível, sem modificações.

Art. 6º Cabe aos órgãos e entidades do poder público, observadas as normas e procedimentos específicos aplicáveis, assegurar a:

I - gestão transparente da informação, propiciando amplo acesso a ela e sua divulgação;

II - proteção da informação, garantindo-se sua disponibilidade, autenticidade e integridade; e

III - proteção da informação sigilosa e da informação pessoal, observada a sua disponibilidade, autenticidade, integridade e eventual restrição de acesso

Já a Lei de Proteção a Dados Pessoais (Lei 13709/2018) estabelece que:

Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios: (...)

V - qualidade dos dados: garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;

As leis, decretos e outros normativos que dão suporte às políticas públicas devem estar definidos em função da abrangência da política, e com ampla divulgação para os cidadãos. Quando isso não acontece, pode representar um risco para a qualidade da base.

O cadastro do CPF, por exemplo, apesar de ter sido definido por força do Decreto-lei nº 401 de 30 de dezembro de 1968, foi ao longo dos anos 80-90, quando a Secretaria da Receita Federal teve forte desenvolvimento tecnológico que processos de gestão da sua base de dados foram se estruturando e tomando maior publicidade. Ainda assim, somente em 2015 o CPF teve regulamentados aspectos de sua gestão, por meio da Instrução Normativa RFB Nº 1548, de 13 de fevereiro de 2015; de lá pra cá, foram 13 alterações documentadas no site da Receita – por meio de outras instruções normativas ou atos declaratórios da comissão de cadastros (COCAD). Essa situação legal coloca risco à usabilidade do CPF por outros órgãos da administração, pois alterações sobre a base podem ser feitas sob a gestão da Receita.

Além disso, a legislação incidente deve estar coerente entre si, para que possa atender às necessidades da política e estabelecer com clareza os requisitos do cadastro. Um exemplo dos conflitos que podem ocorrer e que impactam dimensões da qualidade é o próprio caso da identidade única no Brasil, cuja iniciativa partindo de várias origens e momentos ainda não permite que um único identificador garanta a unicidade dos registros de pessoas no Brasil.

Assim, a verificação da existência de leis precisas, de âmbito adequado, com coerência entre si faz parte da avaliação de qualidade de cadastros de pessoas.

#### **6.2.1.8. Usabilidade:**

Conforme atenta OLSON (2003), um valor não pode ser preciso se o usuário do valor não for capaz de dizer o que ele é, o que significa ou o contexto no qual ele é apresentado – caso contrário, a interpretação e o uso do dado pode ser inadequado. O critério de qualidade que visa garantir a compreensão do dado para o seu uso adequado é a usabilidade.

Conforme visto anteriormente neste documento, a qualidade das bases de dados cadastrais é avaliada por seus consumidores – e, em geral, há mais de um perfil de consumidor para esses dados. De fato, podemos identificar alguns perfis de consumo (Leite et al, 2017):

- a) o usuário público-alvo do cadastro;
- b) os administradores/operadores do cadastro;
- c) os responsáveis pelo controle de qualidade e auditorias sobre o cadastro;

- d) outros órgãos da administração pública, que utilizam o cadastro como suporte para outras políticas.

Para o público-alvo do cadastro, as interfaces dos aplicativos – ou das pessoas – são ponto chave para garantir a compreensão dos dados. O vocabulário utilizado, o formato do cadastro, o meio pelo qual interagem na coleta e atualização de dados impacta significativamente na precisão dos dados que fornecem (Leite et al, 2017).

Para os administradores/operadores do cadastro é preciso a compreensão da semântica dos dados, bem como o conhecimento sobre o negócio e a competência para operar a política suportada pelo cadastro (Leite et al, 2017).

O Ministério do Desenvolvimento, Previdência e Trabalho, por exemplo, realiza pesquisas e campanhas regulares junto ao seu público-alvo sobre o cadastro único. Além disso, os entrevistadores dispõem de Manual explicativo para coleta de informações junto ao público-alvo – em geral, formado por famílias de baixa escolaridade. Ainda assim, não se pôde localizar nenhum estudo sobre imprecisão de dados coletados ou informados pelo aplicativo em função de erro de compreensão sobre semântica ou objetivo de informação a ser fornecida para o cadastro.

Para os responsáveis pela qualidade do cadastro, e aqueles que realizam as auditorias sobre eles – como o caso do TCU – é preciso também o conhecimento da política e da semântica dos dados. Além disso, é fundamental que os dados estejam disponíveis, facilmente/rapidamente recuperáveis, aptos a manipulação e aplicação em diversas tarefas.

O escritório de auditora dos Estados Unidos (GAO), por exemplo, defende que a confiabilidade dos dados utilizados para as auditorias é um dos requisitos essenciais do trabalho do auditor, e associa a acurácia, completude e pertinência dos dados como as dimensões de maior importância. (GAO, 2019) Por isso, argumenta, para que os dados em bases de dados possam ser utilizáveis em auditorias, os metadados sobre eles precisam estar disponíveis e serem claros – especialmente sobre as dimensões de qualidade elencadas.

Para os outros entes da administração que precisem dos dados, é preciso de modo especial conhecer o contexto em que foram criados para que a apreensão desses dados para outras políticas seja feita de forma adequada (Enap, 2019).

Em todos os casos de consumidores, há um insumo em comum que precisa estar adequado para suportar a qualidade dos dados quanto à usabilidade: os metadados – isto é, as “informações sobre os dados”. Este conjunto de informações



costuma ser frequentemente negligenciado, e isso impacta diretamente a qualidade dos cadastros (DAMA, 2003). Segundo preconizam os processos definidos no DMBok, deveria ser mantido um repositório para cada atributo de uma base de dados, informando descrição, valores possíveis, fontes dos dados, informações de atualização (essencialmente informações do dicionário de dados), além de níveis de qualidade da base. Essas informações são insumos para garantir uma usabilidade dos dados.

#### **6.2.1.9. Transparência, sigilo e privacidade**

Em cadastros sobre pessoas, várias categorias de informações pessoais podem ser consideradas sensíveis ou críticas para a segurança pessoal ou relações sociais (Leite et al, 2017). O controle de acesso aos dados de uma base de cadastro sobre pessoas é um requisito para que o sigilo e a privacidade dos dados sejam atendidos. No entanto, não é possível avaliar o sigilo apenas olhando os dados, e também não é simples avaliar se controle de acesso está adequado – mesmo quando avaliamos os sistemas de informação que operam sobre os dados (MOSES et al, 2022).

Por isso, é mais eficiente adotar uma estratégia conjunta de verificar as definições de sigilo e privacidade para o cadastro, bem como testar o acesso aos aplicativos que operacionalizam o cadastro (quando isso for possível).

Caso não haja definições claras sobre controles de acesso, sigilo, integridade e privacidade podem ser violados. A política suportada pelo cadastro pode ser afetada pela ausência desses controles, levando ao comprometimento também de outras dimensões de qualidade por meio de alteração indevida de dados, situações não previstas de inconsistência, fraudes e falhas na execução da política.

Um exemplo de normativo sobre controle de acesso aos dados é a Portaria nº 502, de 29 de novembro de 2017, que institui a Política de Controle de Acesso aos dados do Cadastro Único para Programas Sociais.

A respeito dos controles, há requisitos básicos e relacionados às políticas de segurança em geral do gestor dos dados:

- a) O gestor possui uma política de segurança da informação documentada e uma política de confidencialidade de informações pessoais?

- b) Há listas de controle de acesso a dados, funções de usuário e níveis de segurança definidos para determinar quem pode ter acesso aos dados e visões do cadastro?
- c) Como é feita a autenticação dos usuários e dos administradores/operadores do cadastro? E de outros órgãos da administração pública?

No caso brasileiro, a Lei Geral de Proteção de Dados (LGPD) estabelece parâmetros para o compartilhamento de dados pessoais, bem como papéis e responsabilidades ao gestor público que executa tratamento desses dados. A violação desses comandos legais não compromete a qualidade individual dos dados, mas pode comprometer a eficiência/eficácia da política em particular, bem como a confiabilidade da base de dados como um todo.

De outro lado, a Lei de Acesso à Informação preconiza que aos dados públicos seja dada a transparência, desde que não afete o sigilo necessário às informações. No caso dos cadastros públicos sobre pessoas, algumas informações devem estar disponíveis de forma ampla, em consonância com os preceitos de ambas as leis:

- a) os processos de gestão e governança aplicados ao cadastro, e eventuais relatórios gerais produzidos (estatísticas, resultados, etc);
- b) normativos legais e infralegais que organizam a gestão e governança do cadastro;
- c) quais os órgãos da administração pública têm acesso a quais dados do cadastro;
- d) quais dados estão armazenados sobre os cidadãos, de maneira ampla (ou seja, o “esquema” do banco de dados do cadastro) e o conteúdo dos dados armazenados (nesse caso apenas para o titular dos dados);

Embora esses controles estejam definidos em diplomas legais, a diversidade de formas de gestão nem sempre corresponde à realidade brasileira. Dois exemplos emblemáticos podem ser citados:

- a) O TSE, gestor do título de eleitor, disponibiliza em sítio próprio todos os dados sobre o cadastro, normativos e estatísticas. Além disso, dúvidas apresentadas na ouvidoria do órgão, postadas durante a realização desse estudo, tiveram resposta em 24h.
- b) A Receita Federal, gestora da base do CPF, não disponibiliza informações amplas sobre a gestão do cadastro. Consultada via Fala.Br (plataforma de

acesso a informações do executivo federal), após recurso frente à negativa de acesso a normas de “qualidade do cadastro”, respondeu apenas com um dicionário de dados em documento não oficial, e ponderou que “o acesso às normas de controle de qualidade do cadastro não interessa ao público externo, nem mesmo à Receita em geral, pois trata somente de procedimentos e projetos internos”. Essa situação se torna mais significativa se considerarmos que o CPF é, atualmente, o número de referência para identificar cidadãos no Brasil.

#### 6.2.1.10. Uma palavra sobre sobreposição e agregação de dimensões de qualidade...

Como chamou atenção Myers (2013) e Olson (2003), não há uniformidade nos estudos sobre as dimensões de qualidade mais importantes, bem como há sobreposição de conceitos em muitas definições. Nos exemplos aqui tratados, é possível identificar alguma sobreposição entre consistência, unidade e acurácia, bem como legalidade e transparência/sigilo/privacidade.

De todo modo, é importante notar que o objetivo é decompor a qualidade em termos avaliáveis e estratégicos para o contexto em questão – cadastros públicos sobre pessoas – a fim de permitir o gerenciamento e o aprimoramento da qualidade, e gerar cadastros **confiáveis** para suporte às políticas públicas.

Dentre as possíveis definições de confiabilidade, neste trabalho concorda-se com Pavon et al (2019) que um conjunto de informações confiável é aquele em que, entre outros aspectos, as bases de dados associadas têm a qualidade requerida pelos usuários, de tal forma que seja possível gerar informação útil a partir deles. Nesse sentido, a confiabilidade tem sido associada à resiliência a erros ou falhas e a quão bem os dados estão na avaliação das dimensões aceitas como estratégicas para a qualidade (LOSHIN, 2011; COLEMANN, 2012; PAVON et al, 2019; MOSES et al, 2022). Além disso, a confiabilidade tem sido interpretada como um acordo entre expectativas e capacidade da organização responsável por sua gestão (COLEMANN, 2012). Essas definições são apropriadas para aplicação nos cadastros públicos sobre pessoas, nos quais há uma grande diversidade de políticas, gestores e capacidades, bem como requisitos de qualidade que serão mensuráveis e estratégicos em pesos diferentes, de acordo com o contexto. Assim, para uma aplicação, uma base pode ser

confiável, e para outra, não ter confiabilidade suficiente – mas em ambos os casos, é preciso avaliar para saber.

### 6.2.2. Processos que afetam a qualidade dos cadastros

A qualidade dos dados depende do desenho e da implementação dos processos de produção envolvidos na geração e gestão dos dados (WAND & WANG, 1996). A “Figura 5 Ciclo de vida de cadastros públicos sobre pessoas” mostra etapas comuns no ciclo de vida dos cadastros. Falhas em quaisquer dessas etapas podem levar a falhas de qualidade nos dados. Sendo assim, Leite et al (2017) sugerem algumas questões importantes a que os gestores devem atentar:

- a) Ter clareza sobre como o cadastro dá suporte para a política pública (os momentos na política, as necessidades informacionais, etc);
- b) Estimar custos de criação e manutenção do cadastro (infraestrutura, contratos, agentes operadores, etc)
- c) Prover arranjos institucionais adequados para operacionalizar o cadastro, especialmente a designação de gestor (levando em conta abrangência do cadastro, competência e capacidade dos órgãos operadores, suporte legal, etc);
- d) Definir o processo de operacionalização da política e do cadastro, de maneira documentada (por exemplo, usando um modelo de processo de trabalho, mantendo atualizado um dicionário de dados) e monitorada;
- e) Manter estrutura de tecnologia para suporte à informatização e gestão/governança dos dados da base do cadastro;
- f) Manter processos de controle, avaliação e melhoria contínua dos processos e dos dados dos cadastros.

A gestão sobre etapas dos processos de produção dos cadastros tem contornos específicos se levamos em conta a burocracia estatal envolvida na sua execução. No livro conhecido e de título sugestivo “Implementação: como grandes expectativas em Washington são frustradas em Oakland”, em 1973 Pressman e Wildavsky já chamavam atenção para as grandes dificuldades não apenas da tradução das intenções em ações, como também da compreensão e engajamento das chamadas “burocracias do nível de rua” nas políticas públicas. Em estudo mais

recente, LIPSKY (2010) chama atenção que os meios informatizados (sistemas e bases de dados) muitas vezes conduzem os processos de trabalho. No caso de cadastros públicos de pessoas, a qualidade pode ser afetada não só pelos processos da política suportados pelo cadastro como também pelos interesses conflitantes dos executores da política.

Segundo Morton (2021), uma maneira de identificar esses conflitos é a construção de uma tabela com o balanço de poder e interesse (BPI) para os stakeholders da política pública, de modo a refletir os conflitos de agência que podem acontecer nos processos de trabalho que viabilizam a política e o cadastro e envolvem burocracia de nível de rua, os clientes e os gestores. O objetivo aqui é explicitar possíveis pontos de falha devido a conflitos de interesse

Para exemplificar o tipo de questões que compõem essas tabelas, apresentam-se a seguir as tabelas com o BPI dentre os cadastros analisados neste trabalho.

Implementação do Cadastro de Pessoa Física				
	Stakeholder	Nível federado de atuação	Interesse	Poder
1	Contribuinte	federal	receber mais e pagar menos	voto
2	Receita Federal	federal	garantir a cobrança do imposto correto	gestão do cadastro; punições previstas em lei (multar, inscrever em dívida ativa); fornecer o recurso orçamentário para o governo
3	Governo Federal	federal	arrecadar mais imposto	definir os impostos a cobrar
4	Órgãos da administração que usam CPF	municipal, estadual, federal	ter cadastro acessível que identifique unicamente uma pessoa	vincular a prestação do serviço público ao fornecimento do CPF
5	Bancos	municipal, estadual, federal	conhecer situação fiscal e de crédito de cada CPF	vincular a prestação do serviço bancário à regularidade do CPF
6	Lojistas	municipal, estadual, federal	conhecer situação de crédito de cada CPF	vincular o financiamento de compra ao crédito do CPF; usar o banco que assume mais riscos de créditos ao menor custo;
7	ongs defensoras da privacidade	municipal, estadual, federal	garantir a privacidade dos dados dos cidadãos	contatos com bases políticas; visibilidade na mídia

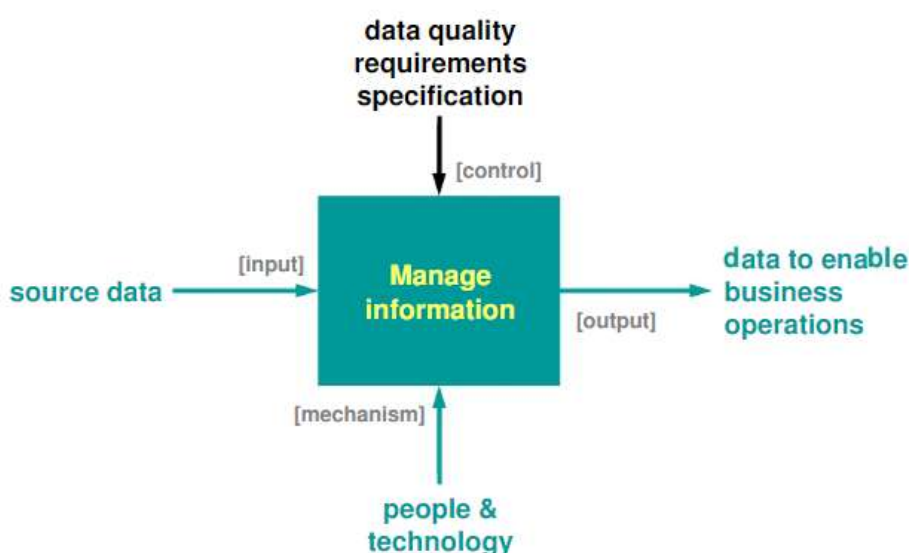
Implementação do Cadastro Único				
	Stakeholder	Nível federado de atuação	Interesse	Poder
1	População de baixa renda	municipal, estadual, federal	Complementar renda	voto
2	INSS	federal		Reconhecer ou negar o direito ao recebimento de benefícios administrados pela Previdência Social
3	Governo Federal	federal		Destinar orçamento e regras para manter o cadastro único
4	Caixa Econômica Federal	municipal, estadual, federal	ter cadastro acessível que identifique unicamente uma pessoa	Liberar o pagamento de benefícios para as pessoas do cadastro único
5	CRAS	municipal		Incluir as pessoas no cadastro único
6	Gestores municipais	municipal	Receber os repasses para o maior número de municípios referente às políticas do cadastro único	Validar os cadastros e repassar os recursos

Implementação da base de eleitores				
	Stakeholder	Nível federado de atuação	Interesse	Poder
1	Contribuinte	federal	receber mais e pagar menos	voto
2	Receita Federal	federal	garantir a cobrança do imposto correto	gestão do cadastro; punições previstas em lei (multar, inscrever em dívida ativa); fornecer o recurso orçamentário para o governo
3	Governo Federal	federal	arrecadar mais imposto	definir os impostos a cobrar
4	órgãos da administração que usam CPF	municipal, estadual, federal	ter cadastro acessível que identifique unicamente uma pessoa	vincular a prestação do serviço público ao fornecimento do CPF
5	Bancos	municipal, estadual, federal	conhecer situação fiscal e de crédito de cada CPF	vincular a prestação do serviço bancário à regularidade do CPF
6	Lojistas	municipal, estadual, federal	conhecer situação de crédito de cada CPF	vincular o financiamento de compra ao crédito do CPF; usar o banco que assume mais riscos de créditos ao menor custo;
7	ongs defensoras da privacidade	municipal, estadual, federal	garantir a privacidade dos dados dos cidadãos	contatos com bases políticas; visibilidade na mídia



De acordo com a norma ISO 9001, o conceito de qualidade compreende medição e controle, melhoria contínua, alcance de metas e objetivos e envolvimento da alta direção. Já na família de normas específicas de qualidade de dados e dados mestre, ISO 8000, a confiabilidade dos dados é vista como um efeito dos níveis de qualidade das bases, descritas em dimensões de acordo com as prioridades da organização, e que só podem ser mantidas mediante um processo de tratamento da qualidade.

**Figura 9 - Qualidade de dados no contexto das organizações**



Fonte: "Data Quality", in ISO 8000.

A partir dessas definições, infere-se que o gerenciamento da qualidade dos dados é um programa, não um projeto. De fato, nenhuma organização que pretenda manter qualidade dos dados que consome consegue fazê-lo sem um esforço contínuo de avaliação e melhoria das dimensões mais importantes para a qualidade dos cadastros. Além disso, o processo de avaliação da qualidade e das expectativas das partes interessadas também precisa ser contínuo, visto que os requisitos sofrem mudanças com o tempo, à medida que as necessidades dos negócios e a influência das forças externas evoluem (OLSON, 2003).

Em sintonia com essa visão, no DAMABok (2015) a gestão de qualidade é parte de um processo de governança de dados – é muito mais eficiente para as

organizações incluir a manutenção da qualidade dos dados nos processos de governança dos dados.

A governança de dados na administração pública pode ser entendida como

o conjunto de políticas, processos, pessoas e tecnologias que visam a estruturar e administrar os ativos de informação, com o objetivo de aprimorar a eficiência dos processos de gestão e da qualidade dos dados, a fim de promover eficiência operacional, bem como garantir a confiabilidade das informações que suportam a tomada de decisão.  
(Enap, 2019)

Já o Decreto 10.046/2019 define governança de dados como “exercício de autoridade e controle que permite o gerenciamento de dados sob as perspectivas do compartilhamento, da arquitetura, da segurança, da qualidade, da operação e de outros aspectos tecnológicos”.

Na mesma linha, a segunda edição do Guia Referencial Básico de Governança do Tribunal de Contas da União - TCU (2015) apresenta um resumo das características sobre gestão e governança: definir direcionamento estratégico, supervisionar gestão, envolver partes interessadas, gerenciar conflitos, auditar o sistema de gestão e controle e promover *accountability* e transparência. Essas características estão em consonância com o conceito apresentado no Decreto 9.203/2017, no qual a governança pública compreende o conjunto de mecanismos de liderança, estratégia e controle postos em prática para avaliar, direcionar e monitorar a gestão, com vistas à condução de políticas públicas e à prestação de serviços de interesse da sociedade.

Pode-se inferir, então, que a qualidade dos dados governamentais está inserida da governança de dados governamentais, que é, por sua vez, parte da governança pública. A boa gestão de dados governamentais em cadastros sobre pessoas, então, é parte de uma boa governança pública.

Finalmente, vale lembrar a visão de Maydanchik (2007):

De fato, nenhuma quantidade de lenha o aquecerá no inverno, a menos que você consiga fazer um fogo decente. O principal produto da avaliação de qualidade de dados é o scorecard de qualidade de dados pois fornece informações abrangentes sobre a qualidade dos dados em um banco de dados e permite tanto análises agregadas quanto detalhadas. Pontuações agregadas ajudam a entender os inúmeros relatórios de erros produzidos durante a avaliação da qualidade dos dados e apontam para a confiabilidade da base de dados.  
(tradução livre)

Assim, qualquer organização que almeja melhorar a qualidade de seus dados deve adotar ou desenvolver um conjunto de dimensões por meio das quais possa mensurar a qualidade e alcançar um consenso sobre as dimensões que podem fornecer um ponto de partida para a construção de um vocabulário comum sobre o tema. A partir daí, é a construção de indicadores relevantes, e a gestão desses indicadores ao longo do tempo que manterão a boa qualidade das bases de dados. (MAYDANCHIK, 2017; OLSON, 2003).

No contexto da atuação do TCU, identificar bases de dados de baixa qualidade aponta achados de auditoria sobre falhas de governança na gestão de políticas públicas. Complementarmente, o objetivo é fomentar melhoria de qualidade dessas bases, para que possam ser utilizadas na avaliação e aperfeiçoamento das próprias políticas públicas que são suportadas por essas bases.

## **7. Síntese da proposta de uma ferramenta de avaliação**

Dada a perspectiva prática de uso de uma tal ferramenta como guia ou ponto de partida para realização de avaliações de qualidade de dados em cadastros públicos de pessoas, optou-se por agrupar os itens de cada dimensão no formato de questões, e listar possíveis fontes e indicadores para obter os insumos para a avaliação.

Também não é necessário que, em toda avaliação, todas as dimensões aqui listadas sejam objeto de análise. De fato, dada a necessidade de avaliação e melhoria contínua dos cadastros, é perfeitamente possível a realização de avaliações parciais em amplitude de critérios, mas que permitam o aprofundamento nas dimensões avaliadas.

Nesse sentido, a tabela a seguir sintetiza as questões discutidas ao longo do item 6 deste trabalho.

## Checklist para avaliação de cadastros públicos sobre pessoas

### Dimensão: Completude

#### proporção de dados armazenados em relação ao total do universo definido para a base

<i>Questões</i>	<i>Possíveis evidências ou indicadores</i>
Qual a cobertura do cadastro relativo ao público-alvo pretendido?	Comparação entre estimativas estatísticas para o público-alvo e a quantidade de registros ativos. Comparar dados do censo ou Pnad continua, por exemplo.
Uma pessoa pode se habilitar e ser incluída no cadastro a qualquer momento?	Documentos sobre o processo do cadastro (normativos, manuais)
Os requisitos para ingressar no cadastro são compatíveis e acessíveis ao público-alvo?	Documentos sobre o processo do cadastro (normativos, manuais)
Os campos de preenchimento obrigatório estão informados para os registros existentes?	Verificação dos campos da base de dados, conforme as regras de negócio e o dicionário de dados.

### Dimensão: Unicidade ou exclusividade

#### definição de identificação única para cada registro e não há duplicidade desses registros;

<i>Questões</i>	<i>Possíveis evidências ou indicadores</i>
Há uma definição clara do atributo que identifica unicamente uma pessoa no cadastro (chave)?	Modelo e Dicionário de dados do cadastro
Há apenas uma ocorrência de cada pessoa no cadastro?	Verificação de unicidade da chave na base de dados; verificação

### Dimensão: Oportunidade ou tempestividade

#### grau em que os dados representam a realidade para o momento em que é necessário;

<i>Questões</i>	<i>Possíveis evidências ou indicadores</i>
Há prazo limite definido para atualização dos dados do cadastro?	Documentos e manuais sobre o cadastro; legislação

Os registros estão dentro do prazo de atualização?	Verificação na base de dados e comparação com o definido no dicionário de dados
O prazo de atualização atende à política suportada e ao público-alvo?	Averiguação junto às definições da política que suporta, se há ciclos, etc. Verificar se o público
O mecanismo de atualização do cadastro é adequado à política e ao público-alvo?	Verificar se há reclamações no Fala.BR ou outra base de reclamações/atendimento. Caso haja aplicativo, verificar na avaliação do App. Comparar com outros serviços prestados pelo gestor para o público-alvo. Considerar fatores como escolaridade, renda, idade, etc
<b>Dimensão: Validade</b>	
<b>conformidade com a sintaxe definida para o dado (formato, tipo, intervalo, ....)</b>	
<i>Questões</i>	<i>Possíveis evidências ou indicadores</i>
Há dicionário de dados definido e documentado para o cadastro que inclua tipo, formato, intervalo e outras características da sintaxe dos campos?	Documentos e manuais sobre o cadastro; legislação
Os campos atendem à sintaxe definida?	Verificação dos campos na base de dados
Os campos de preenchimento obrigatório estão informados para os registros existentes?	Verificação dos campos na base de dados
<b>Dimensão: Precisão ou acurácia</b>	
<b>grau de aderência entre os dados e o objeto do mundo real que representam;</b>	
<i>Questões</i>	<i>Possíveis evidências ou indicadores</i>
Há procedimentos de verificação dos dados do cadastro por meio de circularização com outras fontes?	Documentos e manuais sobre o cadastro; legislação
Há procedimentos de verificação presencial dos dados do cadastro, tais como visitas, provas de vida e outros?	Documentos e manuais sobre o cadastro; legislação
Qual o nível de acurácia dos registros?	Verificação de acurácia por meio de circularização com outras bases; testes in loco.
<b>Dimensão: Consistência</b>	

### ausência de diferença ao comparar várias representações de uma mesma coisa, a partir de uma definição

<i>Questões</i>	<i>Possíveis evidências ou indicadores</i>
A semântica dos atributos está coerente nos diferentes registros?	Comparação entre o dicionário de dados e as informações nos registros da base.
A semântica dos atributos está coerente no tempo?	Comparação entre versões de modelos/dicionário de dados e verificação substantiva nos registros da base
Os campos relacionados entre si de acordo com regras de negócio estão coerentes? (ex: se um campo de naturalidade do país não é coerente com o estado, ou se do sexo feminino e possui certificado de reservista do serviço militar, ou se indica casado e não tem data de casamento, etc)	Verificação de regras de coerência entre campos nos registros da base

### Dimensão: Legalidade

#### Aderência às definições legais; compliance

<i>Questões</i>	<i>Possíveis evidências ou indicadores</i>
As regras do cadastro estão definidas em instrumento legal de amplo conhecimento, compatível com a abrangência do cadastro? (lei, decreto, instrução normativa)	Lei, decreto, instrução normativa ou outro instrumento que define o cadastro. Observar se o alcance do instrumento é compatível com a abrangência do cadastro.
Há normativos, inclusive infralegais, que definem de forma clara as regras e semânticas dos itens do cadastro, de modo a não haver interpretações conflitantes entre os atores?	
Os marcos legais da política que é suportada pelo cadastro estão bem definidos?	
Os dados evidenciam atendimento a esses marcos legais?	Há inconsistência nos dados das bases que evidenciam o não atendimento aos marcos legais
Há legislação conflitante em aspectos do cadastro (gestão, sobreposição de políticas, incentivos conflitantes com o objetivo ou funcionamento do cadastro)	
A LAI e a LGPD estão atendidas?	

Dimensão: Usabilidade	
Facilidade de uso, compreensibilidade	
Questões	Possíveis evidências ou indicadores
O vocabulário e formato do cadastro é adequado e facilita a compreensão do público-alvo?	Verificar se há manuais e guias; verificar se há reclamações no Fala.BR ou outra base de reclamações/atendimento. Caso haja App, verificar na avaliação do App.
O instrumento utilizado para coleta e atualização de dados é adequado ao público-alvo?	Verificar se há reclamações no Fala.BR ou outra base de reclamações/atendimento. Caso haja App, verificar na avaliação do App. Comparar com outros serviços prestados pelo gestor para o público-alvo. Considerar fatores como escolaridade, renda, idade, etc
Há local onde consultar dúvidas frequentes sobre o cadastro?	Verificação de FAQ em sites, vídeos, publicidade ou outros meios para tirar dúvida; verificar se há telefone ou atendimento presencial para dúvidas, se for o caso.
Dimensão: Transparência, Sigilo e Privacidade	
Controle de acesso adequado ao grau de confidencialidade	
Questões	Possíveis evidências ou indicadores
As informações sobre o cadastro estão facilmente disponíveis ao cidadão?	Verificação em sites e grande mídia
Qual o grau de conhecimento do público-alvo sobre o cadastro?	Verificar se há campanhas de divulgação sobre o cadastro
Os dados públicos do cadastro são disponibilizados como dados abertos?	Verificar no cadastro no portal brasileiro de dados abertos ou no site do gestor.
As informações que compõem o cadastro estão classificadas quanto à confidencialidade?	Metadados de segurança Resultado do questionário geral de governança, item: “informações pessoais são identificadas e rotuladas, com vistas a viabilizar adequado

	tratamento e proteção (gov geral 2021)”, respondido pelo gestor
Estão designados o controlador, operador e encarregado do cadastro, nos termos da LGPD?	Informação em site ou obtida junto ao gestor
Há informações de violações de privacidade dos dados do cadastro?	Verificação na base de reclamações do Fala.BR e em estatísticas da ANPD
<b>Dimensão: Processos que afetam a qualidade de dados</b>	
Os custos, contratos e orçamento relacionados ao cadastro estão identificados?	Contratos e orçamento para os processos relacionados ao cadastro
O gestor do cadastro é claramente identificado?	Documentos e manuais sobre o cadastro; legislação.
Há dicionário de dados definido e disponível?	
Os processos de inclusão e atualização de dados, definições e melhorias de estrutura do cadastro estão definidos?	Documentos e manuais sobre o cadastro; legislação.
O próprio gestor realiza avaliações sobre o cadastro? Com que escopo? Com que periodicidade?	Resultados de auditorias de gestor
Há instâncias reguladoras relacionadas com o cadastro? São feitas auditorias no cadastro? Com que escopo? Com que periodicidade?	Resultados de auditorias de instância reguladora
A CGU tem feito auditorias no cadastro? Com que escopo? Com que periodicidade?	Resultados de auditorias de CGU
O TCU tem feito auditorias no cadastro? Com que escopo? Com que periodicidade?	Resultados de auditorias de TCU – coleta possível no DW
Como têm evoluído no tempo os principais achados de auditoria?	Série temporal de achados de auditoria selecionados



## 8. Considerações finais

Ao longo deste trabalho, foi possível observar a multidimensionalidade da qualidade de cadastros públicos sobre pessoas. E, conforme atenta Olson (2016), o custo de baixa qualidade pode ficar escondido e não ser óbvio para aqueles que não estão procurando por problemas. Por isso, um programa contínuo de qualidade é tão importante.

Qualidade de dados nos cadastros públicos de pessoas é importante para as políticas que os suportam. Qualidade não é algo que se consegue com uma única intervenção: é preciso um processo contínuo. Melhoria na qualidade de cadastros públicos de pessoas será conseguida avaliando as dimensões de qualidade. Quais serão as dimensões, quais prioritárias entre elas e quais métricas de cada uma, dependerá dos objetivos do consumidor de dados. Como orientações, foram trazidas aqui dimensões frequentes na literatura e relevantes dentro do cenário dos cadastros brasileiros.

O instrumento de avaliação aqui proposto não pretende ser exaustivo, mas tão somente um apoio para a realização de avaliações de qualidade de cadastros, um guia que sintetiza padrões, boas práticas e a experiência de auditorias do TCU sobre o tema. Com aplicação do instrumento, estudos futuros podem avaliar aspectos mais importantes para melhoria contínua da qualidade, além de ampliar ou adaptar dimensões e seus pesos relativos para a qualidade de cada cadastro.

Também há espaço para aprofundar a questão da síntese de indicadores, de tal forma a, como argumenta Olson (2016), “expressar a confiabilidade de uma base de dados” – e em especial, uma base de cadastro sobre pessoas – de modo a tornar possível comparar a qualidade entre bases ou comparar a qualidade da mesma base, ao longo do tempo. A definição do contexto de qualidade, a importância de cada dimensão, o público-alvo consumidor do indicador são características a serem consideradas nessa construção, além das qualidades próprias de um bom indicador, tais como mensurabilidade, relevância e simplicidade.

Neste trabalho, buscou-se focar no momento de uma avaliação, reforçando sempre a importância de avaliações ao longo do tempo. Estudos futuros também podem avaliar o impacto de intervenções sobre dimensões de qualidade de uma base

na melhoria da qualidade dessa base ao longo do tempo. É possível, ainda, que a experiência permita a definição de padrões mínimos de qualidade para utilização dos registros administrativos sobre pessoas como meio de apoio a políticas públicas, seja para verificação de elegibilidade a benefícios, oferta de serviços público, controle ou mesmo planejamento e aprimoramento de políticas públicas.

No presente estudo, focamos no que consiste a qualidade dos cadastros públicos sobre pessoas. Vimos que as dimensões de qualidade dos cadastros públicos sobre pessoas são fortemente influenciadas pelas relações entre os cadastros. Embora o tema da integração não tenha sido aprofundado aqui, é importante ressaltar sua importância e que, hoje, vários esforços têm sido feitos para integrar cadastros públicos em geral. Entre esses esforços, podemos citar: a definição da “Estratégia de Governança Digital” (Decreto nº 10.332, alterado pelo Decreto nº 10.996, de 14 de março de 2022), que define diretrizes para o tema de governança, incluindo a governança de cadastros; o “Cadastro Base Cidadão” (Decreto 10.046, de 09 de outubro de 2019), que pretende ser um sistema de “dados mestres” para informações sobre cidadãos; o Conecta Brasil, que promove a troca automática e segura de informações entre os sistemas para que o cidadão não tenha que rerepresentar informações que o governo já possua (direito garantido pelas Leis 13.726/2018 e Lei 14.129/2021); e a definição de identificação única para os cidadãos. As potencialidades, acertos e erros dessas iniciativas também oferecem insumos para aprimoramento da ferramenta de avaliação de qualidade de cadastros públicos aqui apresentada.

Em 2020, o TCU avaliou que a integração entre dados no governo brasileiro ainda era baixo (TC 031.158/2020-0). O TCU pode colaborar com a administração pública, induzindo a melhoria contínua dos cadastros por meio de fiscalizações de qualidade e monitoramento dos progressos (e erros) ao longo do tempo. Além disso, sua visão mais geral do efeito de boas práticas, bem como de ausências, pode ser incluída nas recomendações aos gestores, e contribuir para disseminar esse aprendizado.

The only thing you know for sure is that you found a specific number of inaccuracies. The bad news is that there are probably more; the good news is that you found these. If the number you find is significant, you know you have a problem.  
(OLSON, 2003).

## Referências bibliográficas

- BRASIL. Casa Civil. *Avaliação de políticas públicas: guia prático de análise ex ante*. Brasília: Ipea, 2018a, v.1, 192 p. Disponível em: [http://www.ipea.gov.br/portal/index.php?option=com\\_content&view=article&id=32688&Itemid=433](http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=32688&Itemid=433). Acesso em 9 nov. 2018.
- BRASIL. Secretaria Nacional de Renda de Cidadania (Senarc). *Manual do entrevistador 4ª Edição: Cadastro Único para Programas Sociais*. Brasília: Ministério da Cidadania, 2017.
- CLARA, Adriana Maria da Silva. Um estudo comparativo de metodologias de avaliação da qualidade dos dados em sistemas de informação da área de seguridade social. Brasil, 2004.
- Cichy, Corinna & Rass, Stefan. (2019). *An Overview of Data Quality Frameworks*. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2899751.
- DASU, Tamraparni; VESONDER, Gregg; WRIGHT, Jon R. *Data quality through knowledge engineering*. Ninth International Conference on Knowledge Discovery and Data Mining ACM SIGKDD. 2003. <https://dl.acm.org/doi/pdf/10.1145/956750.956844>
- Exposição de motivos do decreto do Cadastro Único. Brasil, 2007.
- ENAP, Escola Nacional de Administração Pública. *Governança de Dados*. Conteudistas: Wellington Luiz Barbosa, Roberto Shayer Lyra. Brasília: Fundação Universidade de Brasília (FUB), Centro de Apoio ao Desenvolvimento Tecnológico (CDT), Laboratório Latitude, 2019.
- FACHIN, Odília. *Fundamentos de Metodologia*. 5ª ed. São Paulo: Saraiva, 2006.
- Guedes, Josefina Aparecida Soares, et al. Josefina Aparecida Soares Guedes, (2) Rita de Cássia Fonseca, (3) Faimara do Rocio Strauhs. Uso de indicadores e métricas para avaliação da qualidade da informação. *Brazilian Journal of Information Science: Research trends*, vol. 15, publicação contínua, 2021
- GAO, United States Government Accountability Office. *ASSESSING DATA RELIABILITY*. USA, 2019.
- Guia brasileiro de análise de dados: armadilhas & soluções / Editores Claudio D. Shikida, Leonardo Monasterio, Pedro Fernando Nery. -- Brasília: Enap, 2021
- Guimarães, Paulo Ricardo Bittencourt. *Métodos Quantitativos Estatísticos*. Curitiba: IESDE Brasil S.A., 2008.
- IBGE, Estatísticas Vitais. <https://www.ibge.gov.br/estatisticas/sociais/populacao/9110-estatisticas-do-registro-civil.html?edicao=27311&t=o-que-e>.
- Jugulum, Rajesh. "Importance of Data Quality for Analytics." (2016).
- LEITE Philipe; GEORGE T; SUN C et al. *Social Registries for Social Assistance and Beyond*. World Bank, Washington, DC, (2017). DOI: 10.1596/28284

- Lipsky, M. (2010). Burocracia de nível de rua: dilemas do indivíduo nos serviços públicos.
- LOSHIN, David. *The Practitioner's Guide to Data Quality Improvement*. Elsevier, 2011.
- Marconi Fernandes de Sousa, *Indicadores, Monitoramento e Avaliação de Políticas Públicas*. Enap, jul 2013
- MENEZES, Karina. Conheça países que já adotaram o documento de identidade digital. 2019. <https://blog.idwall.co/conheca-paises-que-ja-estao-adotando-o-documento-de-identidade-digital/>. Acesso em 01 maio 2022.
- MORTONI, Marcos. *Relações Intergovernamentais*. Notas de Aula. ISC, 2021.
- MOSES, Barr; GAVISH, Lior; VORWECK, Molly. *Data Quality Fundamentals*. Orreily, 2022.
- Myers, Dan. *Dimensions of Data Quality Under the Microscope*. Information Management Online Article, 2013. Acessado em 12/05/2022.
- OLIVEIRA NETO, Osvaldo Sotério de. A importância da carteira de identidade única e do sistema AFIS para a realidade brasileira. *Revista QUALIA - a ciência em movimento*. v. 5, n. 1 (2019). <https://revistas.unifan.edu.br/index.php/RevistaICSA/issue/view/34> Acesso em 02 jun 2022.
- Rahm, Erhard & Do, Hong. (2000). *Data Cleaning: Problems and Current Approaches*. IEEE Data Eng. Bull.. 23. 3-13.
- REGONINI, Gloria. *Capire le politiche pubbliche*. Bologna: Il Mulino, 2001 apud SECCHI, 2008.
- SECCHI, Leonardo. *COQ3: um framework de análise de políticas públicas*. 2008 in 32º Encontro Anual da ANPOCS.
- SECCHI, Leonardo. *Políticas públicas: conceitos, categorias de análise, casos práticos*. 2. ed. São Paulo: Cengage Learning, 2013. [VitalSource Bookshelf].
- SECCHI, Leonardo. *Análise de Políticas Públicas: Diagnóstico de problemas, recomendação de soluções*. Cengage Learning Brasil, 2016. [VitalSource Bookshelf].
- SLIBAR, Barbara; ORESKI, Dijana; BEGICEVI, Nina. *Importance of the Open Data Assessment: An Insight Into the (Meta)Data Quality Dimensions*. SAGE Open. April-June 2021: 1 –18. DOI: 10.1177/21582440211023178
- TALBURT, John & ZHOU, Yinle. (2015). *ISO Data Quality Standards for Master Data*. 10.1016/B978-0-12-800537-8.00011-9.
- TCU, *1º RELATÓRIO DE ACOMPANHAMENTO DE DADOS Medidas de resposta à pandemia do Covid-19*. **TC 016.834/2020-8**. **Relator:** Ministro Bruno Dantas.
- TCU, *RELATÓRIO DE ACOMPANHAMENTO INSTITUTO NACIONAL DO SEGURO SOCIAL. Fiscalização contínua de benefícios Previdenciários 2020*. **TC 036.871/2020-6**. **Relator:** Ministro Bruno Dantas

YANES PAVON, Jessica; SEPULVEDA LIMA, Roberto; DIAZ PANDO, Humberto. La evaluación de la calidad de datos: una aproximación criptográfica. *Comp. y Sist.*, Ciudad de México , v. 23, n. 2, p. 557-568, jun. 2019 . Disponível em <[http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S1405-55462019000200557&lng=es&nrm=iso](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-55462019000200557&lng=es&nrm=iso)>. Acesso em 30 abril 2022. Epub 10-Mar-2021. <https://doi.org/10.13053/cys-23-2-2899>.

### **Missão**

Aprimorar a Administração Pública em benefício da sociedade por meio do controle externo

### **Visão**

Ser referência na promoção de uma Administração Pública efetiva, ética, ágil e responsável