



RELATÓRIO DE ACOMPANHAMENTO DE DADOS

Medidas de resposta à pandemia do Covid-19

Secretaria de Controle Externo da Gestão Tributária, da Previdência e da Assistência Social

SEGUNDO RELATÓRIO DO ACOMPANHAMENTO ESPECIAL DE DADOS

TC 016.834/2020-8

Fiscalização: Registro Fiscalis 70/2020**Relator:** Ministro Bruno Dantas**DA FISCALIZAÇÃO****Modalidade:** Acompanhamento**Ato originário:** Plano Especial de Acompanhamento das Ações de Combate à Covid-19, aprovada na Sessão Plenária do dia 25/3/2020 (TC 016.602/2020-0).**Objetivo da fiscalização:** Empregar análise de dados no acompanhamento das medidas de resposta à crise do Covid-19 nas áreas de Previdência Social, Assistência Social e Administração Tributária, visando subsidiar os acompanhamentos específicos e desenvolver painel de dados que forneça transparência sobre o nível de execução dessas medidas.**Atos de designação:** Portaria de Fiscalização - SecexPrevidência 166, de 15 de abril de 2020 (peça 2), Portaria de Fiscalização - SecexPrevidência 167, de 15 de abril de 2020 (peça 3) e Portaria de Fiscalização – SecexPrevidência 376, de 30 de julho de 2020 (peça 64).**Supervisor:** Rodrigo Otávio Coelho Hildebrand (AUFC, matrícula 9.466-8).**Composição da equipe:**

Nome	Matrícula	Cargo
Luíza da Silva Jaques	10.691-7	AUFC (coordenadora)
Daniel Moreira Guilhon	7.668-6	AUFC
Luiz Henrique Batistuta Gomide	9.981-3	AUFC
Theuryn Saches Loureiro Figueiredo	3.071-6	AUFC

Período de Realização da Fiscalização: 1/4/2020 a 30/9/2020.**DO ÓRGÃO FISCALIZADO****Órgãos fiscalizados (Vinculadores):** Ministério da Cidadania; Ministério da Economia; Instituto Nacional do Seguro Social – INSS (Ministério da Economia).**Responsáveis:**

Min. da Cidadania: Onyx Dornelles Lorenzoni (CPF 210.259.320-72), a partir de 13/2/2020;

Min. da Economia: Paulo Roberto Nunes Guedes (CPF 156.305.876-68), a partir de 1/1/2019;

INSS: Leonardo José Rolim Guimarães (CPF 436.473.754-20), a partir de 29/1/2020;

Vinculação TCU: SecexPrevidência**PROCESSOS CONEXOS****TC 016.827/2020-1.** Acompanhamento Especial das medidas de resposta à crise do Coronavírus para proteção da renda de informais e pessoas de baixa renda.**TC 016.841/2020-4.** Acompanhamento Especial das medidas de resposta à crise do Coronavírus na arrecadação tributária e previdenciária federal.**TC 016.830/2020-2.** Acompanhamento Especial das medidas de resposta à crise do Coronavírus na Previdência Social.**TC 009.922/2019-9.** Acompanhamento da concessão, manutenção e pagamento de benefícios assistenciais do exercício de 2019.

RESUMO

O presente acompanhamento representa uma das 27 ações previstas no Plano Especial de Acompanhamento das Ações de Combate à Covid-19, aprovado pelo plenário do Tribunal de Contas da União, em sessão plenária virtual realizada em 25/3/2020, que prevê a forma de trabalho da Corte de Contas no acompanhamento das medidas adotadas pela administração pública federal nesse momento excepcional de enfrentamento à pandemia de Covid-19.

A consolidação dos dados das diversas iniciativas que foram adotadas nas áreas de Assistência Social, de Previdência Social e de Gestão Tributária como resposta à pandemia de Covid-19 apresenta-se como um grande desafio, dada a grande compartimentalização das bases de dados do Setor Público e a complexidade das políticas envolvidas.

A principal medida objeto deste segundo relatório de acompanhamento com enfoque em dados foi a avaliação da qualidade dos dados do Cadastro de Pessoas Físicas após o processo de regularização de ofício realizado em abril de 2020 pela Secretaria da Receita Federal do Brasil, para permitir a concessão mais célere do auxílio emergencial de proteção social a pessoas em situação de vulnerabilidade devido à pandemia do Covid-19, estabelecido pela Lei 13.982/2020.

O presente relatório apresenta a aplicação de técnicas de estatística, análise de dados e tecnologia da informação na fiscalização do grande volume de dados objeto deste acompanhamento, em busca de indícios de inconsistências na base cadastral do CPF. Os achados de auditoria e encaminhamentos aqui relatados buscam auxiliar na correção dessas situações para aprimoramento do CPF, essencial para a correta identificação dos cidadãos brasileiros e para efetivação de diversas políticas públicas.



SUMÁRIO

1. Introdução	5
2. Visão geral	7
3. Credibilidade.....	8
3.1. Verificação da unicidade.....	8
3.2. Verificação da uniformidade.....	9
3.3. Verificação da acurácia.....	9
3.4. Verificação da completude, validade e consistência.....	9
3.5. Conclusão.....	13
4. Tipologias	13
4.1. Registro de óbito em bases da Administração Pública federal ou estadual.....	13
4.2. Registro de supercentenários	15
4.3. Inscrições de CPF suspensas há 11 anos ou mais.....	17
5. Regularização de ofício.....	20
6. Inscrições regulares do CPF superiores à população brasileira	21
7. Conclusão.....	24
8. Proposta de encaminhamento	25
Apêndice I.....	27

1. Introdução

1. Trata-se do segundo Relatório de acompanhamento (Racom) de dados relacionados às ações de combate à Covid-19 nas áreas de Assistência Social, Previdência Social e Gestão Tributária, vinculadas à atuação da SecexPrevidência. O Racom é decorrente do Plano especial de acompanhamento das ações de combate à Covid-19, aprovado na sessão plenária do dia 25/3/2020 (TC 016.602/2020-0), que objetiva focalizar a atuação dos órgãos e instituições públicas na luta contra a pandemia (peça 1).

2. Por se tratar de acompanhamento focado na análise e cruzamento de dados, ele se integra aos demais acompanhamentos emergenciais desta Secretaria de forma complementar, adicionando testes e procedimentos de auditoria específicos, quando possível.

3. Como o Tribunal de Contas da União conta com um conjunto abrangente de bases de dados e já realizou diversos trabalhos com apoio de técnicas de análise de dados, está em uma posição favorável para contribuir por meio de fiscalizações que forneçam transparência sobre as medidas de resposta à crise do novo Coronavírus e que identifiquem eventuais falhas nos programas e ações emergenciais em curso.

4. Dentre esses trabalhos voltados à auditoria de dados, a SecexPrevidência emprega, desde 2015, a metodologia de fiscalização com uso intensivo de ferramentas e técnicas de análise de dados denominada Fiscalização Contínua de Benefícios (FCB), com o objetivo de identificar, por meio de cruzamentos sistemáticos de bases de dados, indícios de irregularidades em benefícios, 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.

5. No ano de 2020, foram prolatados os Acórdãos 1.123/2020-TCU-Plenário, relativo à FCB da Assistência Social, de relatoria do Ministro-Substituto Marcos Bemquerer, e 1.350/2020-TCU-Plenário, relativo à FCB de Previdência Social, de relatoria do Ministro Bruno Dantas. Juntas, estas fiscalizações avaliaram folhas de pagamento mensais de aproximadamente 49 milhões de benefícios, resultando em um benefício potencial de R\$ 5,2 bilhões, relativo à revisão e ao eventual cancelamento dos benefícios com indícios de irregularidade.

6. Os principais riscos identificados nas medidas de resposta à pandemia Covid-19 são tratados nos respectivos Acompanhamentos Especiais, todos de relatoria do Ministro Bruno Dantas: das medidas de resposta à crise do Coronavírus para proteção de renda de informais e pessoas de baixa renda (TC 016.827/2020-1); das medidas de resposta à crise do Coronavírus na arrecadação tributária e previdenciária federal (TC 016.841/2020-4); e das medidas de resposta à crise do Coronavírus na Previdência Social (TC 016.830/2020-2).

7. Em uma análise preliminar realizada pela equipe de fiscalização do TC 016.841/2020-4, também desta Secretaria de Controle Externo, foram apontados indícios de inconsistências no Cadastro de Pessoas Físicas (CPF) que ensejariam aprofundamento nas análises por meio da aplicação de técnicas de auditoria com uso de análise de dados e de tecnologia da informação. Conforme item 9.1.5 do Acórdão 1.638/2020-TCU-Plenário, de relatoria do Ministro Bruno Dantas, tal aprofundamento se dá no presente TC 016.834/2020-8. Assim, este segundo relatório de acompanhamento de dados foca na qualidade das informações do CPF da Secretaria da Receita Federal do Brasil (RFB).

Deliberação

8. Este trabalho, que tem como fundamento o Plano especial de acompanhamento das ações de combate à Covid-19, foi aprovado na sessão plenária do dia 25/3/2020 (TC 016.602/2020-0).

Objetivo e escopo

9. O presente trabalho tem por objetivo a focalização da atuação dos órgãos e instituições públicas na luta contra a pandemia. O número de CPF é a principal chave para a inscrição dos beneficiários do auxílio emergencial, bem como a principal chave utilizada nos cruzamentos de dados para a aprovação do benefício e controle da elegibilidade. Dessa maneira, ao avaliar a qualidade da base de dados do CPF, objetiva-se, também, contribuir para o fortalecimento do controle do auxílio emergencial.

10. Além disso, o CPF é tido como principal meio de identificação dos cidadãos brasileiros, conforme disposto no Decreto 9.723/2019, e, na prática, de registros públicos e privados do País. Desta forma, a melhoria da qualidade dos dados do CPF pode contribuir para o fortalecimento da identificação de pessoas, afetando diversas políticas públicas e relações privadas, como por exemplo, o Sistema Financeiro Nacional, que utiliza o CPF como chave primária das transações de pessoas físicas.

11. O escopo do trabalho foi a base de dados do CPF da RFB de junho de 2020. Conforme Acordo de Cooperação Técnica entre os órgãos (peça 66), as informações de CPF são enviadas mensalmente ao Tribunal de Contas da União pelo Serviço Federal de Processamento de Dados (Serpro), referentes ao Contrato TCU 46/2019, celebrado para a prestação de serviços técnicos de informática para acesso e fornecimento periódico de dados das bases do CPF e do Cadastro Nacional de Pessoas Jurídicas (CNPJ) (peça 67).

Limitações

12. Apesar do campo de título de eleitor constar da estrutura da base de CPF e do convênio assinado entre TCU e RFB, o Contrato TCU 46/2019 informa que “mesmo constando no convênio o campo ‘título de eleitor’ não será fornecido”, sem especificar as motivações (peça 67, p. 37).

13. Conforme Nota Técnica Cocad/RFB 105, de 21 de agosto de 2020 (peça 82), encaminhada em resposta ao Relatório Preliminar de fiscalização, a RFB informou que não há limitação específica para o não fornecimento do dado de título de eleitor ao TCU. A limitação seria da própria distribuição do dado em geral e originar-se-ia de cláusula do convênio entre a RFB e o TSE, que impediria o compartilhamento do dado com terceiros.

14. No entanto, a limitação trata da ausência da disponibilização do título de eleitor detido pela RFB relativo a um CPF. Esse título de eleitor pode advir de documentação apresentada na inscrição no CPF ou de Declaração de Imposto de Renda de Pessoa Física (DIRPF), por exemplo, ou mesmo de processos de qualificação da base utilizando dados do Tribunal Superior Eleitoral. Desta forma, não se trata de repasse de dados originários do TSE, mas sim dos dados custodiados pela própria RFB.

15. Ainda assim, ressalta-se que o TCU possui Acordo de Cooperação Técnica com o TSE para acesso ao Cadastro Eleitoral (peça 83), e que mantém contato com os órgãos gestores das informações na tentativa de que tais dados sejam disponibilizados para futuros trabalhos de controle externo.

16. Dessa maneira, não foi possível realizar testes de acurácia, que consistem em comparar os dados da base do CPF com outras fontes de informações externas, como a base dos eleitores do Tribunal Superior Eleitoral, devido à ausência do campo chave de Título de Eleitor na base CPF.

Metodologia

17. Os trabalhos foram realizados em conformidade com as Normas de Auditoria do TCU – NAT (Portaria 280, de 8 de dezembro de 2010, alterada pela Portaria 168, de 30 de junho de 2011) e com o Manual de Acompanhamento do TCU (Portaria Segecex 27, de 9 de dezembro de 2016).

18. A metodologia adotada nesta fiscalização compreendeu a realização das etapas de: entendimento do negócio e dos dados; preparação dos dados, que inclui a avaliação de qualidade (credibilidade) dos dados, a higienização e o enriquecimento das bases e a construção de bases qualificadas; a elaboração de modelos de análise, que abrange o desenvolvimento de tipologias, a identificação de padrões de regularidade e irregularidade; a avaliação dos resultados; e as etapas de conclusão e elaboração de propostas de encaminhamento.

19. A credibilidade pode ser entendida como o grau de confiança que o auditor pode ter sobre as informações registradas, utilizando-se de critérios de qualidade de dados. A análise de credibilidade empregada neste trabalho é composta por seis atributos, detalhados na Seção 3.1: completude, unicidade, validade, consistência, acurácia e uniformidade.

20. A construção de bases qualificadas utiliza técnicas de higienização e de enriquecimento para produzir bases derivadas com dados de maior qualidade, que serão utilizadas para análises e cruzamentos de dados nas etapas seguintes. Já as tipologias ou trilhas de auditoria correspondem a cruzamentos de bases de dados e filtros específicos para verificar se a legislação pertinente ao tema fiscalizado está sendo observada pelos responsáveis pela política pública ou cadastro (critérios de auditoria).

21. A descrição pormenorizada dos procedimentos realizados e das bases utilizadas como insumo da fiscalização pode ser encontrada no Apêndice I – Metodologia detalhada.

2. Visão geral

22. O CPF é um banco de dados gerenciado pela RFB, que armazena informações cadastrais de contribuintes obrigados à inscrição no CPF, ou de cidadãos que se inscreveram voluntariamente. Conforme relatado, o trabalho utilizou a base referente ao mês de junho de 2020, enviada pelo Serpro ao TCU conforme o Contrato TCU 46/2019.

23. Em junho de 2020, a base recebida possuía 254,7 milhões de inscrições do CPF, das quais 226,9 milhões estavam em situação “Regular”, conforme indica a Tabela 1.

Tabela 1: Quantidade de inscrições de CPF por situação cadastral

Código	Situação Cadastral	Quantidade
0	Regular	226.957.798
2	Suspensa	8.867.821
3	Titular falecido	16.848.026
4	Pendente de regularização	113.205
5	Cancelada por multiplicidade	1.856.002
8	Nula	18.030
9	Cancelada de ofício	133.229
	Total	254.794.111

Fonte: elaboração própria.

24. Conforme os artigos 17 e 21 da IN RFB 1.548/2015, a inscrição é “Suspensa” quando ocorrer inconsistência cadastral, “Pendente de Regularização” quando houver omissão na Declaração do Imposto de Renda da Pessoa Física (DIRPF), “Cancelada por multiplicidade” quando houver mais de uma inscrição no CPF para a mesma pessoa, “Titular falecido” quando for incluído o ano de óbito e “Nula” quando for constatada fraude. Segundo o artigo 16 da mesma IN, a inscrição no CPF é “Cancelada de Ofício” nas hipóteses de: atribuição de mais de um número de inscrição para uma mesma pessoa física; por decisão judicial; ou por decisão administrativa.

25. Para os testes de credibilidade foram selecionados apenas os registros com situação cadastral “Regular” e para as tipologias foram utilizadas as inscrições com situação cadastral “Regular”, “Pendente de regularização” e “Suspensa”.

3. Credibilidade

26. As informações do banco de dados foram avaliadas quanto à credibilidade, ou seja, o grau de confiança que o auditor pode ter sobre as informações registradas. A análise de credibilidade empregada é composta por seis atributos: completude, unicidade, validade, consistência, acurácia e uniformidade.

27. A **completude** é a verificação da existência de registros com dados faltantes. Há casos em que essa falta é justificável, por exemplo, pode ser que o campo “Título de Eleitor” não precise estar preenchido para todos os registros da base.

28. A **unicidade** do campo verifica se existe duplicidade na chave primária (simples ou composta) da tabela. Alguns exemplos são o número do CPF ou do Cadastro Nacional de Pessoas Jurídicas (CNPJ) nas bases da RFB e a chave composta por Código de Prefeitura, Código de Família e Código de Pessoa na base do Cadastro Único, do Ministério da Cidadania.

29. A **validade** busca identificar se o esquema do banco de dados está sendo respeitado. Verifica-se, por exemplo: se campos que deveriam ser numéricos estão sendo disponibilizados dessa forma; se números de CPF e título de eleitor respeitam suas regras de formação (algoritmo para dígitos verificadores); se campos do tipo “data” apresentam apenas datas válidas; ou se campos com códigos (como “sexo” e “estado civil”) apresentam apenas códigos existentes.

30. A **consistência** é verificada a partir de testes que confrontem diferentes variáveis do banco de dados, que, em tese, devem trazer a mesma informação, como o Código de Endereçamento Postal (CEP), bairro e endereço. Nesses casos, os resultados podem ser que todas as informações prestadas correspondem a uma só situação de fato ou que há problemas de consistência (campos de data de nascimento posterior à data de inscrição no CPF, por exemplo).

31. Também se considera teste de consistência a verificação de regras condicionais de campos inter-relacionados. Se determinado campo de um banco de dados cadastral deve ser preenchido com 1 ou 0, significando que a pessoa identificada sabe ou não o nome de sua mãe, é considerado um erro de consistência a resposta de “não saber” com algum nome informado no campo de “nome da mãe”, bem como a resposta de “saber” com o campo “nome da mãe” nulo.

32. A **acurácia** dos dados é verificada por meio de testes que confrontem os dados obtidos com outras fontes de informação. Números de CPFs e informações cadastrais (“nome”, “nome da mãe” e “data de nascimento”) do CadÚnico ou do título de eleitor podem ser comparados com seus equivalentes nos bancos de dados da RFB, por exemplo.

33. A **uniformidade** busca verificar se campos com informações quantitativas mantêm um padrão de medida. Esses testes podem ser feitos em campos numéricos monetários (descrições em reais, centavos, dólares, cruzeiros) ou baseados em outras unidades (quilo, litro, resma, etc.), conforme o caso.

3.1. Verificação da unicidade

34. O número da inscrição do CPF foi testado como chave primária e apresentou unicidade.

3.2. Verificação da uniformidade

35. Não foram realizados testes de uniformidade, pois não existem campos com unidades monetárias ou com unidades de medida.

3.3. Verificação da acurácia

36. Conforme registrado nas “Limitações”, não foram realizados testes de acurácia na base de CPF, tendo em vista a ausência de outras chaves fortes na base encaminhada, como o título de eleitor detido pela RFB. Considerou-se que o confronto dos dados da pessoa física com outras fontes apenas pelos campos de identificação naturais (nome, data de nascimento e nome da mãe) teria taxa de incerteza maior que a tolerável.

3.4. Verificação da completude, validade e consistência

37. Os atributos de completude, validade e consistência foram analisados por meio da verificação das regras de preenchimento. Os campos que não foram testados, por exemplo, aqueles com preenchimento facultativo, estão marcados com o sinal “-” na Tabela 2, que apresenta os resultados da análise em números absolutos e em percentual.

Tabela 2: Análise de falhas de completude, validade e consistência

Campo	Completude		Validade		Consistência	
	Qtd	%	Qtd	%	Qtd	%
NUM_CPF	0	0,00	0	0,00	-	-
NOME	1	0,00	2.909	0,00	-	-
DATA_NASCIMENTO	9.567	0,00	14.027	0,00	9.700	0,00
IND_SEXO	32.688	0,01	32.688	0,01	0	0,00
NOME_MAE	9.378.708	4,14	174.830	0,08	-	-
NUM_TITULO_ELEITOR	-	-	-	-	-	-
TIPO_LOGRADOURO	0	0,00	-	-	-	-
DESCR_LOGRADOURO	1.058.399	0,47	-	-	-	-
NUM_LOGRADOURO	26.398.379	11,64	-	-	-	-
DESCR_COMPLEMENTO_LOGRADOURO	-	-	-	-	-	-
NOME_BAIRRO	7.908.131	3,49	-	-	-	-
NUM_CEP	0	0,00	22.185.301	9,78	0	0,00
NOME_MUNICIPIO	0	0,00	931.053	0,41	0	0,00
SIGLA_UF	0	0,00	0	0,00	0	0,00
NUM_DDD	-	-	-	-	-	-
NUM_TELEFONE	-	-	-	-	-	-
NUM_FAX	-	-	-	-	-	-
SE_ESTRANGEIRO	0	0,00	0	0,00	-	-
NOME_PAIS_NACIONALIDADE	-	-	0	0,00	8.264	0,00
COD_SITUACAO_CADASTRAL	0	0,00	0	0,00	0	0,00
DESCR_SITUACAO_CADASTRAL	0	0,00	0	0,00	0	0,00
DATA_SITUACAO_CADASTRAL	-	-	0	0,00	0	0,00
DATA_INSCRICAO	53.235.386	23,48	53.235.386	23,48	53.235.519	23,48
ANO_OBITO	-	-	0	0,00	0	0,00

Campo	Compleitude		Validade		Consistência	
	Qtd	%	Qtd	%	Qtd	%
ANO_ULTIMA_ENTREGA_DECLARACAO	-	-	0	0,00	0	0,00

Fonte: elaboração própria.

38. Parte dos problemas identificados são de menor risco, seja pelo baixo percentual de inconsistências ou pela importância da informação. As situações de maior importância são relatadas nos tópicos a seguir.

3.4.1. Registros de nome da mãe com problema de preenchimento

39. Em relação ao campo de nome da mãe (NOME_MAE), o teste de completude identificou 4,14% dos registros em branco ou nulos. Segundo Nota Cocad/RFB nº 89/2020, trata-se de um problema de legado e há ações da RFB para preencher este campo, como o povoamento das informações a partir do título de eleitor realizado em abril de 2020:

a) O campo “nome da mãe” foi inserido no CPF em 1993 para novas inscrições e alterações; portanto, as inscrições anteriores a essa data (legado) –e que não foram atualizadas pelo cidadão –permaneceram com o campo em branco.

b) O procedimento adotado (Povoamento do campo nome da mãe) baseou-se em pesquisa nos sistemas da RFB para identificar casos de contribuintes com inscrição CPF nas seguintes situações:

- 1) Regular, Pendente de Regularização ou Suspensa;
- 2) sem nome de mãe informado (em branco);
- 3) com título de eleitor vinculado;
- 4) com nome de mãe informado na réplica do cadastro eleitoral do Tribunal Superior Eleitoral (TSE);
- 5) com “Nome” e “Data de Nascimento” iguais nos cadastros da RFB e do TSE.

c) A partir dessas informações, foi gerada lista para alimentar o CPF com nome da mãe constante no cadastro eleitoral. Esse procedimento foi realizado de ofício por intermédio de apuração especial.

40. Para o mesmo campo, o teste de validade identificou 0,08% de registros inválidos. Uma das causas é a forma de preenchimento que pode ser utilizada caso não exista a informação do nome da mãe. Por exemplo, são aceitas informações como “MAE INEXISTENTE”, “NAO CONSTA”, “DES CONHECIDA”. “IGNORADO”, “XXX”, dentre muitas outras.

3.4.2. Registros sem a data de inscrição

41. O procedimento de credibilidade com maior quantidade de inconsistências é a completude do campo DATA_INSCRICAO, havendo 53,2 milhões de registros (23,48%) sem essa informação. O campo indica a data em que a pessoa foi cadastrada na base CPF e é importante para comparar com diversas outras datas como data de nascimento, data de criação do título de eleitor, data de óbito, dentre outras. Tais comparações são baseadas na credibilidade do banco de dados e utilizadas para evidenciar trilhas de eventuais erros ou fraudes.

42. Buscando as razões para a falta, elaborou-se a distribuição de preenchimento dessa informação ao longo do tempo, conforme consta na Tabela 3.

Tabela 3: Distribuição de CPFs por data de inscrição

Ano da DATA_INSCRICAO	Quantidade
1990 a 1999	39.201.460
2000 a 2009	59.018.683

2010 a 2019	72.260.273
2020	3.241.996
sem informação	53.235.386
Total CPFs regulares	226.957.798

Fonte: elaboração própria.

43. A Tabela 3 estratifica a quantidade de registros pelos anos da data de inscrição, evidenciando-se, assim, a inexistência de preenchimento com datas anteriores ao ano de 1990.

44. Essa conclusão foi ratificada pela Nota Técnica Cocad/RFB 105, de 21 de agosto de 2020 (peça 82), encaminhada em resposta ao Relatório Preliminar de fiscalização, que informou que o campo de “data de inscrição” foi incluído na base em 10 de novembro de 1990.

45. Pela análise dos dados, no entanto, restaria possível registros de pessoas nascidas a partir de 1990 sem a data de inscrição e contidos nos 53,2 milhões problemáticos. Essa possibilidade caracterizaria que o problema continuou a acontecer depois de 1990 e poderia estar acontecendo até os dias atuais. Para elucidar essa questão, elaborou-se a estratificação pelo ano de nascimento dos 53,2 milhões de registros sem data da inscrição, conforme consta na Tabela 4.

Tabela 4: Distribuição de CPFs sem data de inscrição por data de nascimento

Ano do nascimento	Quantidade
1861 a 1989	53.228.344
1990 a 1999	142
2000 a 2009	8
2010 a 2019	5
2020	-
sem informação do ano de nascimento	6.887
total	53.235.386

Fonte: elaboração própria.

46. A análise minimiza a possibilidade de o problema estar acontecendo atualmente, dado o baixo quantitativo de casos para anos mais recentes (155 registros).

47. Em que pese a importância do campo data da criação do CPF, a situação caracteriza-se como problema de legado, sendo de difícil correção.

3.4.3. Registros de endereço com problemas de preenchimento

48. Em relação aos campos de endereço, o campo de número do logradouro (NUM_LOGRADOURO) possui inconsistências em 11,64% dos registros. Ao analisar os valores preenchidos, constata-se que em diversos casos os dados desse campo foram preenchidos no campo DESCR_COMPLEMENTO_LOGRADOURO.

49. O campo de nome do bairro (NOME_BAIRRO) possui inconsistências em 3,49% dos registros. Constata-se uma correlação entre a ausência da informação para os endereços de zona rural, contudo em diversos registros o campo está preenchido como “zona rural” indicando uma falha na padronização do preenchimento.

50. O campo de número do CEP (NUM_CEP) possui inconsistências em 9,78% dos registros. Esse percentual elevado de inconsistência não é adequado e indica que o CEP registrado na base está diferente da faixa de CEP indicada pelos Correios para o município. Problemas de comunicação entre órgãos públicos e cidadão podem advir dessa falha. Contudo, considerando a ausência de um campo para informar a data da última alteração do endereço na base CPF, a quantidade e a idade dos registros, o índice de inconsistência pode advir das alterações no CEP realizadas pelos Correios.

3.4.4. Ausência de informações de título de eleitor

51. O campo de título de eleitor (NUM_TITULO_ELEITOR), apesar de constar no modelo de dados da base enviado em resposta ao Ofício de Requisição 02-70/2020 (peça 30), não foi objeto dos testes de credibilidade. O campo faz parte das informações constantes no convênio assinado entre TCU e RFB e era fornecido anteriormente, porém, segundo contrato de prestação de serviços com o Serpro “mesmo constando no convênio o campo ‘título de eleitor’ não será fornecido”.

52. Cabe relatar que chaves estrangeiras, como o título de eleitor para o CPF, são de elevada importância para procedimentos de qualificação de dados (*Data Quality*), sendo muito utilizadas em processos de confirmação da identidade de cidadãos, como nas Fiscalizações Contínuas de Benefícios do TCU, e em processos de *Master Data Management* (MDM) como o realizado recentemente pelo INSS no Cnis (Cadastro Nacional de Informações Sociais).

53. O TCU já se posicionou sobre a necessidade de maior integração entre os cadastros de pessoas físicas, como na seguinte deliberação do Acórdão 1.123/2020-TCU-P:

9.9. recomendar, com fundamento no art. 250, inciso III, do Regimento Interno/TCU, ao Comitê Central de Governança de Dados, conforme previsto no art. 21 do Decreto 10.046/2019, à Receita Federal do Brasil e ao Tribunal Superior Eleitoral que envidem esforços para fornecimento dos motivos de suspensão do CPF, bem como do Título de Eleitor associado a um CPF (a exemplo de sua inclusão na solução **blockchain** de CPF da Receita Federal do Brasil), de forma a promover a qualidade e interoperabilidade dos dados relativos aos cidadãos e otimizar a implementação de políticas públicas, nos moldes do previsto na Lei 13.444/2017, que instituiu a Identificação Civil Nacional, e do Decreto 10.046/2019, que criou o Cadastro Base do Cidadão (itens 2.6 e 2.7 do relatório de auditoria);

54. A recomendação será acompanhada no monitoramento do Acórdão 1.123/2020-TCU-P, ainda não iniciado. Aqui, reitera-se a necessidade de que sejam envidados esforços dos órgãos responsáveis pelas principais bases cadastrais de pessoas físicas para que haja interoperabilidade entre as bases, otimização de políticas públicas e identificação única de cada cidadão.

Comentários do gestor

55. Conforme Nota Técnica Cocad/RFB 105, de 21 de agosto de 2020 (peça 82), encaminhada em resposta ao Relatório Preliminar de fiscalização, a RFB informou que não há limitação específica para o não fornecimento do dado de título de eleitor ao TCU. A limitação seria da própria distribuição do dado em geral e originar-se-ia de cláusula do convênio entre a RFB e o TSE. O Título de Eleitor constante do CPF não poderia ser compartilhado com terceiros, por força das cláusulas segunda e terceira constantes do convênio celebrado entre RFB e o Tribunal Superior Eleitoral (TSE) em 26 de julho de 2002:

“CLÁUSULA SEGUNDA –Os convenientes se dispõem a fornecer, reciprocamente, informações de interesse da Secretaria da Receita Federal e da Justiça Eleitoral, constantes de seus arquivos e bancos de dados, conforme segue:

II –DO TSE PARA A SRF:

a) dados constantes do Sistema de Alistamento Eleitoral (SAE –Cadastro de Eleitores);

(...)

CLÁUSULA TERCEIRA –Os convenientes se comprometem a utilizar os dados que lhes forem fornecidos somente nas atividades que, em virtude de lei, lhes compete exercer, não podendo transferir a terceiros as informações econômicos-fiscais ou eleitorais apresentadas de forma individualizada, seja a título oneroso ou gratuito, ou, de qualquer forma, divulgá-los, sob pena de extinção imediata deste Convênio, ressalvada a hipótese de deliberação do TSE, em contrário, relativamente as informações que receber da SRF.”

56. Ainda segundo o gestor, por esse motivo o título de eleitor também não compõe o leiaute atual da b-CPF. Destaca, também, que o descumprimento desta obrigação pode ocasionar a denúncia imediata do convênio, razão pela qual o dado não foi incluído nas distribuições do CPF.

Análise dos comentários do gestor pela equipe do TCU

57. Com relação à ponderação do gestor, cabe destacar que a limitação tratou da ausência da disponibilização do título de eleitor detido pela RFB relativo a um CPF. Esse título de eleitor pode advir de documentação apresentada na inscrição no CPF ou de Declaração de Imposto de Renda de Pessoa Física (DIRPF), por exemplo, ou mesmo de processos de qualificação da base utilizando dados do Tribunal Superior Eleitoral. Desta forma, não se trata de repasse de dados originários do TSE, mas sim dos dados custodiados pela própria RFB.

58. Ainda assim, ressalta-se que o TCU possui Acordo de Cooperação Técnica com o TSE para acesso ao Cadastro Eleitoral (peça 83), e que mantém contato com os órgãos gestores das informações na tentativa de que tais dados sejam disponibilizados para futuros trabalhos de controle externo.

59. Conforme relatado anteriormente, as chaves unívocas são de elevada importância para processos de qualificação de dados, como os empreendidos pelo TCU e por diversos outros órgãos da Administração Pública, sendo essenciais para a concretização de projetos de identificação única do cidadão e sua fiscalização.

3.5. Conclusão

60. Os campos da base do CPF apresentam um alto percentual de credibilidade para completude, validade e consistência. Na presente auditoria, optou-se por um percentual máximo tolerável de 2% para falhas ou inconsistências.

61. Constata-se que campos referentes a endereço possuem percentual de inconsistências maior que o tolerável, contudo esses campos possuem poucos incentivos para sua atualização pelo cidadão, especialmente no caso de pessoas físicas que não declaram Imposto de Renda.

62. O campo referente à data da criação do CPF possui o maior percentual de inconsistências, com 23,4% dos registros sem preenchimento. Contudo, verifica-se que a maior parte diz respeito a registros de nascimentos anteriores a 1990. Considerando que essa situação afetaria registros com mais de 30 anos e, normalmente, os registros com data de criação mais recentes são utilizados em casos de fraude (ex: criação de uma inscrição de CPF para a solicitação de um Benefício de Prestação Continuada ou de uma aposentadoria rural), pode-se afirmar que a situação possuiria baixo risco e parte dessas inconsistências poderia ter origem na evolução do banco de dados ou na implementação de novas regras e funcionalidades, sendo de difícil correção após 30 anos do fato.

4. Tipologias

4.1. Registro de óbito em bases da Administração Pública federal ou estadual

63. O objetivo desta tipologia é identificar indícios de óbito em outras bases da Administração Pública Federal e Estadual e comparar com a situação da inscrição na base de CPF. Para a execução dos testes foram utilizadas as seguintes fontes de informação:

- a) Sistema Nacional de Informações de Registro Civil (Sirc), referência de agosto de 2019;
- b) instituidor de pensão por morte da Folha de Pagamentos do INSS (Maciça), referência de junho de 2020;

c) Cadastro Único (Cadun), com filtro para os registros marcados com indício de óbito, referência de junho de 2020;

d) folha de pagamentos de Estados e Municípios, encaminhadas ao TCU pelos respectivos Tribunais de Contas Estaduais (TCE), referências de janeiro de 2019 até março de 2020;

e) Sistema Informatizado de Controle de Óbitos (Sisobi), referência de maio de 2020;

f) Sistema Integrado de Administração de Recursos Humanos (Siape), referência de junho de 2020; e

g) folhas de pagamento dos Poderes Legislativo e Judiciário, Banco Central e Forças Armadas (Extrasiape), referências de janeiro de 2019 até março de 2020.

64. Em relação aos registros da base CPF, foram considerados apenas os registros em situação “Regular”, “Suspensa” e “Pendente de Regularização”.

65. As bases confrontadas com o CPF passaram por procedimentos de qualificação de dados (*Data Quality*) de pessoas, para mitigar os riscos conhecidos de problemas de identificação, conforme metodologia detalhada no Apêndice I.

Situação encontrada:

66. Foram identificadas 3.359.609 inscrições do CPF com indícios de óbito em bases da Administração Pública Federal ou Estadual com situação cadastral “Regular”.

67. A Tabela 5 apresenta os resultados por base e por situação da inscrição do CPF.

Tabela 5: Resultado da tipologia por base e situação do CPF

Situação cadastral	Sirc	Maciça	Extrasiape	TCE	Cadun	Sisobi	Siape	Total (CPFs distintos)
Regular (0)	1.611.519	990.455	31.135	11.935	173.552	2.591.739	82.979	3.359.609
Suspensa (2)	17.381	218.457	20.434	989	1.119	460.189	108.499	763.216
Pendente de regularização (4)	10.502	1.553	618	136	125	11.249	1.921	12.757

Fonte: elaboração própria.

68. Em certos casos, a inscrição de CPF foi identificada com indício de óbito em mais de uma base, por exemplo, no Sirc e no Sisobi. Dessa maneira a coluna Total foi elaborada para evitar a dupla contagem, considerando apenas os CPFs distintos. Note-se que essa repetição é relevante porque pode indicar uma maior certeza quanto ao óbito. No entanto, para fins de avaliação da necessidade de ajuste do CPF, cada pessoa só deve ser contabilizada uma vez e, por isso, a linha de totalização considera apenas as ocorrências distintas.

Critério:

69. A IN RFB n. 1.548 de 12/02/2015, art. 21 e 17, dispõe que:

Art. 21. A inscrição no CPF será enquadrada, quanto à situação cadastral, em:

I - regular, quando não houver inconsistência cadastral e não constar omissão de DIRPF;

II - pendente de regularização, quando houver omissão de DIRPF;

III - suspensa, quando houver inconsistência cadastral;

V - titular falecido, quando for incluído o ano de óbito;

VII - nula, nos termos do art. 17.

Art. 17. Será declarada nula pela RFB a inscrição no CPF em que for constatada fraude.

Evidências:

70. As evidências estão nos arquivos:
- a) obitos_cadun.txt
 - b) obitos_extrasiape.txt
 - c) obitos_macica.txt
 - d) obitos_siape.txt
 - e) obitos_sirc.txt
 - f) obitos_sisobi.txt
 - g) obitos_TCE.txt
 - h) resultado_consolidado.txt
71. Os *scripts* utilizados estão nos arquivos:
- a) obitos_cadun.sql
 - b) obitos_extrasiape.sql
 - c) obitos_macica.sql
 - d) obitos_siape.sql
 - e) obitos_sirc.sql
 - f) obitos_sisobi.sql
 - g) obitos_TCE.sql
 - h) resultado_consolidado.sql

Causas:

72. Insuficiência das verificações realizadas quanto à existência de indícios de falecimento da pessoa em outras bases de dados, como Sisobi, Sirc, INSS, CadÚnico, INSS e folhas de pagamento de agentes públicos; possíveis falhas na identificação de pessoas em certidões de óbito registradas nas bases do Sisobi e do Sirc; limitado compartilhamento de bases de dados entre órgãos.

4.2. Registro de supercentenários

73. O objetivo desta tipologia é identificar indícios de pessoas com idade flagrantemente incorreta, como pessoas com 200 anos de idade e que estão com situação regular na base do CPF, ou pessoas muita acima da máxima idade já alcançada por seres humanos.

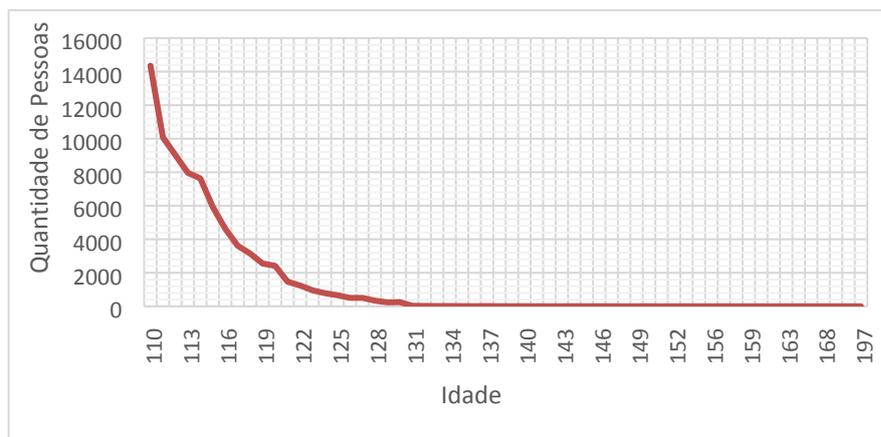
Situação encontrada

74. Segundo definição da *Gerontology Research Group* (GRG), entidade internacional de pesquisa e autenticidade de pessoas velhas, supercentenários são pessoas que, comprovadamente, viveram mais de 110 anos. A GRG também é a responsável por autenticar o recorde mundial de pessoa mais longeva, registrado pelo *Guinness World Record*, publicação que verifica os recordes mundiais em diversas categorias. Segundo o *Guinness*, a maior idade já alcançada por um ser humano foi de 122 anos.

75. Nas análises empreendidas na base de CPF, foram identificadas 5.699 pessoas com idade superior a 122 anos. Desse quantitativo, foram excluídos os casos graves de erro de entrada de dados, como data de nascimento no ano 200, ou 1012.

76. O Gráfico 1 apresenta a curva de frequência de pessoas supercentenárias, por idade, identificada na base do CPF. Registre-se que foram verificadas 72.817 pessoas com idade entre 110 e 122 anos em situação regular. De acordo com os dados atualizados pela GRG em 30/7/2020, atualmente, existem 29 pessoas comprovadamente supercentenárias vivas, sendo apenas uma delas no Brasil (<http://supercentenarian-research-foundation.org/TableE.aspx>).

Gráfico 1: Histograma da quantidade de pessoas supercentenárias com CPF regular por idade



Fonte: dados de CPF, consolidados pela equipe do TCU.

77. Questionada acerca de mecanismos de prova de vida, a RFB informou (peça 82) que não dispõe de sistema próprio para essa finalidade, porém está efetuando estudos técnicos utilizando a plataforma App CPF Digital e avaliando programa com esse fim da Secretaria de Governo Digital (SGD) do Ministério da Economia e INSS.

78. Apesar de não haver necessidade de prova de vida para o cadastro do CPF, é necessário que a RFB crie outros mecanismos de controle para identificar a ocorrência de óbitos além do convênio com a Arpen/Brasil, modificando sua situação cadastral e evitando a utilização de números de CPF para perpetração de possíveis fraudes.

Critério:

79. Registro de pessoas com idade acima de 122 anos, que corresponde ao recorde mundial de pessoa mais velha registrado pelo *Guinness World Record*.

Evidências:

80. As evidências estão no arquivo `supercentenarios.txt`, gerado pelo `script cpf_supercentenarios.sql`.

Causas:

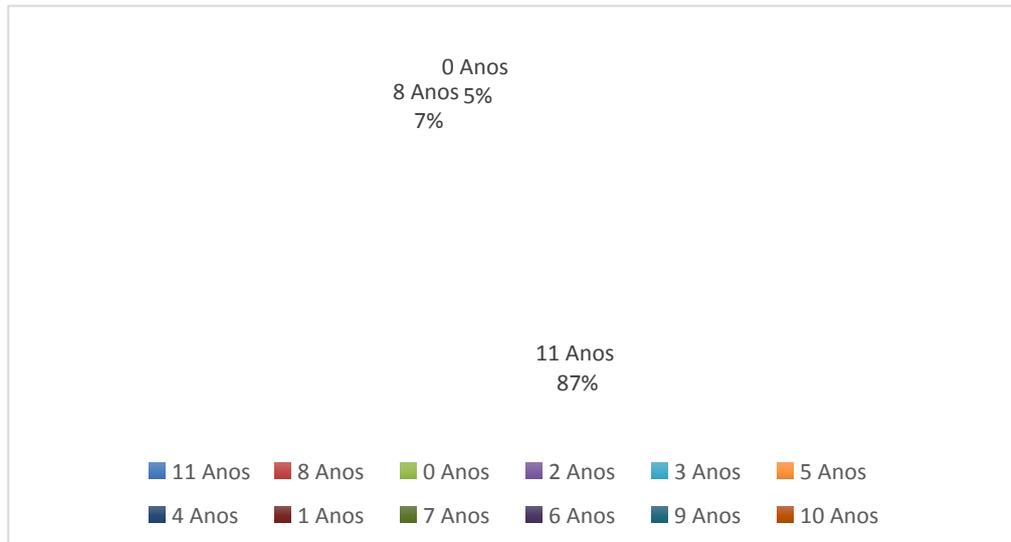
81. A falta de integração do CPF com outras bases públicas (como Sirc, Sisobi, Maciça, folhas de pagamento de agentes públicos e CadÚnico) pode estar gerando lapso temporal entre atualizações cadastrais, principalmente quanto a dados de óbito. Além disso, problemas de qualidade nos registros e comunicações de óbitos no Brasil também podem contribuir para a ocorrência dessa situação.

4.3. Inscrições de CPF suspensas há 11 anos ou mais

Situação encontrada

82. De 8.867.821 CPFs suspensos, identificou-se que 7.706.686 CPFs se encontram nessa situação há mais de 11 anos (87%), conforme estratificação apresentada no Gráfico 2.

Gráfico 2: CPFs suspensos por tempo na situação



Fonte: dados de CPF, consolidados pela equipe do TCU.

83. De acordo com a IN RFB 1.548/2015:

Art. 12. A suspensão da inscrição será realizada pela RFB quando houver **inconsistência cadastral**.

Parágrafo único. Será dada ciência da suspensão por meio do:

I - “Comprovante de Situação Cadastral no CPF”, conforme modelo constante do Anexo V desta Instrução Normativa, disponível no sítio da RFB na Internet, no endereço <http://www.receita.fazenda.gov.br>;

II - “Comprovante de Situação Cadastral no CPF” acessado por meio do aplicativo “APP Pessoa Física” para dispositivos móveis; ou

III - pelo serviço de atendimento telefônico da RFB.

(grifos nossos)

84. Não obstante a previsão da IN RFB 1.548/2015 de suspensão do CPF quando houver “inconsistência cadastral”, segundo informações na Nota Cocad/RFB 92, de 31 de julho de 2020 (peça 72), a suspensão de um CPF pode advir também de suspeitas de fraudes, conforme motivos apresentados na Tabela 6.

Tabela 6: Motivos de suspensão de CPF, em julho de 2020

Código	Motivo
1	Título de eleitor com nome e data de nascimento divergente entre base CPF e base TSE
2	Idade maior ou igual a 19 anos e sem título na base CPF
7	Espólio em qualquer parte do nome
9	Suspensão automática, sem título de eleitor e sem motivo de dispensa
10	Título de eleitor com nome da mãe divergente entre base CPF e base TSE

11	Título de eleitor inexistente na base TSE, com regularização na RFB
13	Regularização somente na RFB (suspensão de ofício)
15	CPF com indícios de fraude, com regularização exclusiva na RFB
17	Inconsistência de endereço
20	Título de eleitor vinculado a estrangeiro ou a pessoa menor de 16 anos
21	NI-CPF com nome da mãe não informado
23	Indício de existência de óbito
24	Indício de multiplicidade de inscrições no CPF

Fonte: Nota Cocad/RFB 92/2020 (peça 72).

85. Assim, verifica-se que a situação de suspensão, na prática, é utilizada tanto para identificar problemas cadastrais mais simples, como inconsistência de endereço, como problemas muito mais graves, como CPF com indícios de fraude ou com suspeita de óbito.

86. Ainda segundo a Nota Cocad/RFB 92/2020:

Esclarecemos, ademais, que, embora ocorra melhoria contínua da base, **não há, legalmente, definição de prazo máximo de permanência de um CPF na situação cadastral suspensa.**

O CPF é suspenso quando houver inconsistência cadastral, tais como, o dado informado na base da RFB diverge do constante em outras bases. Nesse caso, **a notificação ao cidadão se dá por meio do comprovante de situação cadastral, conforme dispõe o Art. 12, da Instrução Normativa RFB 1.548, de 13 de fevereiro de 2015.**

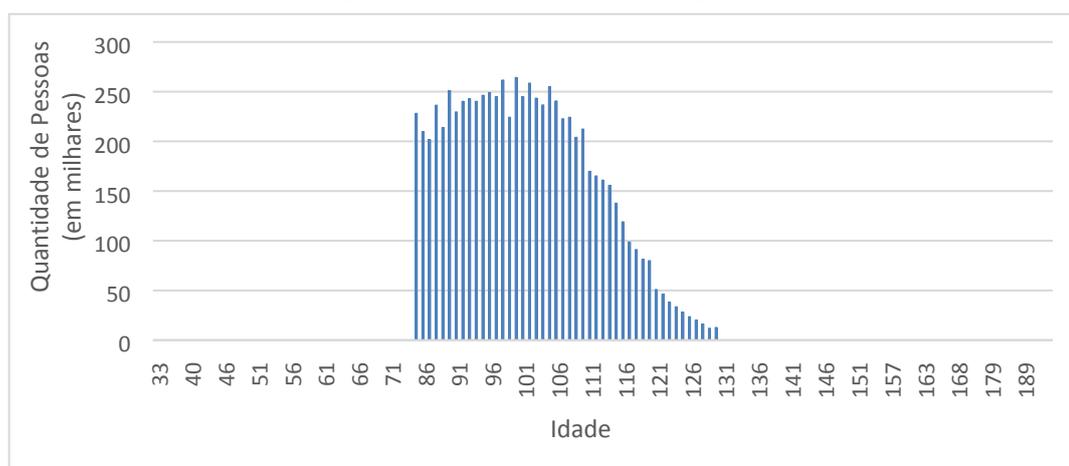
(grifos nossos)

87. Pelas informações do gestor, não há notificação ativa do cidadão em caso de suspensão de seu CPF, havendo a previsão somente de que ele possa consultar a situação pelo sítio eletrônico, aplicativo ou serviço de atendimento telefônico da RFB. Ainda, a RFB afirma que não haveria prazo máximo para um CPF permanecer suspenso, por ausência de previsão legal.

Análise

88. O Gráfico 3 apresenta uma estratificação por idade para CPFs suspensos há 11 anos ou mais, foram excluídas do gráfico as pessoas sem data de nascimento (34.324).

Gráfico 3: CPFs suspensos há mais de 11 anos por idade (em milhares)



Fonte: dados de CPF, consolidados pela equipe do TCU.

89. Percebe-se que há alta concentração de pessoas com idade avançada. Importante destacar que em todos os casos relatados, a data da situação cadastral é 8/3/2009, demonstrando que houve algum tipo de avaliação em massa da situação cadastral do CPF específica nessa data.

90. Em resposta a questionamento desta equipe de fiscalização, a RFB informou que na referida data houve apuração especial para criar novos motivos de suspensão – “CPF omissos até 2007

(ano base 2006) com ano de óbito” e “Título de Eleitor inexistente na base TSE ou com dados divergentes entre as bases CPF e TSE”) e adequar o cadastro ao novo conceito de situação suspensa constante do art. 21 da IN RFB 864/2008, que passou a considerar suspensão em caso de ocorrência de inconsistência cadastral (peça 82).

91. Como os motivos de suspensão trazem uma pluralidade de situações fáticas, em ciclos passados das Fiscalizações Contínuas da SecexPrevidência foram observados pagamentos de benefícios sociais para pessoas com CPF suspenso. Por exemplo, na Fiscalização Contínua de Benefícios Assistenciais de 2019, foram identificados 308.718 beneficiários do Programa Bolsa Família e 217.813 beneficiários do Benefício de Prestação Continuada (BPC) com o CPF suspenso.

92. Acerca dessas questões, o item 9.2 do Acórdão 12.162/2018-TCU-2ª Câmara, de relatoria do Min-subst. André Luis de Carvalho, determinou ao Ministério da Cidadania que fosse apresentado plano de ação para estabelecer as condições que devem ser observadas para manutenção da pessoa no Cadastro Único e pagamento de benefícios de acordo com a situação cadastral do CPF do beneficiário.

93. O Ministério da Cidadania, porém, já se manifestou na Nota Técnica 08/2020 DECAU/SAGI (peça 73, p. 1-8), afirmando que, por meio de tratativas junto à RFB, procurou receber as informações detalhadas de situação cadastral do CPF, via solução *blockchain*. Entretanto, em 16 de julho de 2019, a RFB, por meio do Ofício 41/2019 – RFB/COCAD (peça 74), informou sobre a impossibilidade de repassar os motivos de suspensão do CPF.

94. Dado que nem todos os casos são relativos a indícios de fraude, podendo o CPF estar suspenso por falta de alguma informação no cadastro da pessoa ou algum dado incorreto na RFB, o Ministério da Cidadania argumenta que tal informação impacta no tratamento a ser dado no ajuste do registro no Cadastro Único e na consequente manutenção ou não de benefícios.

95. Considerando os argumentos apresentados, o Acórdão 1.123/2020-TCU-P, de relatoria do Min-subst. Marcos Bemquerer trouxe recomendação no item 9.8.1 para que os órgãos conduzam as ações técnicas necessárias para compartilhamento de informações de cadastros de cidadãos, incluindo Título de Eleitor e motivo da situação do CPF, utilizando sistemática que garanta a interoperabilidade entre as bases, nos moldes do previsto na Lei 13.444/2017, que instituiu a Identificação Civil Nacional, e do Decreto 10.046/2019, que criou o Cadastro Base do Cidadão.

96. Nesse sentido, uma outra solução que poderia ser aventada seria o estudo de um convênio entre os órgãos para que os processos de revisão e atualização previstos para o Cadastro Único realizados pelos Centros de Referência de Assistência Social (Cras) pudessem ser aproveitados para regularização da situação do CPF ou, pelo menos, para o devido encaminhamento do cidadão às entidades competentes.

97. Destaca-se que já existem entidades conveniadas à RFB que prestam serviços de atendimento relativos ao CPF, inclusive para regularização da situação cadastral suspensa, conforme IN RFB 1.548/2015:

Anexo III da Instrução Normativa RFB nº 1.548, de 2015 (incluído pela Instrução Normativa RFB 1.746/2017)

CPF – ATENDIMENTOS NO BRASIL

Local de atendimento:

a) **Correios, Banco do Brasil ou Caixa Econômica Federal**, nos casos de inscrição, alteração e **regularização da situação cadastral suspensa**;

(grifos nossos)

98. Sobre a ausência de previsões legais específicas sobre prazo máximo para que o CPF permaneça suspenso e para notificação ativa do cidadão, entende-se que haveria a possibilidade de a própria RFB regulamentar os procedimentos cabíveis por instrumento infralegal, autorizada por legislações vigentes, conforme a Lei 9.799/1999:

Art. 16. Compete à Secretaria da Receita Federal dispor sobre as obrigações acessórias relativas aos impostos e contribuições por ela administrados, estabelecendo, inclusive, forma, prazo e condições para o seu cumprimento e o respectivo responsável.

99. E de acordo com o Decreto 9.580/2018:

Art. 32. As pessoas físicas ficam obrigadas a se inscrever no CPF, na forma, no prazo e nas condições estabelecidos pela Secretaria da Receita Federal do Brasil do Ministério da Fazenda (Lei nº 4.862, de 29 de novembro de 1965, art. 11 ; Decreto-Lei nº 401, de 1968, art. 1º e art. 2º ; e Lei nº 9.779, de 1999, art. 16).

100. Assim, há proposta para determinar à RFB a apresentação de plano de ação visando estabelecer procedimentos para notificação ativa do cidadão em caso de suspensão de seu CPF, bem como regulamentar prazo limite para um CPF constar na situação “suspensa” após a notificação do cidadão, incluindo no processo de trabalho a forma de tratamento desses CPFs após o prazo estabelecido.

101. Como motivação para a proposta, cabe relatar que, conforme noticiado pela imprensa no mês de abril de 2020, milhares de pessoas tiveram dificuldades de realizar o pedido do auxílio emergencial devido a inconsistências no CPF, levando a uma necessidade de regularização de ofício do CPF pela RFB também emergencial, tratada no Capítulo 5. Conforme também noticiado, muitas dessas pessoas só descobriram que o CPF estava com pendências ao tentar obter o benefício durante a pandemia e não conseguir.

Critérios

102. Lei 9.799/1999, art. 16; Decreto 9.580/2018, art. 32; IN RFB 1.548/2015, art. 12.

Evidências

103. As evidências estão no arquivo `cpf_tempo_suspenso.txt`, gerado pelo `script tempo_sit_cadastral.sql`.

Causas

104. Ausência de regulamentação de prazos para um CPF permanecer suspenso; ausência de mecanismos ativos de notificação do cidadão de que seu CPF foi suspenso; baixos incentivos para o cidadão regularizar a situação de seu CPF em situações normais; limitado compartilhamento de bases de dados e processos entre órgãos.

5. Regularização de ofício

105. Conforme noticiado pela imprensa no mês de abril de 2020, milhares de pessoas tiveram dificuldades de realizar o cadastro do auxílio emergencial devido a inconsistências no CPF (<https://g1.globo.com/bemestar/coronavirus/noticia/2020/04/09/contribuintes-fazem-filas-em-posto-da-receita-federal-para-regularizar-cpf-no-rio.ghtml>).

106. A RFB encaminhou a Nota Cocad 37/2020 (peça 70) para esclarecer os procedimentos realizados na base, em caráter excepcional, durante o enfrentamento da pandemia da Covid-19.

107. As inscrições de CPF “Suspensas” que possuíam inconsistências associadas à justiça eleitoral (divergências entre dados cadastrais da base do TSE e do CPF, ausência de título de eleitor, etc.) ou pendências menores (situação antiga de ausência de Declaração Anual de Isento e redefinição

de códigos municipais e de CEP) foram marcadas como “Regulares”. Nesse procedimento foram regularizadas 12,2 milhões de inscrições, conforme a Tabela 7.

Tabela 7: Ajustes em massa realizados na base de CPF, em abril de 2020

	1ª Lista	2ª Lista	Total
Regularizadas	11.700.150	581.482	12.281.632
Inscrições que estavam regulares	1.793.260	139.351	1.932.611
Inscrições em situação diferente de suspensa	68.258	537	68.795

Fonte: Nota Cocad 37/2020 (peça 70, p. 7).

108. Ainda conforme informações da RFB, foram regularizadas cerca de 480 mil inscrições com situação “Pendente de Regularização” por omissão de Declaração de Imposto de Renda de Pessoa Física (DIRPF) de pelo menos um ano. Nesse caso, foram regularizadas as inscrições com situação de renda mais baixa, de menor interesse fiscal. A RFB ressaltou que a regularização da inscrição é temporária e permanece a pendência para emissão de certidão de regularidade fiscal.

109. Segundo informado nas Notas Omissos IRPF 01 e 02/2020 da Codac/RFB e na Nota Cocad/RFB 92/2020 (peças 70 e 72), a alteração da situação cadastral de ofício para “Regular” foi temporária e a reversão já estaria em andamento via apuração especial, cujas ações seriam executadas em agosto de 2020. Apesar de solicitadas no Ofício de Requisição 02-70/2020 (peça 30), não foram fornecidas documentações detalhadas sobre o cronograma ou as ações que serão empreendidas pelo Serpro na reversão.

110. Em relação ao campo de nome da mãe, houve o povoamento de 210 mil registros a partir das informações do cadastro eleitoral. Conforme Nota Cocad/RFB 92/2020 (peça 72), “o povoamento do campo nome da mãe, a partir das informações do TSE, terá repercussão permanente na base CPF e não será revertido, pois agregará mais qualidade ao cadastro da RFB”.

111. Por fim, o atendimento pelos canais virtuais para a regularização da inscrição foi facilitado por meio de Caixas Corporativas (e-mail) e pelo Chat RFB.

112. As medidas tomadas pela RFB tiveram o objetivo de facilitar o acesso ao Auxílio Emergencial durante o enfrentamento da pandemia da Covid-19. Quando ocorrer a normalização da situação, deve-se realizar o acompanhamento da reversão das situações cadastrais e de como isso eventualmente poderá afetar a qualidade do cadastro.

6. Inscrições regulares do CPF superiores à população brasileira

113. O Acórdão 1.638/2020 – TCU – Plenário, de relatoria do Min. Bruno Dantas, dispõe que: 9.1.5. foram encontrados indícios de irregularidades graves na gestão da base de dados de CPF mantida pela Receita Federal, que contava com cerca de 12,5 milhões de registros ativos além da população brasileira estimada pelo IBGE para o mesmo período. Tendo em vista a necessidade de utilização de técnicas de auditoria com o auxílio de recursos de tecnologia da informação para aprofundar a avaliação dos indícios encontrados, a análise será realizada no TC 016.834/2020-8, relativo ao Acompanhamento Especial das medidas de resposta à crise do Coronavírus para as áreas de Previdência Social, Assistência Social e Administração Tributária com Análise de dados

114. Dessa forma, realiza-se a análise das respostas encaminhadas pela RFB referentes aos indícios apontados no TC 016.841/2020-4, relativo ao Acompanhamento Especial das medidas de resposta à crise do Coronavírus na arrecadação tributária e previdenciária federal.

115. O Relatório referente ao Acórdão 1.638/2020 informa que:

“Chama a atenção o fato de que, ao final do mês de abril de 2020, depois da regularização em massa efetuada pela Receita Federal, o número de cadastros de CPF na situação de regular, que correspondiam a 223.850.498 registros, superavam em mais de 12,5 milhões a população estimada pelo IBGE para o mesmo período, de pouco mais de 211,4 milhões de pessoas.

74. O crescimento do número de CPF registrados na Receita Federal já era esperado por diversos fatores. Até 2014, o registro no cadastro de CPF só poderia ser exigido para os maiores de 18 anos. Com a edição da Instrução Normativa RFB 1548 de 13/2/2015, e suas alterações posteriores, a idade mínima obrigatória para registro de dependentes, para fins do Imposto sobre a Renda da Pessoa Física, foi gradualmente reduzida para 16 anos em 2015, 14 anos em 2016, 12 anos em 2017 e, desde 2018, mantém-se no patamar de 8 anos.”

116. A Nota/Cocad nº 57/2020 (peça 69), em relação à emissão de novas inscrições e sobre sua baixa informa que:

É característica do processo de assimilação dos dados oriundos do Registro Civil que as inscrições advindas de nascimentos e as inscrições tardias (e.g. jovens sem CPF) são incorporadas em maior número e mais facilmente do que são expurgados os óbitos. Ademais, inscrições são efetuadas por diferentes conveniados (Caixa Econômica Federal, Banco do Brasil, Correios) e pela internet, diferentemente das “baixas” por óbito, que advém primordialmente do próprio Registro Civil.

Por um lado, as inscrições são fortemente incentivadas pelos benefícios advindos da inserção do indivíduo na base CPF. Por outro, a baixa por óbito, quando indevida, na prática, retira o indivíduo da vida civil, impedindo-o de peticionar direitos frente a diferentes órgãos. Portanto a “baixa” do CPF deve ocorrer com critérios de muito maior rigorismo. Por fim, os dados de óbito advindos do Registro Civil, e consolidados na Central de Informações do Registro Civil (CRC), tendem a apresentar maior dificuldade de identificação unívoca. A uma, pela qualidade desse dado (sobretudo aquele anterior a 2015); a duas, pela simples inexistência do dado, nos casos de serventias que não reportam o ocorrido à CRC.

117. Constata-se uma assimetria nos incentivos para uma pessoa emitir o CPF, devido a sua necessidade para participar de programas de distribuição de renda, como o Bolsa Família, exercer o direito de voto ou obter uma conta corrente no sistema bancário, em relação à comunicação do falecimento por familiares, que inclusive pode fazer com que um benefício seja cessado, como um benefício previdenciário.

118. Conforme informado na Nota Cocad 92/2017 (peça 69), em 2017, a RFB iniciou uma nova sistemática de consulta diária à base de óbitos da Associação Nacional dos Registradores de Pessoas Naturais do Brasil (Arpen/Brasil).

119. Ainda em 2017, conforme a Nota Cocad 57/2020 (peça 69) houve o tratamento dos legados das inscrições, com a padronização do preenchimento de “Cancelada por óbito sem espólio” e “Cancelada por encerramento de espólio” para “Titular Falecido” e a alteração da situação para “Titular Falecido” para as inscrições com o ano de óbito informado e que estivessem em situação cadastral “Regular”, “Pendente de Regularização” e “Suspensa”.

120. Em novembro de 2017, teve início a Ação Especial Escalonada (APES) que alterou situação cadastral dos CPFs “Regular” com óbito informado, para “Suspenso”; em seguida esses CPFs suspensos tiveram a situação cadastral alterada para “Titular falecido”, conforme cronograma da Tabela 8.

Tabela 8. Cronograma da Ação Especial Escalonada

Data da APES	Ano de Óbito	Data da APES	Ano de Óbito
nov/17	Até 2004	mai/18	2012
dez/17	2005-2007	jun/18	2013

jan/18	2008	jul/18	2014
fev/18	2009	ago/18	2015
mar/18	2010	set/18	2016
abr/18	2011	out/18	2017

Fonte: Nota Cocad 57/2020 (peça 69)

121. Sobre as iniciativas empreendidas pelo gestor para aprimoramento de dados do CPF, a Nota Cocad 57/2020 (peça 69) esclarece que:

O CPF, regido pela Instrução Normativa RFB nº 1.548, de 13 de fevereiro de 2015 vem sendo continuamente aprimorado pela Receita Federal (RFB), tornando-se um verdadeiro *númerocidadão*. Esse aprimoramento, gradual e consistente, materializa-se por meio de uma série de ações e investimentos que visam à expansão da quantidade de cadastrados e ao aumento da integridade dessa base.

Nesse diapasão, a RFB tem ampliado os canais de atendimento de CPF, trabalhado junto às entidades produtoras primárias do dado cadastral e disponibilizado a base para diferentes instituições de modo a subsidiar suas políticas e necessidades

122. Considerando as informações prestadas pelo gestor, constata-se que a gestão da base tem sido realizada de maneira a incluir novos processos e controles, permitindo que o cadastro evolua com o tempo. Ademais, constata-se que a principal fonte de informações para a marcação de óbitos, a partir de 2017, é a base da Central de Informações do Registro Civil (CRC) por meio de convênio com a Arpen/Brasil.

123. A Seção 4.1 apresentou os resultados dos cruzamentos de informações com bases da Administração Pública Federal e Estadual, nos quais foram identificadas 3.359.609 inscrições de CPF com indícios de óbito. Deve-se ressaltar que, desse total, 3.107.103 foram identificados nas bases do Sirc ou Sisobi, que são os principais registros responsáveis pelo armazenamento dos dados dos registros civis de nascimento, casamento, óbito e natimortos da Administração Pública Federal.

124. Considerando que a RFB, na Nota/Cocad 57/2020, reconhece que existem maiores dificuldade para registrar um óbito, maiores esforços são necessários para identificá-los em outras fontes de informação. Por exemplo, poderiam ser realizados esforços para identificar os instituidores de pensão por morte na folha de pagamentos do INSS, pois nesse caso, existe o incentivo de registrar o óbito para o recebimento de uma pensão.

125. Conforme informações da RFB na Nota/Cocad 92, de 31 de julho de 2020:

A fonte de informação para suspender um CPF por indício de óbito são os dados constantes do legado de óbito do registro civil. Via de regra, os dados do legado não permitem a vinculação do registro de óbito ao CPF. A ausência de dados como data de nascimento do falecido, por exemplo, inviabiliza a correta vinculação. Nesses casos, para evitar danos sérios à vida civil do cidadão, o CPF não passa de pronto à situação Titular Falecido, e irá migrar para a situação suspensa por indício de óbito. A RFB recebeu recentemente da ARPEN/BR arquivo contendo recorte do legado de óbito, entretanto, devido às inconsistências nele contidas, não foi possível efetivar com segurança as suspensões correspondentes.

126. Observa-se que as bases de registro civil (CRC, Sirc e Sisobi) possuem poucos registros de certidões de óbito anteriores a 2015 com a qualidade necessária. Esse legado de certidões de óbito, que não estão armazenadas de forma estruturada ou contém problemas de qualidade, pode ser um dos motivos para o elevado número de inscrições de CPF regulares em relação à população do país.

127. Em relação ao Sirc, a Fiscalização Contínua de Benefícios da Assistência de 2019 (TC 009.922/2019-9) apontou diversos problemas de qualidade de dados. Em essência, os achados do Sirc concentraram-se: na incompletude da base de certidões registradas, resultante principalmente da baixa disponibilidade de dados de certidões mais antigas; na indisponibilidade de dados relativos

a averbações de divórcios e anotações; e em problemas de qualidade dos dados, especialmente os relativos à identificação da pessoa como CPF, declaração de nascido vivo (DNV) e declaração de óbito (DO).

128. As deficiências da base foram endereçadas aos órgãos responsáveis, por meio dos itens 9.5 e 9.6 do Acórdão 1.123/2020-TCU-P, de relatoria do Min-subst. Marcos Bemquerer. Conforme relatado naquele trabalho, é necessária a atuação integrada de diversos órgãos para aprimorar as informações de nascimentos, casamentos e óbitos dos registros públicos e fazer com que elas conversem adequadamente com outras bases cadastrais da Administração Pública.

129. Apesar disso, a partir dos processos de qualificação de dados (*Data Quality*) realizados no âmbito da FCB de Assistência Social de 2019 e neste trabalho, foi possível utilizar os dados cadastrais dos indivíduos com elevado grau de certeza para realizar os cruzamentos pelas equipes de fiscalização nos dois trabalhos.

130. Conforme demonstrado pelos achados apontados na Seção 4.1, é necessário pontuar que problemas históricos de qualidade de dados cadastrais, bem como de maior compartilhamento de bases entre órgãos, ainda precisam ser enfrentados para melhoria das informações no Setor Público. A atuação dos órgãos de controle interno e externo, em conjunto com os gestores das bases, pode contribuir para o aprimoramento contínuo das informações e controles sobre os cadastros.

131. A partir do Decreto 9.723/2019 a base CPF da RFB está caracterizada como o principal cadastro de pessoas do Brasil, o que de alguma forma conflita com a sua origem, pois tratava-se inicialmente apenas de cadastro de contribuintes, ou seja, com forte apelo tributário. Essa mudança de enfoque histórica pode trazer dificuldades para uma boa atualização da base. É inevitável que a RFB, diante do quadro normativo posto, adote postura mais adequada em gerir um cadastro de informação de pessoas em geral no Brasil, mesmo que tal informação não tenha efeitos tributários.

132. Nesse sentido, há proposta de encaminhamento para recomendar à RFB, com fundamento no art. 250, inciso III, do Regimento Interno do TCU, que adote medidas para obter acesso a outras bases que tragam informações de falecimentos ou possam servir para comprovar nascimentos, a fim de aprimorar continuamente a base cadastral do CPF. Poderiam contribuir para essa qualificação, por exemplo, as bases do Sistema Nacional de Registros Cíveis – Sirc, do Sistema de Controle de Óbitos – Sisobi, das folhas de pagamentos de benefícios do INSS – Maciça, do Cadastro Único, das folhas de pagamento de agentes públicos federais – Siape/Extrasiape e das folhas de pagamento de agentes públicos estaduais e municipais.

7. Conclusão

133. O presente Relatório de acompanhamento (Racom) de dados marca o segundo ciclo de avaliação das ações de combate à Covid-19 nas áreas de Assistência Social, Previdência Social e Gestão Tributária, vinculadas à atuação da SecexPrevidência.

134. Neste trabalho, realizou-se a avaliação da base de dados do CPF, gerida pela RFB, em atendimento à deliberação constante no item 9.1.5 do Acórdão 1.638/2020-TCU-P, de relatoria do Min. Bruno Dantas.

135. Foram empreendidas análises de credibilidade da base, de avaliação de tipologias (ou trilhas de auditoria de dados), do processo de regularização de ofício realizado em 2020 e dos quantitativos de inscrições do CPF frente à população brasileira.

136. A análise de credibilidade identificou que os campos da base de dados do CPF apresentam um alto percentual de credibilidade para completude, validade e consistência, considerando um percentual máximo tolerável de 2% para falhas ou inconsistências. Constata-se que campos referentes

a endereço possuem percentual de inconsistências maior que o tolerável, contudo esses campos possuem poucos incentivos para sua atualização, especialmente para pessoas físicas não obrigadas a declarar Imposto de Renda.

137. O campo referente a data da criação da inscrição do CPF possui o maior percentual de inconsistências, com 23,4% dos registros sem preenchimento da informação. Considerando que essa situação afeta, em geral, registros com mais de 30 anos e, normalmente, os registros com data de criação mais recentes são utilizados em casos de fraude (ex.: criação de uma inscrição de CPF para a solicitação de um Benefício de Prestação Continuada ou Aposentadoria Rural), a situação possuiria um risco reduzido.

138. A Seção 4.1 apresentou os resultados dos cruzamentos de informações com bases da Administração Pública Federal e Estadual, nos quais foram identificadas 3.359.609 inscrições de CPF regulares com indícios de óbito. Desse total, deve-se ressaltar que 3.107.103 (92%) registros foram identificados nas bases do Sirc ou do Sisobi, que são os principais registros responsáveis pelo armazenamento dos dados dos registros civis de nascimento, casamento, óbito e natimortos da Administração Pública Federal. Dessa maneira, considerando que a RFB reconhece que existem maiores dificuldades para registrar um óbito, deveria envidar esforços para identificá-los em outras fontes de informação.

139. Os dados dos registros de óbito expedidos a partir de 1973 deveriam ter sido inseridos no Sirc, em até cinco anos, a partir da publicação da Lei 11.977/2009, que dispõe:

Art. 39. Os atos registrais 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.

140. Contudo, constata-se que as bases do Sirc, do Sisobi e do CRC (Arpen/Brasil) possuem poucos registros das certidões de óbito anteriores a 2015. Esse legado de certidões de óbito, que não estão armazenadas de forma estruturada, pode ser um dos motivos para o elevado número de inscrições de CPF regulares em relação à população do país.

141. As deficiências da base do Sirc foram endereçadas aos órgãos responsáveis, por meio dos itens 9.5 e 9.6 do Acórdão 1.123/2020-TCU-P, de relatoria do Min-subst. Marcos Bemquerer. Conforme relatado naquele trabalho, é necessária a atuação integrada de diversos órgãos para aprimorar as informações de nascimentos, casamentos e óbitos dos registros públicos e fazer com que elas se integrem adequadamente com outras bases cadastrais da Administração Pública.

142. A partir dos resultados obtidos, há propostas de encaminhamento das listas de indícios e metodologia empregadas para revisão de controles pelo órgão gestor e possíveis incorporações de novas bases ao processo, visando ao aprimoramento contínuo da base de CPF.

143. Em relação à regularização de ofício das inscrições de CPF, o procedimento da RFB atendeu emergência oriunda da pandemia do Covid-19, em que o foco era diminuir o risco de exclusão de pessoas ao auxílio emergencial. Conforme enfatizado pela equipe, o procedimento é temporário e deve ser realizado um acompanhamento pelo TCU das ações de reversão.

8. Proposta de encaminhamento

144. Ante o exposto, submetem-se os autos à consideração superior propondo:

a) **determinar** à Secretaria da Receita Federal do Brasil (RFB), com fulcro na Lei 8.443/92, art. 43, I c/c art. 250, inciso II, do Regimento interno do TCU, que:

- a.1) **no prazo de 120 dias**, indique as providências ou os controles que serão tomados para reduzir o número de inconsistências identificadas no Cadastro de Pessoas Físicas (CPF) quanto:
- a.1.1) à avaliação de credibilidade;
 - a.1.2) aos registros de óbito em bases da Administração Pública Federal ou Estadual que se encontrem em situação regular;
 - a.1.3) aos registros de supercentenários;
 - a.1.4) às inscrições suspensas há 11 anos ou mais;
- a.2) **no prazo de 120 dias**, apresente plano de ação visando estabelecer procedimentos para notificação ativa do cidadão em caso de suspensão de seu CPF, bem como regulamentar prazo limite para um CPF constar na situação “suspensa” após a notificação do cidadão, incluindo no processo de trabalho a forma de tratamento desses CPFs após o prazo estabelecido;
- b) **recomendar** à Secretaria da Receita Federal do Brasil (RFB), com fundamento no art. 250, inciso III, do Regimento Interno do TCU, que
- b.1) adote medidas para obter acesso a outras bases que tragam informações de falecimentos ou possam servir para comprovar nascimentos, a fim de aprimorar continuamente a base cadastral do CPF, a exemplo do Sistema Nacional de Registros Cíveis – Sirc, do Sistema de Controle de Óbitos – Sisobi, das folhas de pagamentos de benefícios do INSS – Maciça, do Cadastro Único, das folhas de pagamento de agentes públicos federais – Siape/Extrasiape e das folhas de pagamento de agentes públicos estaduais e municipais - TCE;
- c) **encaminhar** à Secretaria da Receita Federal do Brasil (RFB) os indícios identificados e a metodologia detalhada de avaliação do Cadastro de Pessoas Físicas (CPF);
- d) **restituir** os autos à SecexPrevidência para continuidade do acompanhamento.

Secex Previdência, DCAD, em 26/8/2020.

(Assinado eletronicamente)

Luiz Henrique Batistuta Gomide

AUFC – Mat. 9.981-3

(Assinado eletronicamente)

Theuryn Saches Loureiro Figueiredo

AUFC – Mat. 3.071-6

(Assinado eletronicamente)

Luíza da Silva Jaques

AUFC – Mat. 10.691-7

Coordenadora

(Assinado eletronicamente)

Daniel Moreira Guilhon

AUFC – Mat. 7.668-6

Apêndice I

Metodologia detalhada

1. Os trabalhos da Diretoria de Análise de Dados e Tecnologia da Informação (Dcad/SecexPrevidência), em geral, seguem 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).
2. O Crisp-DM sugere um 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.
3. Tal como preceituado pelo Crisp-DM, as etapas neste trabalho foram realizadas de forma cíclica, 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 metodologia deste trabalho contou também com atividades típicas das fiscalizações do TCU, como requisições e encaminhamentos aos órgãos gestores de programas e bases avaliados.
4. Este segundo Racom de dados avaliou a base de CPF segundo metodologia empregada há pelo menos 5 (cinco) anos em processos de Fiscalização Contínua da SecexPrevidência para avaliação de políticas da sua clientela, como os pagamentos do INSS, Cadastro Único e Bolsa Família. A metodologia inclui procedimentos de avaliação de credibilidade e tipologias (ou trilhas de auditoria).
5. A credibilidade dos dados consiste em uma série de testes realizados para que se estabeleça o grau de confiabilidade dos bancos de dados. 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 chaves que deveriam ser únicas; 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 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.
6. Já 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 ou cadastro.
7. Para a tipologia de super centenários, as pessoas com situação do CPF regular foram filtradas pelo critério de idade, indicando-se aquelas com idades flagrantemente incorretas.
8. Já para a tipologia de óbitos, realizou-se o confronto das informações do CPF com outras bases disponíveis no TCU com dados de falecidos, como: Sirc, Sisobi, Maciça, CadÚnico e de instituidores de pensões no Setor Público.
9. Para minimizar os casos de falsos positivos na etapa de análise, foram construídas bases derivadas com dados de maior qualidade utilizando técnicas de qualificação de dados (*Data Quality*), como higienização e enriquecimento de dados.
10. Essa qualificação foi realizada por meio da avaliação dos dados da pessoa frente aos dados cadastrados na base de CPF da RFB. Assim, após os cruzamentos das bases pelo CPF, avaliou-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 (se existir). Essa avaliação gera uma nota de similaridade, que permite classificar os registros referentes a uma pessoa como iguais, distintos, inconclusivos ou nulos.

11. Os registros em que a nota de similaridade gerada permite concluir, com elevada certeza, que se trata da mesma pessoa, são utilizados nas análises empregadas pelas tipologias para identificação de indícios de erros.

12. A seguir serão detalhados os cruzamentos realizados para o achado de óbitos, relativos às diversas bases utilizadas (item 2.1).

Sirc

13. O Sistema Nacional de Informações de Registro Civil (Sirc) é responsável pelo armazenamento dos dados dos registros civis de nascimento, casamento, óbito e natimortos.

14. Os registros do SIRC foram comparados com os registros do CPF, utilizando o campo CPF como chave, e, posteriormente, os dados cadastrais foram comparados para verificar se correspondem ao mesmo indivíduo, conforme metodologia de qualificação já apresentada.

Instituidor de pensão por morte do INSS (Maciça)

15. A folha de pagamentos do INSS possui o registro dos instituidores de pensão por morte. Os registros dos instituidores foram comparados com os registros do CPF, utilizando o campo CPF como chave, e, posteriormente, os dados cadastrais (conjunto formado pelo nome do titular, nome da mãe, data de nascimento) foram comparados para verificar se correspondem ao mesmo indivíduo, conforme metodologia de qualificação já apresentada.

Instituidor de pensão nas folhas de pagamento de Estados e Municípios (TCE)

16. Os registros das bases foram comparados com os registros do CPF, utilizando o campo CPF como chave. As folhas de pagamentos Estaduais e Municipais possuem uma limitação por não conter o campo de nome da mãe e diversos registros não possuem a informação da data de nascimento. Dessa maneira, os campos de nome e data de nascimento foram comparados utilizando a distância de *levenshtein* e multiplicados pelos pesos de 0,75 e 0,25, respectivamente, para o cômputo da nota final. Para os registros que não tinham a data de nascimento, somente a nota relativa ao nome foi utilizada.

17. Os dados cadastrais foram considerados como referentes a pessoas distintas se a nota final foi abaixo de 50, inconclusivos se a nota final ficou entre 50 e 75 e referentes à mesma pessoa se a nota final foi igual ou acima de 75.

Cadastro Único

18. O Cadastro Único (Cadun) é responsável por armazenar as informações sobre as famílias brasileiras em situação de pobreza e extrema pobreza. A Caixa Econômica Federal, responsável pela operacionalização do CADUN, por meio de controles internos, realiza a marcação dos registros com suspeita de óbito. A partir desse subconjunto de registros, os dados das bases foram comparados utilizando o campo CPF como chave e, posteriormente, os dados cadastrais foram comparados para verificar se correspondem ao mesmo indivíduo, conforme metodologia já apresentada.

Sisobi

19. O Sistema Informatizado de Controle de Óbitos (Sisobi) é responsável por armazenar os registros civis de óbito. Os registros do Sisobi foram comparados com os registros do CPF, utilizando o campo CPF como chave, e, posteriormente, os dados cadastrais foram comparados para verificar se correspondem ao mesmo indivíduo, conforme metodologia já apresentada.

Instituidor de pensão nas folhas de pagamento de Siape e Extrasiape

20. O Sistema Integrado de Administração de Recursos Humanos (Siape) possui os dados da folha de pagamento do Poder Executivo, enquanto os conjuntos das folhas de pagamento do Poder Legislativo e Judiciário, Banco Central e Forças Armadas não possuem centralização, sendo consolidados pelo TCU em um conjunto de dados chamado de Extrasiape. A partir de marcações nessas bases, é possível identificar se um indivíduo é instituidor de pensão por morte ou está falecido.

21. Os registros das bases foram comparados com os registros do CPF, utilizando o campo CPF como chave. Os campos de nome, nome da mãe e idade foram comparados utilizando a distância de *levenshtein* e multiplicados pelos pesos, conforme a Tabela 1, para calcular a nota final da comparação. Os dados cadastrais foram considerados como referentes a pessoas distintas se a nota final foi abaixo de 50, inconclusivo se a nota final foi igual ou acima de 50 e abaixo de 75 e referentes às mesmas pessoas se a nota final foi igual ou acima de 75.

Tabela 1: Peso das variáveis

Campo	Situação	Nome	Nome da mãe	Data de nascimento
Peso	Com nome da mãe	38,10	38,10	23,80
Peso	Sem nome da mãe	75,00	-	25,00

Fonte: elaboração própria.