

**LUÍZA DA SILVA JAQUES**

**Avaliação da qualidade de dados do Sistema Nacional de  
Registros Cíveis como insumo para a Fiscalização Contínua de  
Benefícios**

**Brasília**

**2020**

**LUÍZA DA SILVA JAQUES**

**Avaliação da qualidade de dados do Sistema Nacional de  
Registros Cíveis como insumo para a Fiscalização Contínua de  
Benefícios**

Trabalho de conclusão do curso de pós-graduação  
*lato sensu* em Análise de Dados para o Controle  
realizado pela Escola Superior do Tribunal de  
Contas da União como requisito para a obtenção do  
título de especialista.

Orientador: Prof. Dr. Eduardo Chaves Ferreira.

**Brasília**

**2020**

## REFERÊNCIA BIBLIOGRÁFICA

JAQUES, Luíza da Silva. **Avaliação da qualidade de dados do Sistema Nacional de Registros Cíveis como insumo para a Fiscalização Contínua de Benefícios**. 2020. Trabalho de Conclusão de Curso (Especialização em análises de dados para o controle) – Escola Superior do Tribunal de Contas da União, Instituto Serzedello Corrêa, Brasília DF.

## CESSÃO DE DIREITOS

NOME DO AUTOR: Luíza da Silva Jaques

TÍTULO: Avaliação da qualidade de dados do Sistema Nacional de Registros Cíveis como insumo para a Fiscalização Contínua de Benefícios.

GRAU/ANO: Especialista/2020

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.

---

Luíza da Silva Jaques

luizasj@tcu.gov.br

### Ficha catalográfica

Jaques, Luíza da Silva Jaques

Avaliação da qualidade de dados do Sistema Nacional de Registros Cíveis como insumo para a Fiscalização Contínua de Benefícios. / Luíza da Silva Jaques; orientador, Eduardo Chaves Ferreira, 2020.

55 p.

Monografia (especialização) – Escola Superior do Tribunal de Contas da União, Curso de Especialização em análises de dados para o controle, Brasília, 2020.

Inclui referências.

1. Análise de dados para o controle. 3. Fiscalização contínua. 4. Sirc. 5. Assistência. 6. Previdência. I. Eduardo Chaves Ferreira. II. Escola Superior do Tribunal de Contas da União. Especialização em análises de dados para o controle. III. Título.

LUÍZA DA SILVA JAQUES

**Avaliação da qualidade de dados do Sistema Nacional de  
Registros Cíveis como insumo para a Fiscalização Contínua de  
Benefícios**

Trabalho de conclusão do curso de pós-graduação lato sensu em análise de dados para o controle realizado pela Escola Superior do Tribunal de Contas da União como requisito para a obtenção do título de especialista.

Brasília, 27 de março de 2020.

**Banca Examinadora:**

---

Prof. Eduardo Chaves Ferreira, Dr.  
Orientador  
Instituto Serzedello Corrêa, TCU

---

Prof. Rodrigo Otávio Coelho Hildebrand.  
Instituto Serzedello Corrêa, TCU

## **AGRADECIMENTOS**

À minha família pelo amor e apoio incondicionais, e pela compreensão quanto à necessidade de aprimoramento profissional aliada à constante vontade de fazer a diferença, ainda que diminuta, no mundo.

Aos amigos e colegas do TCU, especialmente da Diretoria de Análise de Dados e Tecnologia da Informação - DCAD, da Secex Previdência, que me auxiliaram nas dificuldades enfrentadas durante a elaboração deste trabalho final, compartilhando conhecimentos e acolhendo, com paciência, dúvidas sobre o processo de trabalho.

Aos professores do curso que dedicaram seu tempo para planejar as atividades de ensino e orientar os alunos, em especial ao Professor Eduardo Chaves Ferreira, pelo apoio e oportunas contribuições para a melhoria do presente trabalho.

Aos amigos e colegas de pós-graduação, grandes aliados nesta jornada de aprimoramento e enfrentamento dos diversos percalços para concretização deste objetivo.

## RESUMO

Em razão da relevância social e da alta materialidade dos valores envolvidos, o Tribunal de Contas da União (TCU) tem buscado, nos últimos anos, atuar na fiscalização da concessão, da manutenção e do pagamento de benefícios de forma sistêmica e ampla. Esta atuação materializou-se no processo de trabalho conhecido como Fiscalização Contínua de Benefícios (FCB), que utiliza técnicas de estatística, análise de dados e tecnologia da informação para analisar o grande volume de dados objetos da fiscalização em busca de indícios de erros ou fraudes. O processo busca maximizar a efetividade do Controle Externo, acompanhando sistematicamente programas e políticas públicas – como Cadastro Único, Programa Bolsa Família, benefícios previdenciários e benefícios de prestação continuada (BPC) –, por meio de trilhas de auditoria aplicadas a seus bancos de dados. Dado o caráter continuado da FCB, a cada novo ciclo de fiscalização busca-se o aperfeiçoamento do processo por meio da incorporação de informações inexploradas até aquele momento e da utilização de novas ferramentas e técnicas de análise de dados que potencialmente alcancem uma maior acurácia nos resultados obtidos. Para o quinto ciclo de FCB, referente ao exercício de 2019, o TCU obteve dados do Sistema Nacional de Registros Cíveis (Sirc), que possui informações sobre certidões de nascimento, casamento e óbito dos cartórios de pessoas naturais. Esses eventos constituem marcos fundamentais para diversos benefícios, além de contribuir para a qualificação e o enriquecimento dos dados de identificação de cidadãos e de seus vínculos com outras pessoas. O presente trabalho buscou aplicar metodologia de avaliação da qualidade de dados à base do Sirc, a fim de possibilitar sua utilização como insumo para a Fiscalização Contínua de Benefícios.

**Palavras-chave:** Análise de dados para o controle. Fiscalização contínua. Auditoria Contínua. Avaliação da qualidade de dados. Sistema Nacional de Registros Cíveis. Sirc. Assistência Social. Previdência Social.

## ABSTRACT

Due to the social relevance and the high materiality of the amounts involved, the Federal Court of Accounts (TCU - Brazil) has sought, in recent years, to act in a systemic and wide inspection of the concession, maintenance and payment of benefits. This action materialized in the work process known as Continuous Benefit Inspection (FCB), which uses statistical techniques, data analysis and information technology to analyze the large volume of data subject to inspection in search of evidence of errors or fraud. The process seeks to maximize the effectiveness of External Control, systematically accompanying public programs and policies – such as Brazil's Cadastro Único, Bolsa Família program, social security benefits and Brazilian non-contributory pension program (BPC) –, through audit trails applied to its databases. Given the continued nature of the FCB, every new cycle of auditing seeks to improve the process through the incorporation of information that has not been explored up to that moment and the use of new data analysis tools and techniques that potentially achieve greater accuracy of the results obtained. For the fifth FCB cycle, referring to the 2019 financial year, TCU obtained the National System of Civil Registries (Sirc) database, which has information on birth, marriage and death certificates from natural registry offices. These events are fundamental milestones for several benefits, in addition to contributing to the qualification and enrichment of the identification of citizens and their links with other people. The present work sought to apply a data quality assessment methodology to the Sirc database, in order to enable its use as an input for the Continuous Benefit Inspection (FCB).

**Keywords:** Data analysis for control. Continuous inspection. Continuous audit. Data quality assessment. Nacional System of Civil Records. Sirc. Social assistance. Social Security.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>11</b>
<b>2</b>	<b>FISCALIZAÇÃO CONTÍNUA DE BENEFÍCIOS.....</b>	<b>14</b>
2.1	PREVIDÊNCIA SOCIAL .....	15
2.2	ASSISTÊNCIA SOCIAL.....	16
2.3	METODOLOGIA DA FISCALIZAÇÃO CONTÍNUA DE BENEFÍCIOS .....	17
2.3.1	<b>Entendimento do negócio e dos dados.....</b>	<b>17</b>
2.3.2	<b>Obtenção e preparação dos dados.....</b>	<b>18</b>
2.3.3	<b>Elaboração de modelos de análise .....</b>	<b>21</b>
2.3.4	<b>Avaliação dos resultados .....</b>	<b>22</b>
2.3.5	<b>Conclusões e propostas de encaminhamentos .....</b>	<b>22</b>
<b>3</b>	<b>METODOLOGIA DE AVALIAÇÃO DA QUALIDADE DE DADOS.....</b>	<b>23</b>
3.1	COMPLETUDE (OU INTEGRIDADE).....	24
3.2	UNICIDADE (OU IDENTIDADE) .....	25
3.3	VALIDADE.....	26
3.4	CONSISTÊNCIA.....	26
3.5	ACURÁCIA.....	27
<b>4</b>	<b>SISTEMA NACIONAL DE REGISTROS CIVIS (SIRC) .....</b>	<b>27</b>
4.1	ESTRUTURA DE GOVERNANÇA.....	30
4.2	LEGISLAÇÃO .....	30
4.3	ANÁLISE DESCRITIVA DA BASE DE DADOS.....	31
<b>5</b>	<b>AVALIAÇÃO DA QUALIDADE (CREDIBILIDADE) DE DADOS DO SIRC</b>	<b>32</b>
5.1	UNICIDADE E COMPLETUDE DE DADOS DE IDENTIFICAÇÃO.....	33
5.2	VALIDADE, ACURÁCIA E QUALIFICAÇÃO DE DADOS DE IDENTIFICAÇÃO	36
5.3	COMPLETUDE DA BASE DE DADOS.....	39
5.4	VALIDADE DOS TERMOS DE MATRÍCULA DAS CERTIDÕES.....	42
<b>6</b>	<b>APLICAÇÃO DOS DADOS DO SIRC ÀS ANÁLISES DA FCB.....</b>	<b>43</b>



6.1	CREDIBILIDADE DA DOCUMENTAÇÃO DE CERTIDÕES NO CADASTRO ÚNICO	43
6.2	ENRIQUECIMENTO DOS DADOS DE IDENTIFICAÇÃO DO CADASTRO ÚNICO	47
6.3	TIPOLOGIAS DE INDÍCIOS DE FALECIMENTO.....	50
<b>7</b>	<b>CONCLUSÃO .....</b>	<b>52</b>
	<b>REFERÊNCIAS.....</b>	<b>54</b>



## 1 INTRODUÇÃO

Diz-se que os dados são a matéria-prima do século 21: um recurso que permite novos serviços e produtos inimagináveis e que estão apenas esperando para serem explorados (HALONEN, 2012).

O volume dos dados e a capacidade de análise aumentaram consideravelmente nos últimos anos, alimentando rápidos avanços tecnológicos, com algoritmos cada vez mais sofisticados e maior poder computacional, promovendo mudanças disruptivas nos negócios, nos governos e na sociedade em geral.

Nesse cenário de *big data analytics*, apresenta-se como um grande desafio a geração de resultados precisos e de alto valor para pessoas e organizações. De acordo com CAI e ZHU (2015), dados de alta qualidade são condição prévia para analisar e usar big data e garantir o valor dos dados. Da mesma forma, a qualidade dos dados também é essencial para os processos de tomada de decisões das organizações, essencialmente baseados em informações.

O termo qualidade depende do contexto em que é aplicado. Segundo SEBASTIAN-COLEMAN (2012), qualidade de dados é sobre os dados atenderem às expectativas implícitas ou explícitas das pessoas que os usarão. Como alguém julga a qualidade dos dados depende do que essa pessoa espera dos dados. Ainda segundo a autora, as expectativas podem ser complexas e baseiam-se não apenas no que os dados devem representar, mas também no motivo pelo qual uma pessoa precisa dos dados e em como pretende usá-los.

No que diz respeito à prestação de serviços públicos, a enorme quantidade de informações produzidas pelos entes governamentais faz com que seus sistemas administrativos constituam importante fonte de informação sobre a população e suas necessidades, servindo como base para implementação de políticas públicas e também para acompanhamento da efetividade dessas políticas.

Não obstante a existência de diversas iniciativas para aprimorar os dados produzidos e custodiados pelos entes públicos, a necessidade de prestação continuada de serviços essenciais e o elevado volume de dados fazem com que, muitas vezes, a qualidade desses dados seja impactada sem que os detentores da informação consigam detectar, mensurar ou tomar medidas para evitar sua deterioração.

A melhoria da qualidade dos registros administrativos pode ser induzida por contribuições dos usuários ativos desses dados, sejam eles da sociedade civil, da academia, de

órgãos públicos de controle, do próprio órgão detentor da informação, dentre outros atores interessados. Por meio de demandas desses usuários, é possível a identificação de problemas de qualidade dos dados e, potencialmente, a correção dessas inconsistências pelos detentores da informação, aumentando a qualificação das informações.

Nesse sentido, o Tribunal de Contas da União (TCU) tem se inserido como ator relevante nesse processo, realizando diversas ações de controle nos últimos anos com o apoio maciço de análises de dados, que resultaram em deliberações com vistas a aperfeiçoar a prestação de políticas públicas usuárias dessas bases de dados, além de contribuir para a qualidade dos próprios dados.

Exemplo de ação de controle dessa natureza é a Fiscalização Contínua de Benefícios (FCB), realizada desde 2015, pela atual Secretaria de Controle Externo da Gestão Tributária, da Previdência e da Assistência Social (Secex Previdência). Trata-se de acompanhamento anual, apoiado intensamente no uso de ferramentas de tecnologia da informação, que explora bases de dados governamentais internalizadas pelo TCU em busca de situações que indiquem irregularidades ou ineficiências na concessão e no pagamento de benefícios sociais. No exercício de 2018, o acompanhamento fiscalizou mais de R\$ 500 bilhões anuais em pagamentos realizados pelo Governo Federal nas funções de previdência, assistência e trabalho, tendo um benefício de controle esperado de aproximadamente R\$ 2 bilhões.

As principais bases avaliadas nas áreas de assistência e previdência são: o Cadastro Único para programas sociais do Governo Federal (Cadastro Único); a folha de pagamentos do Programa Bolsa Família (PBF); e a folha de pagamentos previdenciários (Maciça) e do benefício de prestação continuada (BPC).

Dado o caráter continuado da FCB, a cada novo ciclo de fiscalização busca-se o aperfeiçoamento do processo por meio da incorporação de outras informações inexploradas até aquele momento e da utilização de novas ferramentas e técnicas de análise de dados que potencialmente alcancem uma maior acurácia nos resultados obtidos.

Para o quinto ciclo de FCB, referente ao exercício de 2019, o TCU obteve dados do Sistema Nacional de Registros Cíveis (Sirc), que possui informações sobre certidões de nascimento, casamento e óbito dos cartórios de pessoas naturais. Esses eventos constituem marcos fundamentais para diversos benefícios previdenciários, assistenciais e trabalhistas, além de contribuírem para a correta identificação dos beneficiários.

O presente trabalho busca incorporar as análises dos dados do Sirc à Fiscalização Contínua de Benefícios, tratando do problema de pesquisa: De que forma os registros cíveis de

nascimento, casamentos e óbito do Sirc podem contribuir no acompanhamento da concessão, da manutenção e do pagamento de benefícios sociais?

A principal motivação para a iniciativa é agregar novas análises ao processo cíclico propiciado pela FCB: o acompanhamento contínuo de cadastros e programas identifica deficiências de qualidade de dados e falhas na execução de políticas; essas constatações geram propostas de melhorias e correções, que potencialmente aprimoram os dados que serão reavaliados em ciclo posterior; estas bases de maior qualidade permitem o desenvolvimento de trilhas de auditoria novas ou com maior acurácia nos resultados.

Dentre os objetivos específicos, pretende-se: i) avaliar a qualidade dos dados do Sirc; ii) avaliar a qualidade dos dados relativos a certidões de bases da assistência social (Cadastro Único); iii) enriquecer a identificação de pessoas beneficiárias de programas sociais; iv) verificar indícios de benefícios concedidos em desacordo com a legislação aplicável utilizando os dados de registro civil.

De suma importância destacar que as situações apresentadas neste estudo de caso são também objeto de relatório de fiscalização do TCU (TC 009.922/2019-9, de relatoria do Ministro Marcos Bemquerer), no qual são tratadas mais profundamente as causas, as legislações aplicáveis e as propostas de encaminhamentos para os problemas identificados, com a finalidade de aprimoramento de políticas públicas e bases de dados envolvidas.

Naturalmente, as situações aqui apresentadas estão sujeitas a apreciação do Ministro relator e do Pleno deste Tribunal. Este trabalho de conclusão de curso focou essencialmente na apresentação do método e dos resultados alcançados utilizando-se a metodologia da Fiscalização Contínua de Benefícios para os dados do Sirc.

Nos capítulos a seguir serão detalhados a base teórica, as análises empreendidas e os resultados alcançados, conforme a seguinte divisão: o Capítulo 2 apresenta a Fiscalização Contínua de Benefícios da Secex Previdência; o Capítulo 3 faz uma compilação de metodologia de avaliação da qualidade de dados; o Capítulo 4 apresenta o sistema Sirc; o Capítulo 5 apresenta os resultados obtidos com a avaliação de qualidade dos dados do Sirc; o Capítulo 6 detalha a aplicação dos dados do Sirc à FCB; e, por fim, o Capítulo 7 traz as conclusões do trabalho.

## 2 FISCALIZAÇÃO CONTÍNUA DE BENEFÍCIOS

O Tribunal de Contas da União iniciou em 2015 a metodologia de fiscalização com uso intensivo de ferramentas e técnicas de análise de dados denominada Fiscalização Contínua de Benefícios – FCB, aplicando-a às funções de governo da Previdência Social, do Trabalho e da Assistência Social (BRASIL, 2019).

Realizada em formato de Relatório de Acompanhamento (Racom) e com periodicidade anual, a FCB tem como objetivo identificar, por meio de cruzamentos sistemáticos de bases de dados, indícios de irregularidades em benefícios concedidos, propondo, quando couber, ações de controle e deliberações que mitiguem o risco de pagamentos indevidos e que busquem aperfeiçoar os controles internos das instituições fiscalizadas.

A FCB encontra-se em seu quinto ciclo de fiscalização pela Secex Previdência, havendo dois processos de controle externo em curso com o objetivo de fiscalizar a concessão, a manutenção e o pagamento de benefícios: TC 009.922/2019-9, relativo a benefícios assistenciais, de relatoria do Min. Marcos Bemquerer; e TC 012.592/2019-6, relativo a benefícios previdenciários, de relatoria do Min. Bruno Dantas. Em 2019, houve a segregação da função Trabalho, que passou a ser fiscalizada por unidade específica, a recém-criada Secretaria de Controle Externo do trabalho e entidades paraestatais (SecexTrabalho) – TC 028.803/2019-1, de relatoria do Min Marcos Bemquerer.

A utilização de análise de dados com uso massivo de Tecnologia da Informação representa uma mudança de paradigma com relação a modelos de auditoria tradicionais, que comumente utilizam métodos de amostragem para examinar um conjunto de dados e reduzir o risco de auditoria. Amostragem em auditoria pode ser descrita como a aplicação de procedimentos em um subconjunto dos dados, denominada amostra, para que seja possível formar conclusão sobre a totalidade dos itens analisados.

A principal motivação para o uso de técnicas de análise de dados na fiscalização de benefícios sociais se dá pela alta materialidade das políticas públicas e pelo elevado volume de dados objetos de fiscalização, que tornariam uma auditoria tradicional utilizando amostragem bastante complexa. Com a aplicação de ferramentas estatísticas e de TI, é possível analisar todo o universo de concessões e pagamentos fiscalizado e não apenas uma amostra.

Quanto à materialidade de recursos e à volumetria dos dados fiscalizados, tomando a função da Assistência Social, em setembro de 2018, o Cadastro Único continha 26,9 milhões de famílias e 74,5 milhões de pessoas em situação de vulnerabilidade. Já as transferências do PBF alcançaram, em média, 13,7 milhões de famílias, com um orçamento de R\$ 29,4 bilhões,

e o BPC contava com cerca de 4,6 milhões de beneficiários e pagamentos que somaram R\$ 55,2 bilhões, no exercício de 2018.

Quanto aos valores de benefícios potenciais, também para a FCB Assistência, em 2018, a exclusão de beneficiários identificados com indícios de irregularidade no Bolsa Família alcançaria R\$ 531 milhões e do Benefício de Prestação Continuada seria de R\$ 670 milhões.

O caráter contínuo e sistêmico da FCB tende a propiciar maior tempestividade e eficácia às ações de controle externo, tornando-as mais próximas da realidade do jurisdicionado e possibilitando que sejam agregados novos procedimentos a cada ciclo de fiscalização, sem que se perca o conhecimento adquirido nos processos anteriores.

A seguir serão apresentadas, de forma sucinta, as funções da Previdência e da Assistência Social, bem como a metodologia de trabalho empregada na FCB.

## 2.1 PREVIDÊNCIA SOCIAL

A Previdência Social brasileira divide-se em regimes previdenciários especializados: o Regime Geral de Previdência Social (RGPS); os Regimes Próprios de Previdência Social (RPPS) – federal, estaduais, distrital e municipais; os Encargos Financeiros da União com militares inativos e pensionistas; e o Regime de Previdência Complementar (RPC), cada um deles definido constitucionalmente para proteger um conjunto de trabalhadores, obrigatória ou facultativamente.

Os trabalhos da FCB relativos à Previdência concentram-se nos benefícios do RGPS, cujas políticas são elaboradas e acompanhadas pela Secretaria de Previdência do Ministério da Economia e executadas pelo Instituto Nacional do Seguro Social (INSS) com o apoio da Empresa de Tecnologia e Informações da Previdência Social (Dataprev). O RGPS tem caráter contributivo e, dentre os seus contribuintes, encontram-se empregadores, empregados assalariados, domésticos, autônomos, contribuintes individuais e trabalhadores rurais.

O TCU conta com acordo de cooperação com o INSS para recebimento da base de dados de pagamentos mensais relativos aos benefícios previdenciários e de prestação continuada (BPC), denominada Maciça. Além disso, os dados relativos às concessões desses benefícios também são recebidos mensalmente e internalizados pelo TCU.

## 2.2 ASSISTÊNCIA SOCIAL

A Assistência Social é direito constitucional de todo o cidadão que dela necessitar, independentemente de contribuição à Seguridade Social. Está organizada por meio do Sistema Único de Assistência Social (Suas) e objetiva garantir a proteção social e apoiar indivíduos e famílias no enfrentamento de dificuldades, por meio de serviços, benefícios e programas.

Um dos principais instrumentos para execução de programas de Assistência Social é o Cadastro Único para Programas Sociais do Governo Federal (Cadastro Único), que possibilita a identificação e a caracterização de famílias de baixa renda, funcionando como porta de entrada para famílias em situação de pobreza acessarem diversas políticas públicas. A inscrição no Cadastro Único é condição para concessão dos benefícios do Programa Bolsa Família (PBF), do Benefício de Prestação Continuada (BPC), do Programa Minha Casa Minha Vida (MCMV), da Tarifa Social de Energia Elétrica, do Criança Feliz, dentre outros.

O Bolsa Família é um programa de transferência de renda que atende famílias que vivem em situação de extrema pobreza – renda *per capita* mensal de até R\$ 85 – ou de pobreza – renda *per capita* mensal de R\$ 85 a R\$ 178 –, desde que tenham crianças ou adolescentes de 0 a 17 anos. O programa paga um benefício básico de R\$ 85 mensais e até cinco benefícios variáveis por família, a depender do atendimento de compromissos nas áreas de educação e saúde, como frequência escolar e vacinação.

Já o Benefício de Prestação Continuada garante o pagamento de um salário mínimo para pessoas com 65 anos ou mais ou com deficiência. Podem participar do programa pessoas cuja renda familiar *per capita* seja inferior a um quarto do salário mínimo vigente. Não é necessário ter contribuído para recebê-lo, porém não há pagamento de pensão e não pode ser acumulado com outros benefícios da Seguridade Social, como aposentadorias e pensões.

Os trabalhos da FCB relativos à Assistência concentram-se nos benefícios do Programa Bolsa Família (PBF) e de Prestação Continuada (BPC), por representarem os maiores gastos na área de Assistência, e nos dados do Cadastro Único, de inscrição obrigatória para recebimento dos principais benefícios assistenciais.

As bases de dados do Cadastro Único e de pagamentos do PBF são obtidas pelo TCU via Ofício de Requisição encaminhado ao órgão gestor, Ministério da Cidadania. Já os dados relativos ao BPC são recebidos pelo TCU em conjunto com os benefícios previdenciários da Maciça, via acordo de cooperação com o INSS.



## 2.3 METODOLOGIA DA FISCALIZAÇÃO CONTÍNUA DE BENEFÍCIOS

A Fiscalização Contínua de Benefícios segue fluxo de trabalho análogo ao da metodologia de mineração de dados conhecida por Crisp-DM, sigla do inglês *Cross-Industry Standard Process for Data Mining* (CHAPMAN *et al*, 2000).

O Crisp-DM sugere processo de trabalho que ataca problemas de mineração de dados progressivamente, constituindo-se das etapas: entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e implantação. Cada etapa do Crisp-DM é subdividida em tarefas, mas sua estrutura de trabalho não é rígida. O fluxo de execução é realizado de forma iterativa, ou seja, em ciclos, que se repetem até que os objetivos da mineração sejam atingidos.

Tal como preceituado pelo Crisp-DM, as etapas de análise de dados da FCB são realizadas de forma circular, partindo-se do entendimento do negócio e dos dados, passando por etapas de recebimento e preparação das bases, até a proposição, avaliação e implantação de modelos de análise que permitam atingir os objetivos da fiscalização.

Evidentemente, por estar incluída em processo de controle externo e sujeita a legislação e ritos próprios, a FCB conta também com atividades típicas de fiscalizações do TCU, como requisições e encaminhamentos aos órgãos gestores de programas e bases avaliados.

Os tópicos a seguir apresentarão as etapas da metodologia de trabalho da FCB, incluindo aspectos típicos de análise de dados e de controle externo.

### 2.3.1 Entendimento do negócio e dos dados

Inicialmente, o entendimento do negócio e dos dados foi alcançado por meio de fiscalizações do tipo levantamento, previstas no art. 238 do Regimento Interno do TCU (BRASIL, 2020), em que se objetivou conhecer os órgãos e programas auditados, bem como a legislação aplicável às informações.

Além disso, a cada novo ciclo da FCB são empreendidos esforços adicionais para expandir a compreensão sobre dados e legislação aplicável, principalmente ao se incluir bases ainda não exploradas como insumos da fiscalização. Esse é precisamente o caso da base de dados do Sirc, adicionada no quinto, e atual, ciclo da FCB e objeto do presente trabalho de pesquisa.

### 2.3.2 Obtenção e preparação dos dados

Após o processo de entendimento, tem início a obtenção das bases de dados junto aos órgãos gestores, que é alcançada por meio de acordos de cooperação ou requisições específicas da equipe de auditoria, após negociação quanto à forma de disponibilização.

A internalização das informações em bancos de dados do Tribunal é feita por equipes especializadas em “Extração, Transformação e Carga”, ou ETL (do inglês *Extract, Transform, Load*), da área de Tecnologia da Informação do TCU. É essencial que as informações estejam disponíveis e sejam carregadas tempestivamente e de maneira continuada, seguindo os procedimentos de segurança da informação adequados.

Em sequência, é realizada a preparação de dados para as análises que serão empreendidas. As principais atividades desta etapa são: avaliação de credibilidade, higienização e enriquecimento para construção de bases derivadas de comparação (identificação) de pessoas.

A credibilidade dos dados consiste em uma série de testes realizados para que se estabeleça o grau de confiabilidade dos bancos de dados recebidos. São avaliados seis atributos: completude, unicidade, validade, consistência, acurácia e uniformidade.

De forma resumida: a completude verifica se existem registros com dados faltantes; a unicidade busca identificar multiplicidade em campos chave que deveriam ser únicos; a validade verifica o respeito a regras definidas em esquemas ou modelos de dados; a consistência confronta variáveis diferentes que deveriam, em tese, trazer a mesma informação e também verifica se as regras condicionais de campos relacionados estão sendo respeitadas; a acurácia confronta os dados com outras fontes de informação; e a uniformidade verifica se campos com informações quantitativas mantêm um mesmo padrão de medidas.

Em ferramentas de mercado, livros e trabalhos acadêmicos, essa atividade é comumente denominada avaliação da qualidade de dados. A diferença de nomenclatura deve-se ao fato de o termo credibilidade incorporar um viés de ceticismo, próprio de trabalhos de auditoria.

Segundo a Norma Brasileira de Contabilidade Técnica de Auditoria 200 (NBC TA 200), o ceticismo profissional é definido como “a postura que inclui uma mente questionadora e alerta para condições que possam indicar possível distorção devido a erro ou fraude e uma avaliação crítica das evidências de auditoria”. Ainda segundo a Norma, o ceticismo profissional é necessário para a avaliação crítica das evidências de auditoria (CFC, 2016).

Dessa forma, a análise de credibilidade empregada pela FCB pode ser vista como uma avaliação da qualidade dos dados fiscalizados, acrescida do ceticismo necessário para execução de trabalhos de auditoria, como os realizados pelo TCU.

A revisão teórica de cada atributo de credibilidade, bem como um detalhamento metodológico mais amplo quanto a critérios de avaliação de qualidade de dados serão apresentados no Capítulo 3.

Seguindo na preparação de dados, um produto essencial da FCB é a elaboração de uma base qualificada para comparação (identificação) de pessoas. Essa atividade utiliza técnicas de higienização e de enriquecimento para produzir uma base com dados de maior qualidade, que será utilizada para análises e cruzamentos de dados na etapa seguinte.

Quanto mais completa e qualificada for a identificação das pessoas registradas nos cadastros para recebimento de benefícios sociais, maior é a possibilidade de se fazer a correta identificação de suas necessidades e vulnerabilidades e, conseqüentemente, implementar ações específicas para essa população.

Da perspectiva do controle, a correta identificação de pessoas permite que sejam implementadas, com grau de acerto mais elevado, trilhas de auditoria para avaliar o cumprimento dos requisitos de políticas públicas. As trilhas de auditoria utilizam o cruzamento de bases de dados de várias origens para determinar se o ato fiscalizado está de acordo com as regras da política.

Por exemplo, é possível verificar se a declaração de renda familiar *per capita* registrada no Cadastro Único está de acordo com outros registros de rendimentos da família e se esses rendimentos reais atendem aos requisitos para pagamento dos benefícios do Programa Bolsa Família ou de Prestação Continuada.

Nesse sentido, o cruzamento de dados de várias origens somente é possível com a utilização de campos que sejam comuns aos dados envolvidos na operação, como número do CPF ou do título de eleitor, denominados chaves. Porém, de nada adianta esse batimento de bases, se outros dados como nome e nome da mãe indicarem se tratar de pessoas diferentes.

Assim, surge a necessidade de se aprimorar a qualidade dos dados, a fim de que sejam minimizados os casos de falsos positivos na etapa de análise. Para esse aprimoramento, são utilizadas técnicas de *Data Quality*, como higienização e enriquecimento de dados.

A higienização (ou saneamento) de dados objetiva remover informações incorretas da base, como dados duplicados, inválidos ou inconsistentes. Já o enriquecimento de dados permite adicionar informações que se encontravam ausentes na base original, utilizando-se outras fontes de dados consideradas mais confiáveis. Percebe-se que essas atividades estão intimamente relacionadas com a avaliação de credibilidade, visto que os resultados desta guiam a preparação de uma base de dados mais qualificada.

Assim, a FCB buscou aprimorar as bases do CadÚnico e da Maciça, por meio do cruzamento com duas das maiores bases cadastrais de pessoas disponíveis no TCU: a base da RFB, em que as pessoas são identificadas pelo CPF, e do TSE, em que as pessoas são identificadas pelo Título de Eleitor (TE). Aproveitou-se que a base da RFB possui informações sobre o Título de eleitor e que a base do TSE possui informações sobre o CPF, para que pudessem ser feitos cruzamentos diretos e também cruzados, utilizando-se destas chaves identificadoras.

Após os cruzamentos das informações, avalia-se o grau de similaridade entre campos identificadores de uma pessoa em cada base, como: nome, nome da mãe, nome do pai, data de nascimento, CPF e Título de eleitor. Essa avaliação gera uma nota de similaridade, que permite classificar os registros referentes a uma pessoa como iguais, distintos, inconclusivos ou nulos.

Para os registros que não obtiveram uma nota adequada na primeira etapa para o número de CPF, é realizada, então, uma etapa complementar de busca de CPFs, visando achar o valor correto para este campo identificador.

Emprega-se a técnica de comparação denominada *Fuzzy String Matching*, que identifica sequências que tenham correspondência aproximada, dado um certo padrão, para buscar na base de CPF da RFB os beneficiários ainda sem CPF identificado. O padrão utilizado foi uma composição do *soundex*<sup>1</sup> dos campos: primeiro nome da pessoa; segundo nome em diante da pessoa; primeiro nome da mãe; segundo nome em diante da mãe. Nesse caso, o uso deste padrão permite associar pessoas com pequenas diferenças de grafia ou abreviações nos nomes, privilegiando o primeiro nome.

Dessa forma, a maior parte dos beneficiários pode ser corretamente identificada a partir da comparação com outras bases de dados pelos campos chave ou pela busca ativa nessas bases utilizando padrões aproximados. As bases qualificadas de comparação de pessoas do CadÚnico

---

<sup>1</sup> Função que converte uma cadeia de caracteres alfanumérica em código de quatro caracteres que se baseia no som da cadeia de caracteres quando ela é falada.

e da Maciça são, então, consolidadas e servirão de insumo para a etapa seguinte, de construção de modelos para análise dos dados.

Importa destacar que a metodologia e os resultados dessas etapas de credibilidade e de qualificação de bases são enviados para avaliação dos gestores, com vistas ao aprimoramento das informações cadastrais e de seus controles internos.

### **2.3.3 Elaboração de modelos de análise**

A etapa de elaboração de modelos consolida diferentes atividades de análise de dados, como o desenvolvimento de tipologias, a identificação de padrões de regularidade e irregularidade e a construção de indicadores (gerenciais, de negócio, de projeto e de risco). As atividades podem ser executadas independentemente das demais, de forma paralela.

As tipologias (ou trilhas de auditoria) correspondem a cruzamentos de bases de dados e filtragens específicas para verificar se a legislação pertinente ao tema fiscalizado está sendo observada pelos responsáveis pela política pública.

Na evolução da FCB, grande parte do esforço aplicado tem sido na elaboração de tipologias, que visam identificar possíveis irregularidades. Entretanto, vale ressaltar que os resultados das tipologias constituem indícios que, por si só, não podem ser considerados evidências, sendo essencial a realização de verificações complementares para comprovar a irregularidade.

Na operacionalização dessa fase, além dos critérios de materialidade, relevância e risco, deve-se considerar o acesso às informações, com a tempestividade necessária.

Nesse sentido, a FCB é beneficiada pela existência de solução no Tribunal denominada Laboratório de Informações de Controle (LabContas). O LabContas constitui um ambiente colaborativo de análise de dados, mantido por equipes especializadas de TI e que apoia o Controle Externo por meio da disponibilização de bases de dados e ferramentas de análise.

Diversas são as bases auxiliares utilizadas nas tipologias da FCB, podendo-se citar, por exemplo: a Relação Anual de Informações Sociais (Rais); o Sistema Integrado de Administração de Recursos Humanos (Siape); o Sistema Integrado de Administração Financeira do Governo Federal (Siafi); o Cadastro Nacional de Pessoa Jurídica (CNPJ); e o Sistema Informatizado de Controle de Óbitos (Sisobi).

Outra atividade importante é a identificação de padrões de regularidade e irregularidade. Essa atividade objetiva extrair padrões e conhecimento de grandes bases de dados, utilizando

alguma técnica de mineração de dados, de forma a maximizar a atuação do Tribunal na detecção de benefícios que trazem alta probabilidade de serem irregulares. A análise de padrões pode, por si só, indicar possíveis irregularidades, como pode também constituir insumo para desenvolvimento de novas tipologias.

A construção de indicadores possibilita o acompanhamento das políticas públicas, bem como dos trabalhos da FCB, sob diversos aspectos. Os indicadores gerenciais focam na exploração e conhecimento das bases de dados, subsidiando o planejamento e a execução de fiscalizações. Os indicadores de projeto propiciam o acompanhamento da evolução da Fiscalização Contínua, registrando o fluxo de trabalho e os resultados obtidos a cada ciclo. Já os indicadores de risco fornecem uma métrica para priorização da atuação do Tribunal e dos jurisdicionados.

#### **2.3.4 Avaliação dos resultados**

A etapa de avaliação analisa os resultados obtidos na FCB, validando ou não suas conclusões prévias, de forma a minimizar a ocorrência de falsos positivos, isto é, casos em que um benefício regular é apontado como irregular, e falsos negativos, que significariam casos irregulares não identificados.

A avaliação consiste em reavaliar os critérios e analisar amostras de benefícios com indícios de irregularidade identificados nas etapas anteriores com o objetivo de quantificar a acurácia de tipologias e modelos. Idealmente, deve-se realizar a avaliação em conjunto com o gestor da política, uma vez que este possui um conhecimento das regras de negócio muito mais consolidado que a equipe de auditoria.

Os resultados dessa fase retroalimentam a fase de análise, agregando novas rotinas às já estabelecidas e melhorando o poder de acerto dos modelos empregados pela equipe de Fiscalização Contínua. Os acertos reforçam as análises, enquanto os erros permitem que sejam realizados refinamentos até que se obtenha modelo de análises mais acertado.

#### **2.3.5 Conclusões e propostas de encaminhamentos**

Nessa etapa, os produtos das fases anteriores são consolidados e traduzidos em ações do TCU, denominadas propostas de encaminhamentos. Essas propostas de encaminhamento podem ser internas, como processos para disseminação de conhecimento ou abertura de auditorias específicas, ou externas, como recomendações e determinações aos gestores para melhoria de controles e revisões de benefícios.

Exemplificando, listas de indícios de irregularidade podem resultar em determinações para revisão de benefícios no relatório de Acompanhamento; indicadores gerenciais negativos em determinada região ou espécie de benefício podem indicar a necessidade de modificação no planejamento da unidade no período subsequente; riscos pontuais podem ensejar a abertura de representações ou auditorias específicas, dentre outras possibilidades.

### **3 METODOLOGIA DE AVALIAÇÃO DA QUALIDADE DE DADOS**

Os dados são impactados por vários processos ao longo de seu ciclo de vida, a maioria dos quais podem afetar sua credibilidade.

MAYDANCHIK (2007) classifica os processos que podem causar problemas nos dados em três categorias: processos que internalizam dados externos; processos que manipulam bases de dados internas; e processos que tornam as informações imprecisas com o passar do tempo sem que sejam feitas alterações diretas nos dados.

A primeira categoria descrita diz respeito a processos de internalização de dados. Em alguns casos, os dados externos podem já estar imprecisos em sua origem, com o processo de internalização replicando os erros. Em outros, os erros podem ter sido introduzidos no processo de extração, transformação e carga (*extration, transformation and loading* - ETL). Elevados volumes de dados tenderiam a aumentar os problemas, segundo o autor.

Os processos internos de manipulação de dados constituem a segunda categoria, que inclui processos rotineiros, atualizações de sistema e de dados e atividades *ad-hoc*. Boa parte desses procedimentos careceriam de tempo, recursos e metadados confiáveis para o entendimento dos impactos na qualidade, levando, portanto, a inúmeros problemas de dados.

A terceira categoria engloba processos que tornam dados imprecisos no decorrer do tempo, sem qualquer transformação física aplicada a eles. Esse impacto na qualidade poderia acontecer quando o objeto do mundo real descrito pelos dados é alterado, sem ser capturado pelos processos de coleta, tornando os dados armazenados obsoletos e incorretos.

Diversos estudos propõem modelos para avaliar a qualidade de dados. A maioria deles elenca um conjunto de atributos ou características que representariam dados de qualidade, como: acurácia, atualidade, precisão, confiabilidade, atualização, integridade, relevância, acessibilidade, interpretabilidade, completude, validade, consistência, credibilidade, disponibilidade, correção, unicidade, dentre outros (WANG et al., 1996; SEBASTIAN-COLEMAN, 2012; SVOLBA, 2012; PIPINO et al., 2002; HOBBERMAN, 2015).

Segundo SEBASTIAN-COLEMAN (2012), a avaliação da qualidade dos dados pode ser definida por dois fatores relacionados: quão bem os dados atendem a expectativas e propósitos de seus usuários e quão bem eles representam os objetos, eventos e conceitos que são criados para representar. Assim, a escolha de critérios para avaliação de qualidade de dados pode ser tida como um processo subjetivo, que leva em conta os objetivos da análise e o juízo do usuário da informação.

No âmbito de fiscalizações do Tribunal, para que as conclusões obtidas por meio de análise de dados sejam úteis e precisas, é essencial que se utilize uma metodologia para avaliar a qualidade dessas bases de dados, bem como os instrumentos e controles utilizados para coleta dessas informações, sob pena de invalidar os resultados obtidos.

A avaliação de credibilidade da Fiscalização Contínua de Benefícios (FCB) do TCU utiliza um conjunto de critérios para verificar se as bases de dados utilizadas são críveis e passíveis de confronto com outras fontes de informação. O termo qualidade é o mais utilizado na literatura e no mercado para se referir ao macroprocesso de avaliação de bases de dados, enquanto o termo credibilidade utilizado pela FCB abrange também o ceticismo próprio de trabalhos de auditoria. Para fins deste trabalho, os dois termos são utilizados de forma similar.

A seguir será apresentado um conjunto possível de atributos para avaliação da qualidade de bases de dados, alguns deles selecionados na metodologia de trabalho da Fiscalização Contínua de Benefícios, do TCU.

### 3.1 COMPLETUDE (OU INTEGRIDADE)

De acordo com REDMAN (1996), a completude (do inglês *completeness*) refere-se ao grau em que os valores estão presentes em um conjunto de dados. Os valores podem estar ausentes por diferentes motivos, alguns esperados, outros não. O desafio da completude é saber se os valores devem ou não estar presentes (SEBASTIAN-COLEMAN, 2012).

SVOLBA (2015) define a completude de dados como a métrica de se a informação que deveria existir na base de dados existe para todas as observações.

Assim, em uma primeira perspectiva, a completude pode ser mensurada pela proporção entre os campos preenchidos e os não preenchidos em uma tabela, isto é, se existem campos que deveriam estar preenchidos e que não estão. Deve-se levar em conta que, em muitos casos, os dados podem estar registrados com valores padrão, como “ignorado” ou “00000”, o que pode representar uma incompletude do ponto de vista de negócios.

Ainda segundo SVOLBA (2015), a completude não é relevante apenas para tabelas isoladas em que o número de valores ausentes em uma coluna é contado; também é relevante



para comparações entre tabelas. Os dados podem estar incompletos ainda que nas tabelas respectivas não ocorrer nenhum valor ausente.

Nesse sentido, a completude pode ser enxergada também com relação ao universo esperado de dados, ou seja, se todas as observações que deveriam ser registradas estão no banco de dados. Exemplos de incompletudes desse tipo são meses faltantes em bases de séries temporais ou dados hierárquicos presentes na tabela “pai” e ausentes da tabela “filha”.

Em uma terceira perspectiva, a completude pode ser vista com relação a requisitos e metadados do modelo de dados. Conforme HOBBERMAN (2015), a integridade dos requisitos significa que todos os requisitos solicitados aparecem no modelo de dados e nada a mais. Já a integridade dos metadados significa que todas as informações descritivas que envolvem o modelo são capturadas, como formato, restrições e dicionário de campos.

Para SEBASTIAN-COLEMAN (2012), em resumo, um conjunto de dados completo deve atender a pelo menos três condições: deve ser definido para incluir todos os atributos desejados (largura); deve conter a quantidade desejada de dados (profundidade); e seus atributos devem ser preenchidos na extensão desejada (densidade).

### 3.2 UNICIDADE (OU IDENTIDADE)

Conforme MAYDANCHIK (2007), as chaves de identidade garantem que toda entidade do mundo real (pessoa, coisa, local, conceito ou evento) representada nos dados seja identificável de forma única e possa ser distinguida de todas as outras entidades do mesmo tipo. Elas fornecem uma ligação fundamental entre os dados e os objetos do mundo real.

Ainda segundo o autor, uma regra de identidade é uma regra de qualidade de dados, que valida que todo registro em uma tabela de base de dados corresponde a uma e apenas uma entidade no mundo real e que não existam dois registros que referenciem a mesma entidade.

Usualmente, o mecanismo utilizado para promover a identidade de registros em bases relacionais são as chaves primárias, compostas por uma coluna ou uma combinação de colunas cujos valores distinguem uma linha das demais dentro de uma tabela (HEUSER, 2009).

A existência de chave primária por si só, no entanto, não garante a unicidade, pois em muitos casos são criadas chaves substitutas (*surrogate keys*) geradas por computador para identificar cada registro em uma tabela. Nesse caso, pode ser necessário avaliar outras informações para identificar precisamente aquela entidade. Essas informações, entretanto,

podem não ser precisas e estão sujeitas a inconsistências como erros de digitação e de grafia, por exemplo.

A unicidade é um dos principais atributos de qualidade tratados nos trabalhos da FCB, sendo primordial para o processo de identificação de pessoas. Muitos são os números de identidade existentes nas bases governamentais brasileiras, como: certidão de nascimento, cadastro de pessoa física (CPF), título de eleitor, passaporte, identidades civis e militares, número de identificação do trabalhador (NIT), número de identificação social (NIS), carteira de trabalho e previdência social (CTPS), dentre outros.

### 3.3 VALIDADE

SEBASTIAN-COLEMAN (2012) define a validade como a conformidade dos valores dos dados a um domínio ou a uma regra de negócio. Pode ser medida diretamente de dentro de um conjunto de dados, não requerendo comparação com objetos do mundo real para confirmação, diferente da acurácia.

Segundo MAYDANCHIK (2007), as regras de qualidade de dados usadas para validar valores de atributos são comumente referidas como restrições de domínio de atributos. Essas restrições limitam os valores permitidos dos atributos de dados. Por exemplo, o domínio de um atributo pode aceitar ou não valores nulos, bem como é possível que especifique uma lista de valores, caracteres válidos ou padrão esperado.

O autor afirma ainda que, apesar da aparente trivialidade, as restrições de domínio de atributo raramente são projetadas corretamente porque geralmente são baseadas em documentação incorreta ou incompleta. As restrições do domínio de atributo poderiam ser deduzidas da análise de metadados, como modelos de dados, dicionários de dados, tabelas de pesquisa e perfis de atributos.

### 3.4 CONSISTÊNCIA

REDMAN (1996) descreve como característica da consistência o fato de que duas coisas sendo comparadas não conflitem. Os problemas de consistência aparecem quando os conjuntos de dados se sobrepõem e representam os mesmos conceitos ou conceitos semelhantes de uma maneira diferente ou quando seu conteúdo específico não corresponde (SEBASTIAN-COLEMAN, 2012).

Conforme esclarece DATE (2004), o sistema não pode garantir a verdade, apenas a consistência. Ou seja, o sistema não pode garantir que o banco de dados contém apenas proposições verdadeiras, o que ele pode fazer é garantir que o banco não contém algo que cause

violação de restrições de integridade ou incongruências. Dessa forma, a consistência também se concentra em análises internas à base de dados, não confrontando os dados com objetos do mundo real.

No âmbito da FCB, a consistência confronta variáveis diferentes que deveriam, em tese, trazer a mesma informação, além de verificar se regras condicionais de campos relacionados estão sendo respeitadas. Exemplificando, campos de data de nascimento e idade de uma pessoa, que devem trazer informações consistentes entre si, podem ser confrontados.

### 3.5 ACURÁCIA

ENGLISH (1999) define acurácia como o grau que os dados refletem com precisão o objeto ou evento do mundo real que está sendo descrito. Assim, medir a acurácia requer a comparação com o objeto real que os dados representam. Na maioria das situações, essa medição é proibitivamente cara (equivale a coleta de dados). No entanto, um grau de acurácia pode ser determinado através da comparação com dados contidos em uma fonte original (SEBASTIAN-COLEMAN, 2012).

Portanto, uma forma usual de medir a acurácia é por meio do confronto com dados de fontes de informação externas, especialmente aquelas que sejam a fonte primária dos dados. Em casos de conflito entre registros que deveriam trazer a mesma informação, elege-se como correta a base de maior confiança, normalmente a fonte originária da informação.

No contexto da FCB, uma aplicação da avaliação de acurácia consiste em confrontar as informações de pessoas constantes nas bases de pagamentos dos benefícios com outras bases de identificação de pessoas, como CPF e Título de eleitor. Esse procedimento é fundamental para mitigar a ocorrência de falsos positivos nos cruzamentos das tipologias, isto é, apontar um pagamento dentro das regras como irregular.

## 4 SISTEMA NACIONAL DE REGISTROS CIVIS (SIRC)

Os registros civis públicos têm como finalidade assegurar a autenticidade, a eficácia e a publicidade de atos jurídicos que dizem respeito à vida civil dos cidadãos, registrando informações sobre os principais fatos na vida de uma pessoa, como nascimento, casamentos e óbito, bem como informações básicas que caracterizam essa pessoa, como nome, sobrenome, filiação, estado civil e data de nascimento.

Conforme LEHMKUHL (2017) os registros civis no Brasil são públicos e seu acesso está disponível a todos, podendo servir para diferentes fins, tanto para o cidadão e seus dependentes, quanto para o governo em pesquisas em distintas áreas do conhecimento, como na elaboração de censos relativos à população.

As informações contidas nos registros civis constituem insumos para planejamento e avaliação de políticas públicas voltadas às necessidades da comunidade. O número de nascimentos ocorrido em determinado período e localidade permite, por exemplo, a identificação de população alvo para campanhas de vacinação, a elaboração de indicadores nas áreas de educação e saúde e a identificação de populações vulneráveis para ações de assistência social.

Não obstante existirem há mais de 100 anos, os registros civis de pessoas naturais no Brasil não contavam com um sistema público centralizado até 2014, ano em que foi instituído o Sistema Nacional de Registros Civis (Sirc) por meio do Decreto 8.270, de 26 de julho de 2014, substituído pelo Decreto 9.929/2019, em 22 de julho de 2019.

Segundo o portal oficial do Sirc (BRASIL, 2020a), o sistema tem por finalidade modernizar a captação e o tratamento dos dados dos registros civis de nascimento, casamento, óbito e natimortos, passando a realizar essas atividades com o apoio de uma plataforma digital que conecta os cartórios a ambientes de governo eletrônico brasileiros. Dentre os objetivos elencados pelo portal, está o de reduzir o sub-registro, promovendo o efetivo exercício da cidadania.

A integração do Sirc com dados de outras bases cadastrais de pessoas naturais pode constituir mais um passo para a identificação unívoca do cidadão, melhorando a prestação de serviços pelos entes governamentais e reduzindo problemas com homônimos, isto é, pessoas com mesmo nome ou muito parecido, que diversas vezes passam por transtornos de serem confundidas com outra pessoa.

Essa comunicação entre as bases cadastrais permitiria acompanhar cada cidadão do nascimento ao óbito, possibilitando uma maior celeridade na concessão de benefícios e na verificação de pagamentos indevidos, por exemplo. Os dados de certidões dos cartórios do Sirc poderiam, assim, ser integrados com bases de serviços hospitalares, do Cadastro de pessoa física (CPF), do Título de eleitor, do Cadastro Único para programas sociais, de passaportes, de carteiras de identidade, dentre outras bases governamentais.

Diversas iniciativas foram propostas ao longo dos anos para compartilhamento e interoperabilidade de bases cadastrais da administração pública, tendo como pano de fundo a necessidade de identificação unívoca, completa e centralizada de cidadãos.

Dentre esses projetos, encontra-se a Identificação Civil Nacional (ICN), criada pela Lei 13.444/2017, que integra a base de dados biométricos da Justiça Eleitoral, a base de registros civis do Sirc e da Central Nacional de Informações de Registro Civil (CRC Nacional), e outras informações contidas em bases de dados de identificação de estados e Distrito Federal, do Instituto Nacional de Identificação ou de outros órgãos.

Outra iniciativa foi a criação, em 9 de outubro de 2019, do Cadastro Base do Cidadão, que servirá como base de referência sobre cidadãos para o Poder Executivo Federal. O Decreto 10.046/2019 elenca como finalidade do Cadastro, dentre outras, viabilizar um meio unificado de identificação do cidadão para prestação de serviços públicos, suportado por soluções tecnológicas interoperáveis, aumentando a confiabilidade dos cadastros por meio de mecanismos de manutenção da integridade das bases de dados para torná-las qualificadas e consistentes.

Tais iniciativas de identificação única de pessoas estão compreendidas em um contexto em que corporações e governos estão cada vez mais em busca de soluções de *Master Data Management* (MDM), que viabilizem o cruzamento de informações de forma qualificada.

LOSHIN (2010) conceitua objetos de dados mestre (*master data objects*) como os principais objetos de negócio usados nos diferentes aplicativos da organização, ou seja, as coisas que mais importam em uma organização. Conforme BERSON e DUBOV (2010), o foco do MDM está na necessidade de limpar “coisas antigas” e criar um conjunto preciso, oportuno e completo de entidades e relacionamentos críticos para gerenciar e expandir os negócios.

Claramente, os dados de cidadãos estão entre as informações mais importantes no contexto de prestação de serviços públicos, sendo, portanto, objetos em potencial de projetos de MDM de entidades e órgãos da administração pública.

Neste contexto, a base de dados do Sirc é um relevante insumo para um projeto de MDM de pessoas naturais, tendo em vista registrar eventos chave na vida de uma pessoa e seus documentos de identificação. Além disso, também pode constituir peça chave para identificação dos vínculos associados àquela pessoa, essenciais para diversas atividades de auditoria, como análise de redes de corrupção, por exemplo. Para que tais análises atinjam os objetivos

esperados, no entanto, é necessário que as informações contidas nas bases de dados sejam fidedignas e críveis, com cobertura ampla dos eventos previstos e consistentes com outras bases de informação existentes. A avaliação de credibilidade da base de dados do Sirc será objeto do Capítulo 5 deste trabalho.

A seguir serão apresentadas estrutura de governança, legislação e análise descritiva da base de dados do Sirc.

#### 4.1 ESTRUTURA DE GOVERNANÇA

Conforme o Decreto 9.929/2019, o estabelecimento de diretrizes para o funcionamento, a gestão e a disseminação do Sirc e pelo monitoramento do uso dos dados nele contidos cabe ao Comitê Gestor do Sirc (CGSirc), composto por representantes das seguintes entidades: Ministério da Justiça e Segurança Pública; Ministério da Defesa; Ministério das Relações Exteriores; Ministério da Economia; Ministério da Cidadania; Ministério da Saúde; Ministério da Mulher, Família e Direitos Humanos; Instituto Nacional do Seguro Social (INSS); e Instituto Brasileiro de Geografia e Estatística (IBGE). A coordenação do CGSirc é exercida de forma alternada, anualmente, pelo Ministério da Economia e pelo Ministério da Mulher, Família e Direitos Humanos.

O INSS é o responsável pelo desenvolvimento, a operacionalização e a manutenção do SIRC, observadas as diretrizes e as deliberações do CGSirc. Assim como outros sistemas do INSS, a solução de Tecnologia de Informação relativa ao Sirc é mantida pela Dataprev, empresa pública vinculada ao Ministério da Economia.

#### 4.2 LEGISLAÇÃO

Os Decretos de criação do Sirc regulamentaram o disposto nos artigos 37 a 41 da Lei 11.977, de 7 de julho de 2009, que determinaram que os atos relativos aos Registros Públicos praticados a partir da vigência da Lei 6.015, de 31 de dezembro de 1973, deveriam ser armazenados em sistema eletrônico no prazo de 5 anos. A Lei previu, ainda, que os atos e os documentos anteriores a esta data também deveriam ser inseridos em sistema eletrônico, mas sem estipulação de prazo.

Segundo o art. 68 da Lei 8.212, de 24 de julho de 1991 (com redação alterada pela Lei 13.846, de 18 de junho de 2019), o titular do cartório de registro civil de pessoas naturais deve remeter ao INSS, em até um dia útil, pelo Sirc ou outro meio que o substitua, a relação de nascimentos, natimortos, casamentos, óbitos, averbações, anotações e retificações registradas

na serventia. Para municípios sem acesso à internet, há previsão de remessa de relação em até cinco dias úteis.

Para os registros de nascimento e natimorto, conforme o mesmo dispositivo, constarão das informações, obrigatoriamente, a inscrição no Cadastro de Pessoas Físicas (CPF), o sexo, a data e o local de nascimento do registrado, bem como o nome completo, o sexo, a data de nascimento, o local de nascimento e a inscrição no CPF da filiação do registrado.

Já para os registros de casamento e óbito, constarão das informações, obrigatoriamente, a inscrição no CPF, o sexo, a data e o local de nascimento do registrado, bem como, caso disponíveis: número do cadastro perante o Programa de Integração Social (PIS) ou o Programa de Formação do Patrimônio do Servidor Público (Pasep); Número de Identificação do Trabalhador (NIT); número de benefício previdenciário ou assistencial, se a pessoa falecida for titular de qualquer benefício pago pelo INSS; número de registro da carteira de identidade e respectivo órgão emissor; número do título de eleitor; número e série da Carteira de Trabalho e Previdência Social (CTPS).

Cabe notar que o conceito de filiação trazido pelo Sirc está alinhado às novas estruturas parentais reconhecidas pela doutrina e jurisprudência, como a multiparentalidade e a parentalidade socioafetiva (SALOMÃO, 2018).

Conforme Provimento CNJ 63/2017, alterado pelo Provimento 83/2019, é autorizado o reconhecimento voluntário da paternidade ou da maternidade socioafetiva de pessoas acima de 12 anos perante os oficiais de registro civil das pessoas naturais. A norma autoriza, ainda, em seu artigo 14, o reconhecimento extrajudicial de paternidade e maternidade socioafetivos, ainda que existam pai e mãe biológicos, sendo possível ter até dois pais e duas mães no campo de filiação registrado em cartório.

Com relação às declarações de nascido vivo (DNV), a Lei 12.662/2012, de 5 de junho de 2012, assegurou a interoperabilidade com o Sirc em seu art. 5º, estabelecendo que os dados colhidos serão consolidados em sistema de informação do Ministério da Saúde.

Conforme Recomendação 40 do CNJ, de 2 de julho de 2019, cabe às Corregedorias de Justiça Estaduais fiscalizar as atividades das serventias de registros civis quanto ao encaminhamento dos dados ao Sirc, tanto em relação ao cumprimento dos prazos fixados em lei, quanto ao integral fornecimento das informações disponíveis no registro. Essa fiscalização poderia acarretar diversas penalidades, como multa e até perda da delegação da serventia.

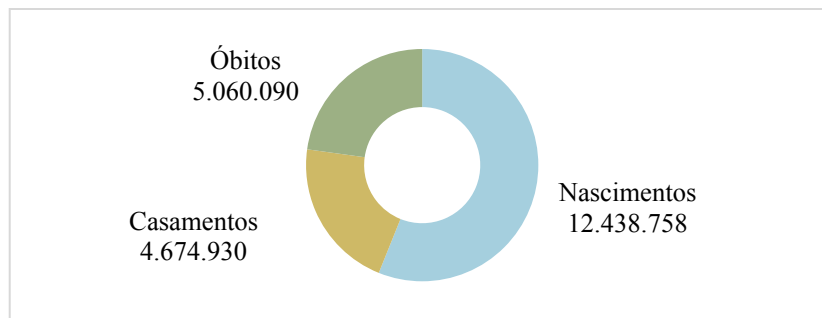
### 4.3 ANÁLISE DESCRITIVA DA BASE DE DADOS

No curso da Fiscalização Contínua de Benefícios Assistenciais de 2019, o INSS encaminhou ao TCU a base de dados do Sirc, atualizada até 1º de agosto de 2019, em resposta ao Ofício de Requisição 02-116/2019.

Cabe mencionar que, segundo Ofício CGGPB/DIRBEN/INSS 383/2019, o acesso aos dados do Sisobi passará a ser operacionalizado por meio do acesso ao Sirc, que poderá incluir também as informações de nascimentos e casamentos. As tratativas para inclusão da base do Sirc no Acordo de Cooperação Técnica com o INSS estão em curso e terão o condão de permitir o recebimento periódico dos dados e a disponibilização para outras áreas do Tribunal.

Para fins de análise, pode-se dividir a base em três grandes blocos de informações, apresentados no Gráfico 1: nascimentos, casamentos e óbitos.

Gráfico 1: Distribuição das certidões no Sirc por tipo, em 1º de agosto de 2019



Destaca-se que o modelo de dados do sistema Sirc, amparado por sua legislação de referência, prevê a incorporação de informações com elevado potencial de contribuir para a identificação unívoca de pessoas e de vínculos, podendo-se citar:

- a) a emissão de CPF em conjunto com a emissão da certidão de nascimento, de forma mais tempestiva, no próprio cartório de registro de pessoas, com campos para composição do termo de matrícula unificado e registro do número da declaração de nascido vivo (DNV);
- b) a identificação das filiações de nascimento – inclusive multiparentais e socioafetivas – com respectivos números de CPF;
- c) os vínculos de casamentos, com CPF e filiações dos cônjuges, bem como alterações de nomes oriundos de casamentos e divórcios; e
- d) a obrigatoriedade de inclusão do CPF na certidão de óbito, com envio em até 5 dias úteis pelos cartórios.

Assim, dado o potencial das informações, foram realizadas análises exploratórias, testes de credibilidade e avaliação da aplicabilidade dos dados do Sirc à Fiscalização Contínua de Benefícios. Os resultados pertinentes serão detalhados nos capítulos a seguir.



## 5 AVALIAÇÃO DA QUALIDADE (CREDIBILIDADE) DE DADOS DO SIRC

O presente capítulo apresenta a avaliação da qualidade dos dados da base do Sistema Nacional de Registros Cíveis (Sirc) com o objetivo de aplicá-los em procedimentos da Fiscalização Contínua de Benefícios (FCB) da Assistência Social. Foram utilizados os critérios de unicidade, completude, validade e acurácia da metodologia de avaliação de qualidade de dados descrita no Capítulo 3.

Os atributos de qualidade podem ser avaliados para a totalidade dos campos e tabelas contidos em um banco de dados ou para um subconjunto destes, considerados mais importantes pelo usuário da informação.

Tendo em vista o potencial do Sirc em contribuir para a identificação de pessoas e de eventos que constituem marcos importantes nas vidas destas pessoas, a avaliação de qualidade teve enfoque principalmente nos campos com informações de documentação, como CPF e matrícula da certidão, ou que tipicamente são capazes de caracterizar uma pessoa, como nome, nome das filiações e data de nascimento.

Em favor de um relato mais conciso, somente são apresentados os resultados da avaliação de qualidade mais interessantes para o estudo de caso em tela.

### 5.1 UNICIDADE E COMPLETUDE DE DADOS DE IDENTIFICAÇÃO

No modelo de dados do Sirc, as tabelas contêm chaves artificiais que garantem fisicamente a propriedade de unicidade de chave primária. Por exemplo, a coluna ID\_CERTIDAO\_OBITO é um campo numérico sequencial que identifica cada registro da tabela de certidões de óbito. Esse campo, no entanto, não tem nenhum significado no mundo real e não garante a unicidade dos eventos representados, que deve ser verificada por meio de outros campos e atributos que identifiquem estes eventos naturalmente.

Conforme a Lei 8.212/1991, que dispõe sobre a organização da Seguridade Social, institui Plano de Custeio, e dá outras providências:

**Art. 68. O Titular do Cartório de Registro Civil de Pessoas Naturais remeterá ao INSS, em até 1 (um) dia útil, pelo Sistema Nacional de Informações de Registro Civil (Sirc) ou por outro meio que venha a substituí-lo, a relação dos nascimentos, dos natimortos, dos casamentos, dos óbitos, das averbações, das anotações e das retificações registradas na serventia. (...)**

**§ 2º Para os registros de nascimento e de natimorto, constarão das informações, obrigatoriamente, a inscrição no Cadastro de Pessoas Físicas (CPF), o sexo, a data e o**

**local de nascimento do registrado, bem como o nome completo, o sexo, a data e o local de nascimento e a inscrição no CPF da filiação. (...)**

§ 3º Para os registros de **casamento** e de **óbito**, constarão das informações, obrigatoriamente, a **inscrição no CPF, o sexo, a data e o local de nascimento do registrado**, bem como, acaso disponíveis, os seguintes dados: (...)

§ 5º O descumprimento de qualquer obrigação imposta neste artigo e o fornecimento de informação inexata sujeitarão o Titular do Cartório de Registro Civil de Pessoas Naturais, **além de outras penalidades previstas, à penalidade prevista no art. 92 desta Lei e à ação regressiva proposta pelo INSS, em razão dos danos sofridos.**

(grifos próprios)

Como se verifica, há previsão legal de registro pelo Sirc dos dados que caracterizam a pessoa registrada na certidão, incluindo CPF, nome completo, data de nascimento e dados das filiações. Essas informações são essenciais para que haja a correta identificação do indivíduo.

Assim, com relação aos critérios de unicidade e completude, para os dados de nascimento, foram examinados os campos de matrícula da certidão, CPF, número da declaração de nascido vivo (DNV) e uma combinação de nome do registrado, data de nascimento e nomes das filiações. A Tabela 1 apresenta os registros de certidões de nascimento com campos ausentes ou em multiplicidade.

Tabela 1: Resultado da verificação de unicidade e completude de certidões de nascimento

<b>Campo chave</b>	<b>Ausentes</b>	<b>%</b>	<b>Em multiplicidade</b>	<b>%</b>
Matrícula da certidão	0	0,0	0	0,0
Cadastro de pessoa física (CPF)	9.748.138	78,4	5.472	0,0
Declaração de nascido vivo (DNV)	3.113.206	25,0	141.867	1,1
Nome, data de nascimento e nomes das filiações	-	-	16.912	0,1

Fonte: elaboração própria.

Apesar de percentualmente baixos, os casos de utilização de CPF e DNV em multiplicidade denotam uma insuficiência nos controles do Sirc e dos sistemas que o alimentam. A possibilidade de emissão de mais de uma certidão com a mesma DNV ou com CPF duplicado pode permitir, em casos extremos, até mesmo o cometimento de crimes e fraudes.

Em matéria veiculada em 10 de novembro de 2019, o programa Fantástico apresentou investigação a respeito de crianças africanas que deixaram o Rio com documentos brasileiros. Segundo a notícia (BRASIL, 2019b):

O esquema usava o primeiro documento de todo brasileiro: a declaração de nascido vivo, ou DNV - um formulário do Ministério da Saúde que já vem numerado. Nesse documento constam informações sobre a nacionalidade, local de nascimento, a data de nascimento e nome dos pais. Mas o nome do pai não é obrigatório, só o da mãe.

É com a DNV que se tira a certidão de nascimento nos cartórios de registro civil. Uma vez levado a cartório, nada impede que um homem vá e se declare: “Eu sou o pai da criança”. Com isso, eles podem inclusive tirar passaporte e podem levar a criança legalmente para o exterior. Foi o que a quadrilha fez.

Tal situação apurada exemplifica o quão grave podem ser as implicações das deficiências de controles na emissão de documentos de identificação.

Para os casamentos, verificou-se a ausência ou a multiplicidade dos campos de matrícula da certidão, CPF e nomes dos cônjuges, conforme Tabela 2. Importante observar que, para cada evento de casamento, são considerados os dados de ambos os cônjuges.

Tabela 2: Resultado da verificação de unicidade e completude de certidões de casamento

<b>Campo chave</b>	<b>Ausentes</b>	<b>%</b>	<b>Em multiplicidade</b>	<b>%</b>
Matrícula da certidão	0	0,0	0	0,0
Cadastro de pessoa física (CPF)	4.318.240	92,4	10.811	0,2
Nomes dos cônjuges	-	0,0	82.267	1,8

Fonte: elaboração própria.

Nota-se que na maior parte dos registros de certidões de casamento não constava o número de inscrição no CPF para pelo menos um dos cônjuges, dificultando a identificação unívoca dos eventos e pessoas a que a certidões se referem.

Ainda no quesito completude dos dados relativos a casamento, observou-se que a base do Sirc não contava com os dados de averbações, anulações e retificações de forma tabulada. Essa situação limitou possíveis verificações quanto a vínculos ativos de casamento, divórcios e a eventualidade de haver pessoas com mais de uma certidão de casamento ativa no mesmo período, por exemplo.

Já para os óbitos, foram examinados os campos de matrícula de certidão, CPF, declaração de óbito (DO) e uma combinação de nome do registrado, data de nascimento e nomes das filiações. A Tabela 3 apresenta os registros de certidões de óbito com campos chave ausentes ou em multiplicidade.

Tabela 3: Resultado da verificação de unicidade e completude de certidões de óbito

<b>Campo chave</b>	<b>Ausentes</b>	<b>%</b>	<b>Em multiplicidade</b>	<b>%</b>
Matrícula da certidão	0	0,0	0	0,0
Cadastro de pessoa física (CPF)	4.250.869	84,0	15.012	0,3
Declaração de óbito	487.757	9,6	16.589	0,3
Nome, data de nascimento e nomes das filiações	-	-	4.473	0,1

Fonte: elaboração própria.

O registro dos dados de identificação em certidões de óbito pelos cartórios é previsto em Lei e é primordial para a suspensão tempestiva de pagamentos de benefícios previdenciários, trabalhistas e assistenciais. Por outro lado, deficiências na correta identificação

da pessoa falecida pode acarretar suspensões indevidas e diversos outros problemas para os beneficiários.

A metodologia para avaliação da completude e unicidade dos campos chave foi desenvolvida utilizando *scripts* em linguagem SQL.

## 5.2 VALIDADE, ACURÁCIA E QUALIFICAÇÃO DE DADOS DE IDENTIFICAÇÃO

Dado o baixo percentual de preenchimento de CPF obtido na análise de completude e considerando que o modelo de dados do Sirc permite que sejam inseridos múltiplos documentos para cada certidão, diferenciando-os pelo tipo de documento, elaborou-se procedimento que desconsidera essa marcação e busca em todos os documentos registrados os CPFs válidos que seriam candidatos a identificar o indivíduo.

A lista de CPFs candidatos é obtida aplicando-se teste para verificar a validade dos números de documento de identidade vinculados às certidões, a partir de regra matemática para geração de CPF utilizada pelo órgão emissor, isto é, pela Receita Federal do Brasil (RFB).

A partir da identificação dos CPFs candidatos, aplicou-se teste de acurácia das informações por meio do cruzamento com a base da RFB, fonte original dos dados de CPF, a fim de se obter o grau de correspondência entre os dados, atribuindo-se uma nota à comparação dos seguintes campos chave de um indivíduo: nome, nome das filiações e data de nascimento.

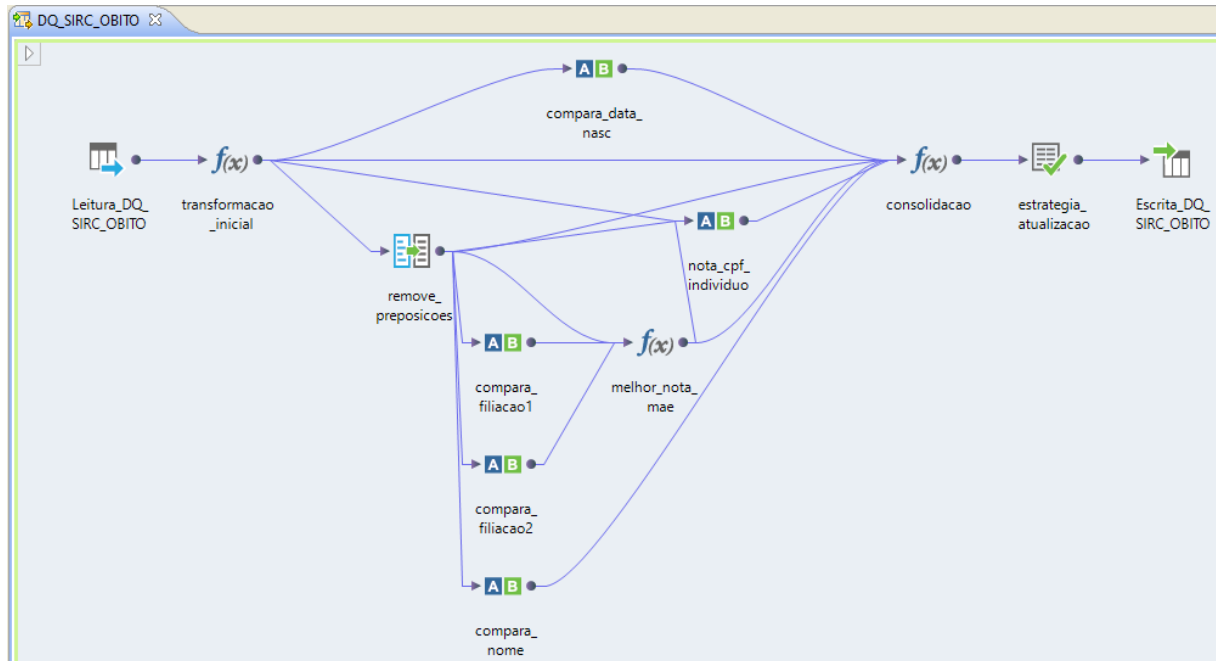
O procedimento utilizou *scripts* em linguagem SQL e mapeamentos no *software Informatica DataQuality* utilizando correspondência de identidade, que atribui nota de 0 (baixa correspondência) a 100 (exatamente iguais) para o grau de similaridade entre as informações da pessoa nas bases do Sirc e de CPF.

As variáveis de texto – nomes da pessoa e das filiações – passaram por etapa para retirar partículas e preposições. As datas de nascimento foram padronizadas para o formato DDMMYYYY.

O modelo de dados do Sirc prevê a possibilidade de registro de dados de múltiplas filiações, inclusive socioafetivas, enquanto a base de dados de CPF só contém o nome da mãe da pessoa. Desta forma, foi necessário comparar cada nome de filiação no Sirc com o nome da mãe no CPF e o melhor resultado obtido foi aproveitado para a avaliação da correspondência de identidade da pessoa.

As Figuras 1 e 2 ilustram o fluxo do procedimento realizado e a estratégia de correspondência dos campos identificadores da pessoa.

Figura 1: Fluxo do procedimento de *Data Quality* para qualificar informações da pessoa



Fonte: Tela do *software Informatica DataQuality*; elaboração própria.

Figura 2: Estratégia de correspondência dos campos identificadores da pessoa

Estratégias:

- Correspondência de Campo
  - Inverter Distância de Hamming
  - Distância de Hamming
  - Editar Distância
  - Distância Bigrama
  - Distância Jaro
- Correspondência de Identidade
  - Wide\_Contact
  - Contact
  - Individual
  - Resident
  - Address
  - Organization
  - Division
  - Household
  - Person\_Name

Campos:

Campos de Entrada	Campos Disponíveis
Person_Name_1 *	NOME_SIRC
Person_Name_2 *	NOME_CPF
Date_1 +	DT_NASC_SIRC
Date_2 +	DT_NASC_CPF
Id_1 +	
Id_2 +	
Attribute1_1	NOME_MAE_SIRC
Attribute1_2	NOME_MAE_CPF
Attribute2_1	
Attribute2_2	
Code_1	
Code_2	

Fonte: Comparação realizada no *software Informatica DataQuality*; elaboração própria.

Como resultado, foram verificados os CPFs que:

- a) estavam preenchidos corretamente, isto é, com um elevado grau de similaridade quando comparados aos dados da base de CPF e com o tipo de documento correto no Sirc;
- b) estavam marcados como outro tipo de documento no Sirc, porém cujas informações contavam com elevado grau de similaridade quando comparadas aos dados da base de CPF, e puderam ser adicionados ou corrigidos;
- c) estavam em branco ou preenchidos com um número inválido;

- d) estavam preenchidos com um número válido, mas que não constavam na base de CPF da RFB; e
- e) estavam preenchidos com informações incorretas ou com dados de outra pessoa.

A partir dos resultados da avaliação de acurácia, foi possível sanear e enriquecer os dados de CPF das pessoas registradas no Sirc, gerando bases mais qualificadas de nascimentos, casamentos e óbitos, que serão utilizadas nos procedimentos aplicados à FCB.

Para saneamento dos dados, os CPFs inexistentes ou que não correspondiam às informações registradas na base da RFB foram excluídos. Além disso, os documentos corretos foram adicionados ou corrigidos sempre que possível.

As Tabelas 4, 5 e 6 apresentam os resultados da qualificação de CPF para as certidões de nascimento, casamento e óbito.

Tabela 4: Resultados da qualificação de CPF do registrado em certidões de nascimento

Descrição	Quantidade de registros	%
a) CPF validado	2.686.993	21,6
b) CPF adicionado ou corrigido	-	0,0
c) CPF invalidado ou nulo	9.748.303	78,4
d) CPF inexistente na base da RFB	14	0,0
e) CPF incorreto ou com dados de outra pessoa	3.448	0,0
<b>Total</b>	<b>12.438.758</b>	<b>100,0</b>

Fonte: elaboração própria.

Tabela 5: Resultados da qualificação de CPF dos cônjuges em certidões de casamento

Descrição	Quantidade de registros	%
a) CPF validado	451	0,0
b) CPF adicionado ou corrigido	2.026.452	21,7
c) CPF invalidado ou nulo	7.046.196	75,4
d) CPF inexistente na base da RFB	15.669	0,2
e) CPF incorreto ou com dados de outra pessoa	165.832	1,8
<b>Total</b>	<b>9.254.600</b>	<b>100,0</b>

Fonte: elaboração própria.

Tabela 6: Resultados da qualificação de CPF do registrado em certidões de óbito

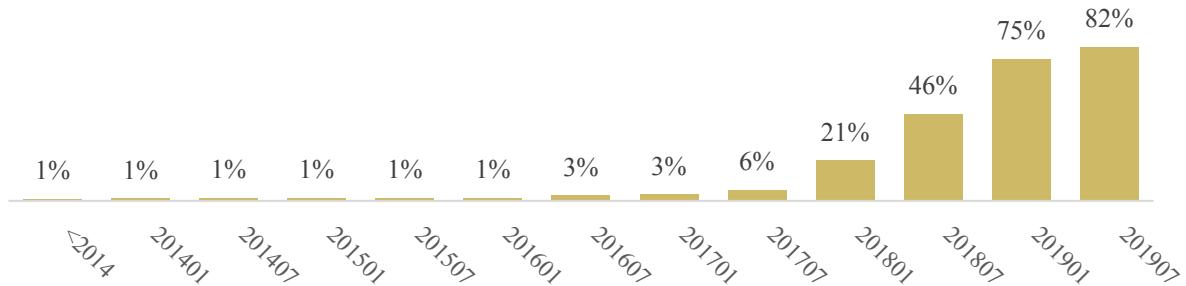
Descrição	Quantidade de registros	%
a) CPF validado	99	0,0
b) CPF adicionado ou corrigido	3.880.968	76,7
c) CPF invalidado ou nulo	1.159.078	22,9
d) CPF inexistente na base da RFB	2.357	0,0
e) CPF incorreto ou com dados de outra pessoa	17.588	0,3
<b>Total</b>	<b>5.060.090</b>	<b>100,0</b>

Fonte: elaboração própria.

É possível notar que, em grande parte do total de registros de certidões de nascimento, não contava CPF válido na base do Sirc (78,4%), em 1º de agosto de 2019, apesar de previsto na legislação (Lei 8.212/1991, art 68, § 2º).

Entretanto, analisando-se o preenchimento de CPF válidos nas certidões de nascimento na série temporal, pode-se perceber que há uma aparente melhora para as certidões emitidas em meses mais recentes, chegando a 82% de preenchimento qualificado, conforme Gráfico 1.

Gráfico 1: Preenchimento qualificado do CPF do registrado em certidões de nascimento por emissão



Fonte: elaboração própria.

Tal situação demonstra uma tendência de melhoria da qualidade dos dados de identificação de CPF nas certidões de nascimento do Sirc para os dois últimos anos, apesar do baixo percentual de preenchimento considerando os dados históricos.

Já para os dados de óbito, foi possível recuperar ou validar a maior parte dos CPFs referentes às certidões a partir dos procedimentos de qualidade de dados.

### 5.3 COMPLETUDE DA BASE DE DADOS

Segundo a Lei 11.977, de 7 de julho de 2009:

Art. 37. Os **serviços de registros públicos** de que trata a Lei nº 6.015, de 31 de dezembro de 1973, observados os prazos e condições previstas em regulamento, **instaurarão sistema de registro eletrônico**. (...)

Art. 39. Os atos registrares praticados a partir da vigência da Lei nº 6.015, de 31 de dezembro de 1973, serão inseridos no sistema de registro eletrônico, no **prazo de até 5 (cinco) anos** a contar da publicação desta Lei. (Vide Decreto nº 8.270, de 2014).

Parágrafo único. Os atos praticados e os documentos arquivados anteriormente à vigência da Lei nº 6.015, de 31 de dezembro de 1973, deverão ser inseridos no sistema eletrônico.

Assim, deveriam ser inseridos em sistema de registro eletrônico todos os atos de registros públicos tratados na Lei 6.015/1973, quais sejam: I – os nascimentos; II – os casamentos; III – os óbitos; IV – as emancipações; V – as interdições; VI – as sentenças declaratórias de ausência; VII – as opções de nacionalidade; VIII – as sentenças que deferirem a legitimação adotiva. Além disso, o dispositivo estabelece que serão averbados, dentre outros: as sentenças que decidirem a nulidade ou anulação do casamento, o desquite e o

restabelecimento da sociedade conjugal; as escrituras de adoção e os atos que a dissolverem; e as alterações ou abreviaturas de nomes.

Desta forma, objetivou-se verificar se o volume de registros constantes na base recebida do Sirc estava dentro do esperado quando confrontado com outras bases cadastrais, como CPF e Sisobi, e ao se analisarem os comandos legais que tratam do prazo para prestação das informações pelas serventias ao Sirc, que preveem o envio de todos os atos registrares, especialmente os praticados a partir de 1º de janeiro de 1976.

O procedimento consistiu em totalizar os quantitativos de registros das bases: do Sirc; do Sirc com CPF qualificado; do Sisobi qualificado; e de CPFs da RFB, com situação cadastral “regular” ou “pendente de regularização”. Utilizaram-se como filtros datas de nascimento ou de óbito anteriores a 1º de agosto de 2019. Os resultados são apresentados nas tabelas a seguir.

Tabela 7: Comparativo do volume de nascimentos nas bases do Sirc e do CPF

Ano de nascimento	Nascimentos no Sirc com CPF qualificado (A)	Nascimentos no Sirc (B)	Nascimentos no CPF (C)	B / C (%)
<1976	4.363	294.664	76.274.475	0,4
1976-2009	7.755	1.336.426	110.402.798	1,2
2010	831	41.260	2.330.303	1,8
2011	1.001	61.728	2.336.824	2,6
2012	1.274	76.589	2.300.444	3,3
2013	1.728	87.288	2.261.926	3,9
2014	2.519	182.455	2.280.658	8,0
2015	6.930	698.335	2.267.951	30,8
2016	63.564	2.596.605	2.365.949	109,7
2017	203.383	2.746.351	2.557.471	107,4
2018	1.207.279	2.786.919	2.810.571	99,2
2019	1.186.671	1.529.518	1.670.971	91,5
<b>Total</b>	<b>2.687.298</b>	<b>12.438.138</b>	<b>209.860.341</b>	<b>5,9</b>

Fonte: elaboração própria.

Tabela 8: Comparativo do volume de óbitos nas bases do Sirc e do Sisobi

Ano de óbito	Óbitos no Sirc com CPF qualificado (A)	Óbitos no Sirc (B)	Óbitos no Sisobi (C)	B / C (%)
<1976	273	14.068	37.919	37,1
1976-2009	28.282	133.939	12.391.784	1,1
2010	5.625	9.037	1.149.536	0,8
2011	12.074	18.285	1.210.378	1,5
2012	19.213	28.746	1.180.054	2,4
2013	23.054	34.089	1.208.489	2,8
2014	47.952	62.915	1.241.415	5,1



2015	193.995	274.928	1.266.946	21,7
2016	881.823	1.237.153	1.325.700	93,3
2017	946.228	1.258.595	1.239.119	101,6
2018	1.058.533	1.253.909	1.230.967	101,9
2019	656.530	716.897	739.580	96,9
<b>Total</b>	<b>3.873.582</b>	<b>5.042.561</b>	<b>24.221.887</b>	<b>20,8</b>

Fonte: elaboração própria.

É possível perceber que ainda é baixo o volume de dados de nascimentos e óbitos registrados no Sirc, alcançando menos de 6% das pessoas registradas na base de CPF e menos de 21% dos óbitos constantes na base do Sisobi.

Ainda, observa-se que o volume maior de registros está concentrado nos períodos mais recentes, notadamente a partir de 2016. Uma hipótese que explica essa situação é que a Resolução CGSirc 2/2015 estipulou o dia 10 de dezembro de 2015 como data limite para que os cartórios de registro cumprissem a obrigação de envio de registros para o Sistema Nacional de Registros Civis (Sirc):

Art. 1º O §4º do art. 1º da Resolução nº 1, de 9 de julho de 2015, passa a vigorar com a seguinte redação: "§4º As serventias de registro civil de pessoas naturais terão até o dia 10 de dezembro de 2015, para dar início ao envio dos dados, na forma desta Resolução".

Assim, apesar de haver previsão legal desde julho de 2009 (Lei 11.977/2009), com regulamentação desde junho de 2014 (Decreto 8.270/2014, alterado pelo Decreto 9.929/2019), somente houve o envio dos dados das serventias de registro civil de pessoas naturais ao Sirc, de forma regular, a partir de dezembro de 2015, com a publicação da Resolução CGSirc 2/2015, que estipulou data limite para início do envio.

De acordo com informações coletadas junto ao INSS, existiriam atos registrais mais antigos que não estariam disponíveis em formato digital e padronizado em alguns cartórios, dificultando a sua informação ao Sirc.

Justificativa similar foi obtida quanto à ausência de dados tabulados de averbações, anotações e retificações. Até a data de extração dos dados obtidos do Sirc, essas informações somente estariam presentes em campo textual de observações, não havendo padronização quanto à disponibilização das informações relevantes de forma estruturada.

#### 5.4 VALIDADE DOS TERMOS DE MATRÍCULA DAS CERTIDÕES

De acordo com o Decreto 7.231/2010, que regulamenta o art. 29 da Lei 6.015/1973, que dispõe sobre os registros públicos e dá outras providências:

Art. 3º As certidões previstas nos arts. 1º e 2º deverão contar com matrícula padronizada e unificada nacionalmente, que identifique o código nacional da serventia, o código do acervo, tipo de serviço prestado, ano do registro, tipo do livro, número do livro, número da folha, número do termo e dígito verificador.

O Provimento CNJ 3, de 7 de novembro de 2009, revogado pelo Provimento CNJ 63, de 14 de novembro 2017, previa que:

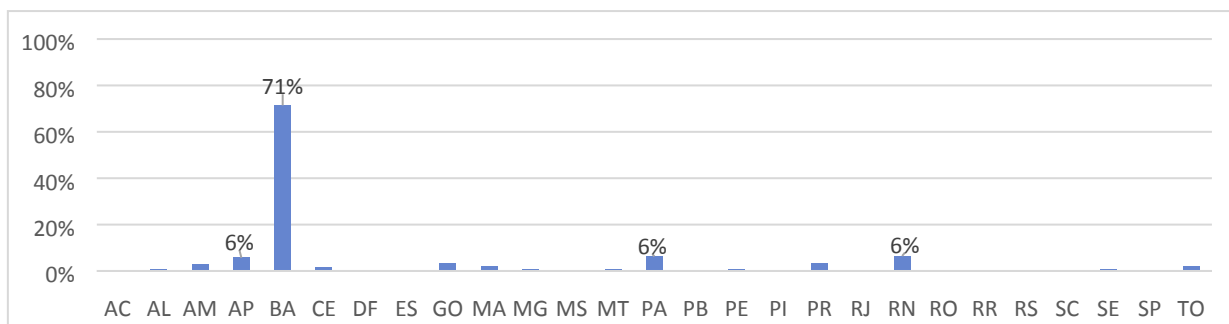
**§ 2º Os Cartórios de Registro Civil das Pessoas Naturais que não possuem acesso a microcomputador deverão lançar duas letras x (xx) no lugar do dígito verificador.** A inexistência do acesso a microcomputador deve ser informada a esta Corregedoria Nacional por meio do endereço físico Pça dos Três Poderes, Anexo I do Supremo Tribunal Federal, sala 356, CEP 70175900, Brasília, DF, ou do endereço eletrônico [justica.aberta@cnj.jus.br](mailto:justica.aberta@cnj.jus.br), anotando-se no ofício: REF Processo n. 58.681. (Grifos próprios)

Assim, é possível observar que o Provimento CNJ 3/2009 previa o preenchimento com “xx” para os casos de cartórios de registro civil que não possuíssem acesso a microcomputador. Porém, tal previsão foi revogada pelo Provimento CNJ 63/2017.

Dessa forma, foi elaborado procedimento para verificar a validade dos termos de matrícula das certidões, a partir de regra de formação matemática para geração do dígito verificador do termo de matrícula (módulo 11), composto pelos campos: número da serventia, número do acervo, número do tipo de serviço, ano de registro, número do tipo de livro, número da folha, número do termo e dígito verificador.

Como resultado, identificaram-se 1.118.030 registros de certidões de nascimento, casamento ou óbito com dígito verificador (DV) preenchido com “xx”, ou seja, não preenchido (5% do total). Desses, 907.432 (81%) referem-se a certidões de serventias da Bahia, o que representa 71% das certidões informadas por cartórios daquela UF (1.271.239), conforme apresentado no Gráfico 2.

Gráfico 2: Percentual de certidões com dígito verificador igual a “xx” por UF



Fonte: elaboração própria.

Observa-se que a ausência de preenchimento do DV inviabiliza a conferência da validade do termo de matrícula de certidão e deveria ser exceção à regra, que é o preenchimento de todos os campos.

## 6 APLICAÇÃO DOS DADOS DO SIRC ÀS ANÁLISES DA FCB

O presente capítulo apresenta as situações em que foi possível adicionar novas análises à Fiscalização Contínua de Benefícios (FCB) da Assistência Social, ou aperfeiçoar as já existentes, utilizando os dados de certidões do Sirc.

A avaliação dos dados do Sirc foi apresentada no Capítulo 5 e seus dados foram aproveitados para avaliação de credibilidade da documentação do Cadastro Único, enriquecimento dos dados de identificação de pessoas do Cadastro Único, bem como insumo de tipologias do Cadastro Único, Bolsa Família e BPC.

Cabe ressaltar que algumas das limitações apresentadas na avaliação da qualidade dos dados do Sirc – como o volume de dados abaixo do esperado e os problemas no preenchimento dos números de inscrição no CPF –, fizeram com que nem todos os testes planejados pudessem ser executados. O Capítulo 7 (Conclusão) traz possibilidades de trabalhos futuros, incluindo uma descrição dos testes desenhados, mas que não puderam ser desenvolvidos.

### 6.1 CREDIBILIDADE DA DOCUMENTAÇÃO DE CERTIDÕES NO CADASTRO ÚNICO

Conforme o art. 7º da Portaria MDS 177/2011 e as regras do bloco 5 do Manual do Entrevistador do Cadastro Único (BRASIL, 2017, p. 74), o responsável familiar (RF) deve informar, obrigatoriamente, o número de CPF ou o título de eleitor. Já os demais componentes da família devem informar, pelo menos, um dos seguintes documentos: Certidão de nascimento, casamento ou registro indígena; Registro de identidade; CPF; Título de eleitor; ou Carteira de

trabalho. Caso o RF seja de família indígena ou quilombola, pode ser informado apenas um dos documentos citados.

Assim, há normativo que permite a apresentação de certidão como documento de identificação no Cadastro Único, em conjunto com outros documentos ou isoladamente. Estão neste último caso apenas pessoas que não sejam o Responsável Familiar ou cuja família seja caracterizada como indígena ou quilombola.

Com relação à coleta dos dados de certidão, a Instrução Operacional 80/Senarc/MDS, de 25 de maio de 2016, reeditada em 24 de setembro de 2018 prevê que:

Com a alteração do Sistema de Cadastro Único, **os municípios devem adequar à coleta da certidão, para que a matrícula das certidões emitidas a partir de 2010 seja coletada de maneira completa, com os 32 dígitos.** Com a implantação da Versão 7.15, não será possível a digitação da matrícula dividida nos itens '3 – Número do livro', '4 – Número da folha' e '5 – Número do termo/RANI'. (Grifos próprios)

Considerando que a emissão de certidões com termo de matrícula unificado ocorreu apenas para emissões a partir de 1/1/2010, o procedimento de avaliação da credibilidade foi realizado em duas etapas.

Para certidões antigas, emitidas até 1/1/2010, verificou-se o preenchimento conjunto dos campos exigidos na inclusão da pessoa do Cadastro Único: nome do cartório, data de emissão, livro, folha, termo, UF e código do município.

Já para as certidões novas, emitidas após 1/1/2010, foi aplicado teste do dígito verificador módulo 11, utilizando a regra de formação do termo de matrícula, de 32 dígitos, composto pelos campos: número da serventia, número do acervo, número do tipo de serviço, ano de registro, número do tipo de livro, número da folha, número do termo e dígito verificador. A composição do termo de matrícula das certidões é apresentada na Figura 3.

Figura 3: Composição do termo de matrícula das certidões emitidas a partir de 1/1/2010



Fonte: Informe para os gestores 211, de 25 de março de 2010, do antigo Ministério do Desenvolvimento Social, atual Ministério da Cidadania (BRASIL, 2020c).

Os resultados foram filtrados para constar apenas os registros com inconsistências nos campos de certidão de nascimento ou casamento, de membros que não sejam o responsável familiar, e que não possuíam preenchidos corretamente outros documentos de identificação dentre os previstos no art. 7º da Portaria MDS 177/2011. Não foram objeto do procedimento os registros indígenas.

Como resultado, foram identificadas 4.767.701 pessoas registradas no Cadastro Único com inconsistências nos dados de certidão de nascimento ou casamento, sendo a certidão o único documento preenchido ou havendo inconsistências também no preenchimento dos outros documentos. Dessas, 1.283.795 certidões são do modelo antigo (anteriores a 1/1/2010) e 3.483.906 certidões são do modelo novo (posteriores a 1/1/2010), conforme Tabela 9.

Tabela 9: Inconsistências no preenchimento de certidões no Cadastro Único

<b>Modelo de certidão</b>	<b>Inconsistências</b>
Antigo (emissão anterior a 1/1/2010)	1.283.795
Novo (emissão posterior a 1/1/2010)	3.483.906
<b>Total</b>	<b>4.767.701</b>

Fonte: elaboração própria.

Em tese, esses registros de pessoas não contam com nenhum documento válido de identificação cadastrado, o que pode indicar a ocorrência de falhas importantes no processo de cadastramento e até, em último caso, possível cometimento de fraude, como a inscrição de pessoas “fantasmas” ou já cadastradas em outra família no Cadastro Único.

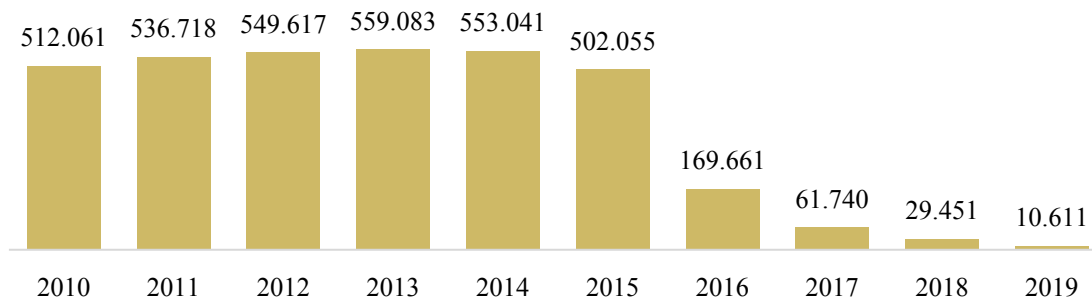
Dentre as possíveis motivações para este cadastro indevido, estão a possibilidade de recebimentos em multiplicidade de benefícios assistenciais, como transferências do Programa Bolsa Família, e até a redução da renda familiar per capita para enquadramento nas regras de elegibilidade do programa.

O maior volume de inconsistências identificadas ocorreu em certidões do novo modelo, por existir regra de formação matemática que possibilita a verificação do número cadastrado, o que não ocorre para as certidões anteriores a 1/1/2010, em que apenas é possível fazer testes mínimos de preenchimento dos campos. Dessa forma, não se pode dizer que as certidões anteriores a 2010 estão mais fidedignas, mas apenas que são escassos os procedimentos de controle que podem verificar sua consistência básica.

É possível, aliás, que se dê o contrário: com a normatização trazida finalmente pelo Decreto 8.270/2014, a qualidade das certidões comece a melhorar. Assim, analisando-se as

inconsistências em certidões do novo modelo por ano de emissão, apresentada pelo Gráfico 3, é possível notar uma aparente melhora para as certidões emitidas em anos mais recentes, notadamente a partir de 2016, no que tange ao mesmo controle.

Gráfico 3: Inconsistências em certidões do Cadastro Único por ano de emissão



Fonte: elaboração própria.

Uma hipótese para a queda mais acentuada de inconsistências a partir de 2016 é que a Resolução CGSirc 2/2015 estabeleceu a data limite para que os cartórios de registro cumprissem a obrigação de envio de registros para o Sistema Nacional de Registro Civil (Sirc) a partir de dezembro de 2015.

Com o advento do Sirc e os comandos dados pelo art. 6º do Provimento 63/2017 do CNJ e pelo art. 23 da Lei 13.846/2019, passou a haver previsão para incluir o número do CPF nos registros de certidões de nascimento, casamento e óbito dos cartórios de registros de pessoas naturais (Lei 8.212/91, art. 68, § 2º), bem como para compartilhamento dos dados de certidões com outros órgãos (Decreto 9.929/2019, art. 7º). Tais previsões podem contribuir para o aprimoramento das validações de certidões registradas no Cadastro Único, bem como possibilitar o enriquecimento de dados de CPF das pessoas cadastradas.

Nesse sentido, foi possível elaborar procedimento adicional para verificar a acurácia da identificação dos indivíduos no Cadastro Único utilizando-se os dados do Sirc. Foram utilizadas a base do Cadastro Único e os dados de certidões de nascimento e casamento do Sirc após processo de qualificação dos dados da inscrição CPF.

Utilizou-se a matrícula das certidões de nascimento ou casamento como chave de comparação, de forma integral ou parcial. O cruzamento parcial considerou os dígitos da 9ª a 30ª posição, enquanto o cruzamento integral desconsiderou apenas os dígitos verificadores.

Escolheu-se essa abordagem por englobar o maior número de combinações possíveis, levando-se em conta a previsão da Instrução Operacional 80/Senarc/MDS, de 25 de maio de 2016, reeditada em 24 de setembro de 2018, que prevê o preenchimento parcial da matrícula da certidão no Cadastro Único:

Nos casos em que os dados da matrícula da certidão emitida a partir de janeiro de 2010 foram coletados de maneira particionada e ainda não foram digitados, para inserir as informações da certidão na Versão 7.15 do Sistema de Cadastro Único, o operador deverá reconfigurar a matrícula de 32 dígitos e inserir no item '6 – Matrícula', observando as seguintes orientações:

1. Incluir 14 zeros nas posições de 1 a 14;
2. Incluir o tipo da certidão na 15ª posição: (1) para nascimento, (2) para casamento;
3. Incluir o número do livro, com cinco dígitos, da 16ª à 20ª posição, completando com zeros à esquerda quando necessário;
4. Incluir o número da folha na posição 21ª à 23ª, completando com zeros à esquerda quando necessário;
5. Incluir o número do termo, com sete dígitos, da 24ª à 30ª posição, completando com zeros à esquerda quando necessário; e
6. Nas duas últimas posições, 31ª e 32ª, incluir a letra X.

Elaborou-se, então, procedimento de *Data Quality* para comparação das informações de pessoas recuperadas do Cadastro Único e do Sirc. Essa etapa atribuiu nota para o grau de similaridade dos registros, considerando os campos de nome, data de nascimento e nome da mãe no CPF e nomes das filiações no Sirc. Para os casamentos, os dados dos dois cônjuges foram testados na comparação, mantendo-se no resultado aquele com melhor nota final.

Como resultado, foram identificados 23.194 registros no Cadastro Único com certidões com dados divergentes. Destes, 8.162 eram certidões com correspondência integral do termo da certidão, o que pode sugerir a existência de erro ou fraude no momento de cadastro.

Certidões de nascimento e casamento são pessoais e o registro de documentação no Cadastro Único de forma irregular pode permitir a ocorrência de fraudes e pagamentos indevidos de benefícios.

## 6.2 ENRIQUECIMENTO DOS DADOS DE IDENTIFICAÇÃO DO CADASTRO ÚNICO

Da perspectiva do controle externo, a correta identificação de pessoas permite que sejam implementadas, com alto grau de acerto, trilhas de auditoria para avaliar o cumprimento de requisitos do Cadastro Único e também de seus programas usuários. Por exemplo, é possível verificar se há registro de óbito para uma pessoa inscrita no Cadastro Único que ainda não tenha sido verificada pelo órgão gestor.

Embora não haja obrigatoriedade de apresentação de todos os documentos, há orientação para que sejam coletados o maior número possível de informações que cada pessoa possuir. Quanto mais completa e qualificada for a identificação das pessoas registradas no

Cadastro Único, maior é a possibilidade de se fazer a correta identificação de sua vulnerabilidade e, com isso, implementar ações específicas (BRASIL, 2017, p. 75).

Nos ciclos anteriores da FCB, desenvolveu-se processo para aprimorar os dados de identificação do Cadastro Único por meio da comparação com outras bases de pessoas disponíveis. Já se utilizava na qualificação dos dados, as bases de CPF da RFB e de título de eleitor do TSE e, neste trabalho, foram incluídos os dados do Sirc.

A primeira etapa da qualificação cruza os dados do Cadastro Único com as bases de CPF e de título de eleitor. Avalia-se, então, o grau de similaridade entre campos identificadores da pessoa em cada base, como: nome, nome da mãe, nome do pai, data de nascimento, CPF e título de eleitor. Essa avaliação gera uma nota de similaridade, que permite classificar os registros referentes a uma pessoa como iguais, distintos, inconclusivos ou nulos.

Para os registros em que não foi identificado um número de CPF com nota aceitável nesta primeira etapa, realiza-se, então, uma etapa complementar de busca de CPFs diretamente na base de CPF da RFB, empregando técnica de comparação denominada *Fuzzy String Matching*, que identifica sequências que tenham correspondência aproximada, dado um certo padrão. O padrão utilizado é uma composição do *soundex* dos campos: primeiro nome da pessoa; segundo nome em diante da pessoa; primeiro nome da mãe; segundo nome em diante da mãe. O uso deste padrão permite associar pessoas com pequenas diferenças de grafia ou abreviações nos nomes, privilegiando o primeiro nome da pessoa e o da mãe.

As duas etapas em conjunto conseguem recuperar dados qualificados de CPF, chave utilizada predominantemente pelas tipologias, para a maior parte das pessoas inscritas no Cadastro Único. No entanto, uma parte considerável das pessoas ainda carece de informações de CPF, já que foram inscritas mediante a apresentação de outros documentos, como certidões de nascimento e casamento, documentos de identidade, carteira de trabalho e NIS.

A primeira etapa de qualificação, que utiliza comparação com as bases de CPF e de título de eleitor, tem bom desempenho no quesito de processamento de dados, tendo em vista utilizar chaves indexadas para comparação, e é aplicada para o conjunto total dos 74 milhões de registros. Ela apresenta bons resultados principalmente para pessoas maiores de 16 anos, que têm maior probabilidade de terem CPF e título de eleitor emitidos.

Já a segunda etapa, de busca por correspondência aproximada, é bastante custosa computacionalmente e gera muitos falsos positivos devido a ocorrência de homônimos. Assim, aplica-se esta etapa somente para o subconjunto das pessoas maiores de 16 anos, utilizando-se como limite mínimo (*threshold*) notas acima de 95% de similaridade.



Na tentativa de contribuir para a identificação das pessoas sem CPF restantes, implementou-se procedimento de qualificação utilizando os dados do Sirc, aproveitando-se da existência de número de certidão no Cadastro Único e no Sirc, bem como da previsão de registro dos dados de CPF e outros documentos no Sirc.

A qualificação utilizando campos de chaves artificiais (*surrogate keys*), como é o caso da qualificação utilizando a matrícula das certidões, tende a ser mais precisa e ter melhor desempenho quando comparada à qualificação utilizando chaves naturais, que têm um grau maior de incerteza.

O procedimento aproveitou a avaliação da acurácia dos dados de certidão do Cadastro Único, utilizando-se a matrícula das certidões de nascimento ou casamento, de forma integral ou parcial, como chave de comparação com o Sirc e obtendo-se notas para o grau de correspondência entre campos de identificação da pessoa.

Obviamente, foram aproveitadas apenas as certidões em que o CPF foi considerado aceitável no processo de qualificação do Sirc relatado na Seção 5.2.

Importa destacar que, para a FCB de 2019, concomitantemente com o trabalho utilizando o Sirc, foram realizados outros ajustes nos procedimentos já existentes de construção das bases de comparação de pessoas. A busca de CPFs por *Fuzzy String Matching* passou a buscar também o CPF de pessoas com dados divergentes nos cadastros. Além disso, a base de comparação do Cadastro Único gerada passou a incluir pessoas com outros documentos cadastrados que não o CPF, aumentando o universo de pessoas objeto de avaliação pelas tipologias.

Como resultado de todo esse esforço, aliado ao processo natural de aprimoramento realizado pelos gestores do Cadastro, foi possível melhorar o percentual de CPFs qualificados da base de comparação, conforme apresentado na Tabela 10.

Tabela 10: Resultado comparativo da qualificação de CPF no Cadastro Único

<b>Resultado da qualificação de CPF</b>	<b>FCB-2018</b>	<b>%</b>	<b>FCB-2019</b>	<b>%</b>
Validado na comparação com CPF (RFB) e título de eleitor (TSE)	50.012.745	67,5	58.622.681	79,2
Adicionado na busca por <i>Fuzzy String Matching</i> na base de CPF (RFB)	2.779.477	3,8	2.378.309	3,2
Adicionado na comparação com certidões do Sirc (INSS)	-	-	185.154	0,3
Não recuperado ou rejeitado	21.696.704	29,3	12.873.840	17,4
<b>Total</b>	<b>74.488.926</b>	<b>100,0</b>	<b>74.059.984</b>	<b>100,0</b>

Fonte: elaboração própria.

Observa-se que foi possível recuperar dados de CPF não identificados pelas demais técnicas utilizando dados de certidões de nascimento e casamento do Sirc. É possível notar, no entanto, que o volume de CPFs recuperados a partir do Sirc foi baixo em comparação com as outras técnicas e bases de dados empregadas.

Dentre as possíveis causas para o resultado não ser percentualmente tão expressivo, estão as deficiências no preenchimento dos dados de certidões no Cadastro Único e ausência de registros de CPF no Sirc, descritas, respectivamente, na Seção 6.1 e no Capítulo 5.

Não obstante tais limitações do cenário atual, o aprimoramento de dados e controles propostos e acompanhados no âmbito da FCB tendem a possibilitar que as análises empreendidas utilizando o Sirc agreguem cada vez mais à qualificação de dados cadastrais das pessoas com o passar do tempo.

Destaca-se que o art. 7º do Decreto 9.929/2019 prevê que os dados contidos no Sirc poderão ser disponibilizados, após a autorização do CGSirc, aos órgãos e entidades da administração pública federal, estadual, distrital e municipal que os solicitarem, observado o disposto no art. 31 da Lei 12.527/2011 (Lei de Acesso à Informação), que versa sobre o tratamento de informações pessoais.

Assim, vislumbra-se que a incorporação dos dados do Sirc nos procedimentos de controle efetuados pelo próprio Ministério da Cidadania pode contribuir para o aprimoramento dos dados cadastrais e enriquecimento da focalização dos benefícios assistenciais.

### 6.3 TIPOLOGIAS DE INDÍCIOS DE FALECIMENTO

Em trabalhos anteriores da FCB, foram aplicadas tipologias para avaliar a existência de indícios de falecimento relativos a pessoas inscritas no Cadastro Único ou beneficiárias de Bolsa Família e BPC, sendo a base do Sistema de Controle de Óbitos (Sisobi) a principal base de dados confrontada.

Durante o curso de 2019, o INSS informou que a disponibilização dos dados de óbito para pessoas jurídicas de direito público, amparada pela Lei 12.527/2011 (Lei de Acesso à Informação – LAI), passará a ser operacionalizada por meio do acesso aos dados do Sistema Nacional de Registro Civil (Sirc), em substituição ao Sisobi.

Como apresentado na Seção 5.3, os volumes de dados de falecimento do Sisobi e do Sirc não têm correspondência exata, existindo dados de óbito presentes no Sisobi que não estão no Sirc e também o oposto.

Da mesma forma, outras bases disponíveis na administração pública contêm marcações de eventos de falecimento da pessoa, não necessariamente capturados pelo Sirc ou

Sisobi. Pode-se citar: CPF com situação cadastral de “titular falecido” (RFB); instituidor de pensão por morte na folha de pagamentos de benefícios previdenciários – Maciça (INSS); e instituidor de pensão por morte na folha de pagamentos de agentes públicos (Siape, Extrasiape e folhas de pagamento de estados e municípios).

Desta forma, aproveitando o estudo dos dados do Sirc realizado neste trabalho, aprofundou-se nas tipologias de verificação de indícios de óbito, elaborando-se procedimentos complementares aos que já existiam. As Tabelas 11, 12 e 13 apresentam os resultados obtidos.

Tabela 11: Pessoas no Cadastro Único com indícios de falecimento

Fonte da informação	FCB – 2018 Suspeita de óbito no órgão gestor		FCB – 2019 Suspeita de óbito no órgão gestor	
	Sim	Não	Sim	Não
Sisobi	179.067	237.141	228.052	72.065
Instituidor de pensão por morte no INSS	39.623	27.011	39.749	23.647
Titular falecido na base de CPF	88.742	29.810	114.164	45.459
Sirc	-	-	208.122	67.904
Agentes públicos instituidores de pensão por morte	-	-	222	219
<b>Total (distintos)</b>	<b>179.361</b>	<b>261.412</b>	<b>229.442</b>	<b>108.308</b>

Fonte: elaboração própria.

Tabela 12: Beneficiários do Programa Bolsa Família com indícios de falecimento

Fonte da informação	FCB – 2018 Suspeita de óbito no órgão gestor		FCB – 2019 Suspeita de óbito no órgão gestor	
	Sim	Não	Sim	Não
Sisobi	16.667	30.014	15.122	15.278
Instituidor de pensão no INSS	1.353	3.619	785	3.182
Titular falecido na base de CPF	7.914	5.589	7.417	9.193
Sirc	-	-	11.035	12.853
Agentes públicos instituidores de pensão por morte	-	-	1	9
<b>Total (distintos)</b>	<b>16.679</b>	<b>34.659</b>	<b>15.156</b>	<b>22.799</b>

Fonte: elaboração própria.

Tabela 13: Beneficiários BPC com indícios de falecimento

Fonte da informação	FCB-2018 Despacho judicial		FCB-2019 Despacho judicial	
	Sim	Não	Sim	Não
Sisobi	119	1.343	85	587

Titular falecido na base de CPF	156	1.980	151	1.481
Instituidor de pensão por morte no INSS	-	-	38	343
Agentes públicos instituidores de pensão por morte	-	-	1	8
Sirc	-	-	17	63
<b>Total (distintos)</b>	<b>275</b>	<b>3.323</b>	<b>268</b>	<b>2.394</b>

Fonte: elaboração própria.

Naturalmente, há que se considerar que em muitos casos os quantitativos de indícios de procedimentos que foram realizados em fiscalizações anteriores tendem a cair nas fiscalizações seguintes, sugerindo que os controles têm sido aprimorados pelos gestores. No entanto, quando se analisa mais detalhadamente, é possível notar que há ganho com as tipologias utilizando as bases e técnicas adicionais.

Cabe ressaltar que uma mesma pessoa pode ter indício de falecimento em mais de uma das bases utilizadas para verificação, tendo sido, nesse caso, contabilizada em mais de uma linha das tabelas. Note-se que essa repetição é relevante porque pode indicar uma maior certeza quanto ao óbito. No entanto, obviamente, para fins de avaliação da necessidade de ajuste do Cadastro Único ou das folhas de pagamento do Bolsa Família e do BPC, cada pessoa só deve ser contabilizada uma única vez e, por isso, a linha de totalização considera apenas as ocorrências distintas.

## 7 CONCLUSÃO

O presente trabalho de conclusão de curso apresentou uma avaliação da qualidade de dados da base de registros civis de pessoas naturais do Sirc, que contém registros de eventos-chave para a identificação de pessoas e execução de políticas públicas, incluindo dados, por exemplo, de: nascimentos, filiações socioafetivas e multiparentais, casamentos, mudanças de nome, óbitos, dentre outros.

Os procedimentos desenvolvidos permitiram que os objetivos propostos fossem alcançados: avaliou-se a qualidade de dados do Sirc (Capítulo 5); avaliou-se a qualidade dos dados relativos a certidões da base do Cadastro Único (Seção 6.1); foram qualificados os dados de identificação de pessoas inscritas no Cadastro Único e beneficiárias de programas assistenciais (Seção 6.2); foram aprimoradas as tipologias de pessoas falecidas por meio da utilização de outras fontes de dados, incluindo o Sirc (Seção 6.3).

Assim, a partir da aplicação da metodologia aos dados do Sirc, foi possível incorporar a base como insumo da Fiscalização Contínua de Benefícios do TCU, gerando indícios relevantes de inconsistências cadastrais e irregularidades em benefícios.

Apesar da utilização de dados do Sirc ter se mostrado viável e a pesquisa ter alcançado seus objetivos, há diversas oportunidades de melhorias que serão endereçadas pela equipe de fiscalização para aprimoramento dos controles e dados das bases, nos processos de controle externo em curso.

Trabalhos futuros podem tirar proveito deste aprimoramento para outros testes de auditoria, como, por exemplo: certidões de nascimento declaradas no Cadastro Único ou na folha de pagamentos previdenciários e inexistentes no Sirc, indicando a possibilidade de registros de pessoas “fantasmas”; pensão por morte concedida sem certidão de óbito do instituidor no Sirc; auxílio maternidade concedido sem certidão de nascimento do filho da titular no Sirc; e omissão de cônjuge, registrado no Sirc e que tenha indício de renda, em famílias que recebem Bolsa Família ou benefício de prestação continuada (BPC).

Além disso, trabalhos futuros podem aproveitar os dados do Sirc para construir soluções de *Master Data Management* (MDM) que consolidem dados de cidadãos, bem como para construção de redes de dados de vínculos entre as pessoas.

Por fim, cabe registrar as principais lições aprendidas com este trabalho: o enorme potencial que os dados de registros civis têm para contribuir para a correta identificação de pessoas e focalização dos beneficiários de políticas públicas; e a necessidade de integração das bases cadastrais de cidadãos existentes na administração pública, tomando os cuidados necessários para implementar mecanismos que as mantenham confiáveis, completas e consistentes.

## REFERÊNCIAS

- BERSON, Alex; DUBOV, Larry. **Master data management and data governance, 2/E**. McGraw Hill Professional, 2010.
- BRASIL. Secretaria Nacional de Renda de Cidadania (Senarc). **Manual do entrevistador 4ª Edição: Cadastro Único para Programas Sociais**. Brasília: Ministério da Cidadania (antigo Ministério do Desenvolvimento Social – MDS), 2017.
- BRASIL. Tribunal de Contas da União. **Ficha-Síntese – Fiscalização Contínua de Benefícios Sociais**. Disponível em: <https://portal.tcu.gov.br/biblioteca-digital/ficha-sintese-fiscalizacao-continua-de-beneficios-sociais.htm>. Acesso em 16 nov. 2019a.
- BRASIL. Fantástico. **Investigação: crianças africanas deixam o Rio com documentos falsos**. Disponível em: <https://g1.globo.com/fantastico/noticia/2019/11/10/investigacao-criancas-africanas-deixam-o-rio-com-documentos-brasileiros.ghtml>. Acesso em 3 dez. 2019b.
- BRASIL, Tribunal de Contas da União. Regimento interno do Tribunal de Contas da União. **Boletim do Tribunal de Contas da União Especial**, n. 1, 2020a.
- BRASIL. **Portal do Sistema Nacional de Informações de Registro Civil**. Disponível em: <http://www.sirc.gov.br/>. Acesso em 20 jan. 2020b.
- BRASIL. Ministério do Desenvolvimento Social (atual Ministério da Cidadania). Informe para os gestores 211, de 25 de março de 2010. Disponível em: [http://www.mds.gov.br/webarquivos/publicacao/bolsa\\_familia/Informes/Informe%20211.pdf](http://www.mds.gov.br/webarquivos/publicacao/bolsa_familia/Informes/Informe%20211.pdf). Acesso em 5 mar. 2020c.
- CAI, Li; ZHU, Yangyong. The challenges of data quality and data quality assessment in the big data era. **Data science journal**, v. 14, 2015.
- CFC, 2016. NBC TA 200 (R1) Objetivos Gerais do Auditor Independente e a Condução da Auditoria em Conformidade com Normas de Auditoria. Brasília: Conselho Federal de Contabilidade.
- CHAPMAN, Pete et al. **CRISP-DM 1.0: Step-by-step data mining guide**. SPSS inc, v. 16, 2000.
- DATE, Christopher J. Introdução a sistemas de bancos de dados. Elsevier Brasil, 2004.

ENGLISH, Larry P. **Improving data warehouse and business information quality: methods for reducing costs and increasing profits**. New York: Wiley, 1999.

HALONEN, Antti. Being open about data. **Analysis of the UK open data policies and applicability of data**, retrieved July, v. 17, p. 2012, 2012.

HEUSER, Carlos Alberto. **Projeto de banco de dados: Volume 4 da Série Livros didáticos informática UFRGS**. Bookman Editora, 2009.

HOBERMAN, Steve. **Data model scorecard: applying the industry standard on data model quality**. Technics Publications, 2015.

LEHMKUHL, Camila Schwinden *et al.* **O acesso à informação no Sistema Nacional de Informações de Registro Civil (SIRC)**. 2017.

LOSHIN, David. **Master data management**. Morgan Kaufmann, 2010.

MAYDANCHIK, Arkady. **Data quality assessment**. Technics publications, 2007.

REDMAN, Thomas C.; BLANTON, A. **Data quality for the information age**. Artech House, Inc., 1997.

SALOMÃO, Marcos Costa. **A Filiação Socioafetiva pela Posse de Estado de Filho e a Multiparentalidade no Provimento 63 do CNJ**. 2018.

SEBASTIAN-COLEMAN, Laura. **Measuring data quality for ongoing improvement: a data quality assessment framework**. Newnes, 2012.

SVOLBA, Gerhard. **Data quality for analytics using SAS**. SAS Institute, 2012.

WANG, Richard Y.; STRONG, Diane M. Beyond accuracy: What data quality means to data consumers. **Journal of management information systems**, v. 12, n. 4, p. 5-33, 1996.